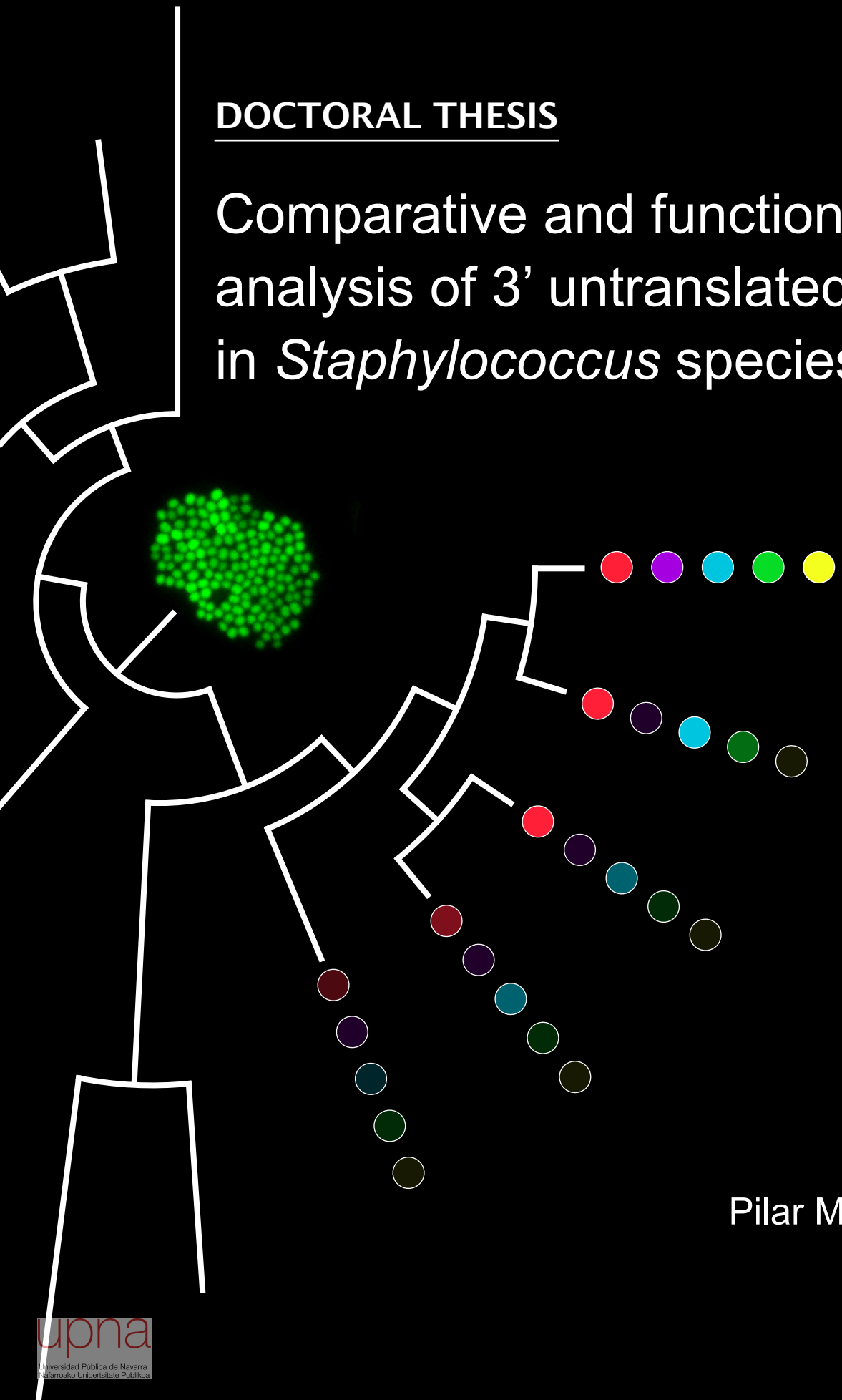
Programa de Doctorado en Biotecnología - Universidad Pública de Navarra
Instituto de Agrobiotecnología - Consejo Superior de Investigaciones Científicas

DOCTORAL THESIS

Comparative and functional analysis of 3’ untranslated regions in *Staphylococcus* species



Pilar Menéndez-Gil

Pamplona, 2020

upna
Universidad Pública de Navarra
Nafarroako Unibertsitate Publikoa

TESIS DOCTORAL

Comparative and functional analysis of 3' untranslated regions in *Staphylococcus* species

Memoria Presentada por

PILAR MENÉNDEZ GIL

Para optar al grado de Doctor por la Universidad Pública de Navarra

Director:

Dr. ALEJANDRO TOLEDO ARANA

Pamplona, 2020

Dr. ALEJANDRO TOLEDO-ARANA, Científico Titular del Consejo Superior de Investigaciones Científicas (CSIC), responsable del Laboratorio de Regulación Génica Bacteriana del Instituto de Agrobiotecnología (IdAB, CSIC-UPNA-GN)

INFORMA:

Que la presente memoria de Tesis Doctoral titulada "**Comparative and functional analysis of 3' untranslated regions in *Staphylococcus* species**" que ha sido elaborada por Doña **PILAR MENÉNDEZ GIL**, ha sido realizada bajo su dirección y que cumple las condiciones exigidas por la legislación vigente para optar al grado de Doctor.

Y para que así conste, firma la presente en Pamplona, a 20 de abril de 2020.

Fdo. Alejandro Toledo Arana

"I know why you're here, Neo. I know what you've been doing... why you hardly sleep, why you live alone, and why night after night, you sit by your computer. You're looking for him. I know because I was once looking for the same thing. And when he found me, he told me I wasn't really looking for him. I was looking for an answer. It's the question that drives us, Neo. It's the question that brought you here. You know the question, just as I did. The answer is out there, Neo, and it's looking for you, and it will find you if you want it to."

Trinity. The matrix, 1999

"The history of evolution is that life escape all barriers. Life breaks free. Life expands to new territories. Painfully, perhaps even dangerously. But life finds a way."

Ian Malcom. Jurassic Park, 1993

A mis padres

A Iñigo

AGRADECIMIENTOS

AGRADECIMIENTOS

Finalmente después de cuatro años y medio esta etapa llega a su fin y en estos días de tanta incertidumbre sobre el futuro me gustaría dar las gracias a todas aquellas personas que de alguna manera me han ayudado a avanzar y acabar esta Tesis.

En primer lugar, me gustaría dar las gracias a Ale por la oportunidad que me ha dado al poder realizar el doctorado con él. Muchas gracias por tener siempre tiempo para resolver dudas y discutir resultados, por todo lo que me has enseñado y por sacarme de los jardines en los que tanto me gusta meterme.

A todo el grupo de Regula Genes que han permitido que el labo sea como una segunda casa. Al grupo inicial formado por Carlos, Arancha y Naiara, al que luego se unieron Laura, Laurene, Leticia, Sergio y Jaio. A Pili y Marta, que aunque hayan sido de las últimas en llegar me han dado una bonita amistad. A Lorena a quien siempre la he considerado parte del grupo.

A Iñigo Lasa y a todas las personas que forman o han formado parte de su grupo: Cristina, Begoña, Maite, Carmen, Saioa, Sonia, Bea, Pedro, Carmen. Gracias por todos los seminarios con sesión de tortilla o pastas!

A Pascale Romby, Isabelle Caldelari y a todo su grupo de Estrasburgo: Stefano, Emma, Javi, Eva, Lucas, Laura, David y Anne-Catherine. Merci beaucoup por abrirme las puertas y hacerme sentir bienvenida.

A todas las personas que forman parte del Idab, en especial a Víctor por ser una gran persona, el Idab no sería lo mismo sin ti.

A mis chics, Ana, Bea, Nat, Lau, Ani y Sara, que siempre están ahí para apoyarme y animarme. A Marina, Portal, Vicky y Carmen por nuestros skypes y a Aaraby por nuestro encuentros anuales en Dk o España.

A toda mi familia, que me demuestra que siempre puedo contar con ella, que me ha enseñado que todas las cosas buenas hay que celebrarlas y que cuando ocurren malas hay que mantenerse unidos y fuertes para poder afrontarlas y seguir adelante.

A Marivi y Moncho, que han sido y son unos geniales padres de adopción, siempre estando cuando los necesito. Gracias a Ana por nuestras sesiones de cine, especialmente en estos días de confinamiento.

A mi hermana, que siempre está presente y que se desespera cuando me intenta explicar sus palabros raros fiscales del día a día. A Adri que llegó hace poco a nuestras vidas y nos conquista a todos con su sonrisa. A Bruce, que aunque ya no esté, no olvidaré la compañía que me dio.

A mis padres, por el increíble apoyo que me dan, que han sufrido esta tesis como si fuese suya y que son los que más ganas tienen de que la acabe. Gracias por dármelo todo y por quererme tanto, si he llegado hasta aquí ha sido gracias a vosotros.

Por último, a Iñigo, que aún estando lejos ha soportado todos mis desahogos denominados “la turra”. Gracias por todo el amor y ánimo que me has dado. Home is wherever I’m with you.

Este trabajo ha sido realizado gracias a la financiación recibida a través de los siguientes proyectos de investigación:

European Research Council Consolidator grant ERC-coG-2014-646869:

“High-throughput *in vivo* studies on post-transcriptional regulatory mechanisms mediated by bacterial 3’-UTRs”.

Ministerio de Economía y Competitividad. BFU2011-56698-P:

“Regulación post-transcripcional mediada por las regiones 3' no traducidas del RNA mensajero en bacterias”.

Consejo Superior de Investigaciones Científicas. PIE-201540I013:

“Regulación post-transcripcional mediada por RNAs y proteínas de unión a RNA en bacterias patógenas”.

TABLE OF CONTENTS

SUMMARY

SUMMARY

A messenger RNA (mRNA) molecule is composed of a coding sequence (CDS) flanked by two untranslated regions (UTRs), the 5'UTR and the 3'UTR, respectively. In eukaryotes, 3'UTRs are key components of post-transcriptional regulatory mechanisms. Shortening or deregulation of these regions is associated with diseases such as cancer and metabolic disorders. Comparatively, little is known about the functions of 3'UTRs in bacteria. Over the past few years, researchers have shown how bacterial 3'UTRs can act as reservoirs for *trans*-acting non-coding RNAs (ncRNAs) as well as regulators of the expression of their own mRNAs. These recent findings place 3'UTRs as important players in the regulation of key bacterial processes such as virulence, iron metabolism and biofilm formation. As a consequence, 3'UTRs could play a broader role than initially anticipated with many questions still unanswered. For example, are 3'UTR sequences preserved within and between bacterial species? How often do orthologous genes from closely-related bacteria contain conserved functional 3'UTRs? Do nucleotide variations affect 3'UTRs functionality and, thus, mRNA expression?

In this Thesis, we performed genome-wide comparative analyses of mRNAs encoding orthologous proteins in the genus *Staphylococcus.* We discovered that most of these mRNAs contained non-conserved 3'UTR sequences. In contrast, 5'UTRs were more conserved than 3'UTRs suggesting an evolutionary bias within 3'UTRs. Transcriptional mapping of

different staphylococcal species confirmed that 3'UTRs were also variable in length. To test if the 3'UTR variability could affect protein expression, we created chimeric mRNAs by fusing the *Staphylococcus aureus icaR, ftnA* and *rpiRc* CDSs with the 3'UTRs of orthologous genes from several staphylococcal species. Northern and Western blot analyses revealed that the nucleotide variations in the 3'UTR sequences altered the mRNA and protein levels. This suggested that the 3'UTRs from orthologous mRNAs may have distinct functional roles. Next, we demonstrated that the differences in the mRNA and protein levels could be explained by the presence or absence of specific regulatory elements within the 3'UTRs. We also showed that sequence variations in the 3'UTRs might occur through different processes, including gene rearrangements, local nucleotide changes and transposition of insertion sequences. Finally, we extended the genome wide comparative analyses to already described functional 3'UTRs and the entire set of mRNAs from *Escherichia coli* and *Bacillus subtilis.* The results suggested 3'UTR variability to be a widespread phenomenon in bacteria.

In summary, this Thesis shows how natural nucleotide variations in 3'UTRs affect mRNA expression. This common occurrence might be responsible for creating different functional species-specific regulatory roles and, ultimately, bacterial diversity through the course of evolution.

Sections of this Doctoral Thesis have been published in:

- Nucleic Acids Research 48 (5): 2544–2563 (2020). **Menendez-Gil P,** Caballero CJ, Catalan-Moreno A, Irurzun N, Barrio-Hernandez I, Caldelari I and Toledo-Arana A**.** Differential evolution in 3'UTRs leads to specific gene expression in *Staphylococcus*.

Other publications in which the candidate has participated:

- Nucleic Acids Research 46 (3): 1345-1361 (2018). Caballero CJ, **Menendez-Gil P**, Catalan-Moreno A, Vergara-Irigaray M, García B, Segura V, Irurzun N, Villanueva M, Ruiz de los Mozos I, Solano C, Lasa I and Toledo-Arana A. The regulon of the RNA chaperone CspA and its autoregulation in *Staphylococcus aureus*.

- Methods in Molecular Biology 2106: 41-58 (2020). **Menendez-Gil P**, Caballero CJ, Solano C and Toledo-Arana A. Fluorescent molecular beacons mimicking RNA secondary structures to study RNA chaperone activity.

RESUMEN

RESUMEN

Una molécula de RNA mensajero (mRNA) está compuesta por una región codificante (CDS) flanqueada por dos regiones no traducidas (UTRs), la 5'UTR y la 3'UTR, respectivamente. En eucariotas, las 3'UTRs son elementos claves en la regulación post-transcripcional. El acortamiento o la desregulación de estas regiones está asociado con diversas enfermedades como el cáncer o trastornos metabólicos. En comparación, el conocimiento de las funciones de las 3'UTRs en bacterias es mucho menor. Durante los últimos años, se ha demostrado que las 3'UTRs bacterianas pueden regular la expresión del mRNA que las contiene al igual que pueden ser reservorios de RNAs no codificantes (ncRNAs). Se les ha vinculado en la regulación de procesos bacterianos esenciales como la virulencia, el metabolismo del hierro y la formación de biofilm. En consecuencia, las 3'UTRs pueden jugar un papel mucho más amplio de lo que se había sugerido inicialmente. Sin embargo, todavía hay muchas cuestiones por resolver, por ejemplo, ¿cuál es el grado de conservación de sus secuencias tanto a nivel intra- como inter-especie? ¿Con qué frecuencia los genes ortólogos de bacterias estrechamente relacionadas mantienen conservadas las 3'UTRs? ¿Cómo afectan las variaciones nucleotídicas a la funcionalidad de las 3'UTRs y, en consecuencia, a la expresión de los mRNAs?

En esta Tesis, por medio de análisis comparativos globales de los mRNAs que codifican genes ortólogos en distintas especies del género *Staphylococcus,* descubrimos que la mayoría de los mRNAs contienen

3’UTRs que no están conservadas. Por el contrario, las 5’UTRs se encuentran más conservadas que las 3’UTRs, sugiriendo un sesgo evolutivo hacia las 3’UTRs. La realización de mapas transcriptómicos de diversas especies de *Staphylococcus* confirmó que las 3’UTRs de genes ortólogos presentaban variaciones en su longitud además de los cambios en sus secuencias. Con el fin de investigar si esta variabilidad afectaba a la expresión proteica, se crearon mRNA quiméricos fusionando las CDSs de los genes *icaR, ftnA* y *rpiRc* de *Staphylococcus aureus* con las 3’UTRs de los mRNAs ortólogos de diferentes especies del mismo género. Los resultados mostraron que las variaciones nucleotídicas en las 3’UTRs alteraban tanto los niveles de mRNA como de proteína. Estos resultados sugerían que las 3’UTRs de genes ortólogos podían tener funciones distintas en cada especie bacteriana. Los cambios en los niveles de expresión podían ser explicados por la presencia o ausencia de elementos reguladores específicos localizados en las diferentes 3’UTRs. Las variaciones en las secuencias de las 3’UTRs podían ocurrir por diferentes procesos incluyendo reordenamientos genómicos, variaciones nucleotídicas locales y transposiciones de secuencias de inserción. Por último, extendiendo los análisis comparativos globales a 3’UTRs funcionales ya descritas, al igual que a los sets completos de mRNAs de *Escherichia coli* y *Bacillus subtilis*, descubrimos que la variabilidad de las 3’UTRs es un fenómeno que se encuentra extendido en las bacterias.

En resumen, esta Tesis demuestra que las variaciones nucleotídicas en las 3’UTRs, que ocurren de manera natural por la evolución, son capaces

de producir cambios en la expresión del mRNA. Esto puede crear funciones reguladoras específicas en una determinada especie, lo que posiblemente podría contribuir a una mayor diversidad entre las especies bacterianas.

Secciones de esta Tesis Doctoral han sido publicadas en:

- Nucleic Acids Research 48 (5): 2544–2563 (2020). **Menendez-Gil P,** Caballero CJ, Catalan-Moreno A, Irurzun N, Barrio-Hernandez I, Caldelari I and Toledo-Arana A**.** Differential evolution in 3'UTRs leads to specific gene expression in *Staphylococcus*.

Otras publicaciones en las que la doctoranda ha participado:

- Nucleic Acids Research 46 (3): 1345-1361 (2018). Caballero CJ, **Menendez-Gil P**, Catalan-Moreno A, Vergara-Irigaray M, García B, Segura V, Irurzun N, Villanueva M, Ruiz de los Mozos I, Solano C, Lasa I and Toledo-Arana A. The regulon of the RNA chaperone CspA and its autoregulation in *Staphylococcus aureus*.

- Methods in Molecular Biology 2106: 41-58 (2020). **Menendez-Gil P**, Caballero CJ, Solano C and Toledo-Arana A. Fluorescent molecular beacons mimicking RNA secondary structures to study RNA chaperone activity.

INTRODUCTION

INTRODUCTION

The central dogma of molecular biology states that the genetic information, encoded in the DNA, is transcribed into messenger RNAs (mRNAs) and then translated into proteins, the main effectors of the different biological processes of a cell (Crick, 1970). Based on that, it was thought that the levels of transcription determined exclusively the quantity of a particular protein. However, huge amount of evidence from last decades showed that gene expression is controlled at several levels. In addition to transcriptional control, gene expression can be modulated at the post-transcriptional and post-translational levels by a wide variety of molecular mechanisms. Altogether, these regulatory layers contribute to determine the final protein levels and its activity (Figure 1). The interplay and coordination between all these layers are essential for the correct adaptation of bacteria to the ever changing environment (Mata *et al.*, 2005; Yang *et al.*, 2014).

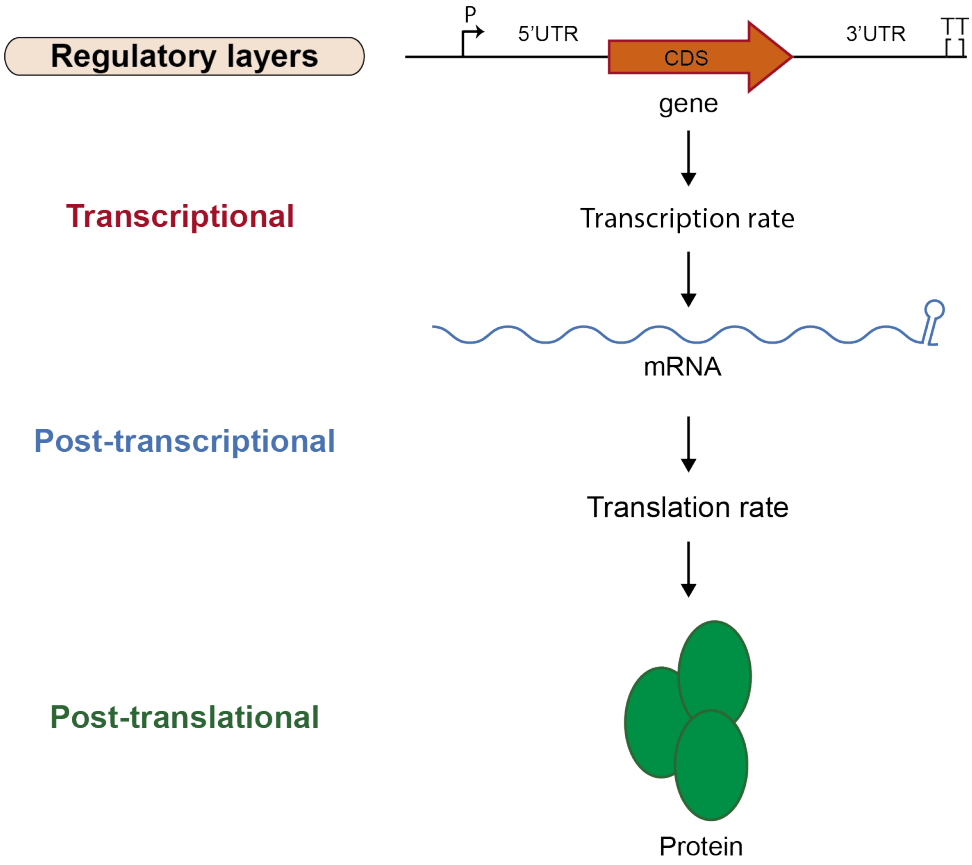


Figure 1. Main regulatory layers in bacteria. The main regulatory processes can be grouped in transcriptional (red), post-transcriptional (blue) and post-translational (green) controls. P, promoter; 5'UTR, 5' untranslated region; CDS, coding sequence; 3'UTR, 3' untranslated region; TT, transcriptional terminator.

Post-transcriptional regulation

Post-transcriptional regulation comprises all the regulatory mechanisms that affect the stability, the structure and the translation of mRNAs (Mata *et al.*, 2005). The development of deep RNA-seq technologies has enabled the scientific community to obtain complete transcriptomes of diverse bacterial models such as *L. monocytogenes, B. subtilis, S. aureus* and *E. coli* (Toledo-Arana *et al.*, 2009; Sharma *et al.*, 2010; Nicolas *et al.*, 2012; Kröger *et al.*, 2012; Ruiz de los Mozos *et al.*, 2013; Conway *et al.*, 2014; Lybecker *et al.*, 2014; Cohen *et al.*, 2016). Thanks to these

technological advances, now it is possible to determine the boundaries of each RNA molecule transcribed in a cell. Interestingly, it was found that the number of transcripts that are not coding for proteins was far larger than previously anticipated. Although the function of these non-coding regions is mostly unknown, it is expected that they are playing a role in post-transcriptional regulation.

Asides from non-coding RNAs, RNA-binding proteins such as ribonucleases (RNases) and RNA chaperones as well as the mRNA itself can modulate mRNA expression (Gripenland *et al.*, 2010; Wagner and Romby, 2015; Holmqvist and Vogel, 2018). Since these regulatory mechanisms are widely diverse, the most relevant paradigmatic examples will be discussed in this section, with a special interest in 3' untranslated regions (3'UTRs), which are the main focus of this Thesis.

Regulatory elements that modulate mRNA expression

Non-coding RNAs

Non-coding RNAs (ncRNAs) are regulatory RNAs that act by base pairing with their target mRNAs, which can be distally encoded from the ncRNA (*trans*-acting mechanism) or encoded in the opposite DNA strand (*cis*-acting mechanism) (Waters and Storz, 2009). These regulatory RNAs are involved in controlling many key processes for bacterial homeostasis including, for example, quorum sensing, virulence, iron metabolism, motility and biofilm formation (Wagner and Romby, 2015).

Trans-acting non-coding RNAs

Although the first studies identified these ncRNAs in intergenic regions of the chromosome, it has been shown that many *trans*-acting ncRNAs can be also originated from 5' and 3'UTRs (Chao *et al.*, 2012). *Trans*-acting ncRNAs can modulate negatively gene expression through different mechanisms. The most common is inhibition of mRNA translation by base pairing to the ribosome binding site (RBS) of the targeted mRNA (Figure 2A). As a consequence, there is a direct competition between the ncRNA and the ribosome for the RBS. If the concentration of the ncRNA is high enough, the ncRNA-mRNA interaction sterically prevents the binding of the ribosome to the mRNA (Gripenland *et al.*, 2010). This mechanism usually leads to RNA degradation, probably because a) the ribosome-naked mRNAs are more accessible to ribonucleases (Caron *et al.*, 2010; Wagner and Romby, 2015) and b) the degradosome machinery is commonly coupled to the ncRNA-mRNA complex (Morita *et al.*, 2005; Morita *et al.*, 2006).

Alternatively, instead of blocking the RBS, ncRNAs can pair other mRNA regions to promote mRNA processing (Figure 2A). Although, in these cases translation initiation is not directly affected, the ncRNA binding causes a decrease of protein expression by reducing the mRNA levels through RNA degradation (Wagner, 2009). This is the case of MicC from *Salmonella* that decreases *ompA* mRNA half-life. MicC binds downstream of the RBS, inside the coding sequence (CDS) of *ompA* mRNA, using a 12

base pairs (bp) interaction that does not interfere with protein translation initiation but recruits RNase E for mRNA degradation (Pfeiffer *et al.*, 2009).

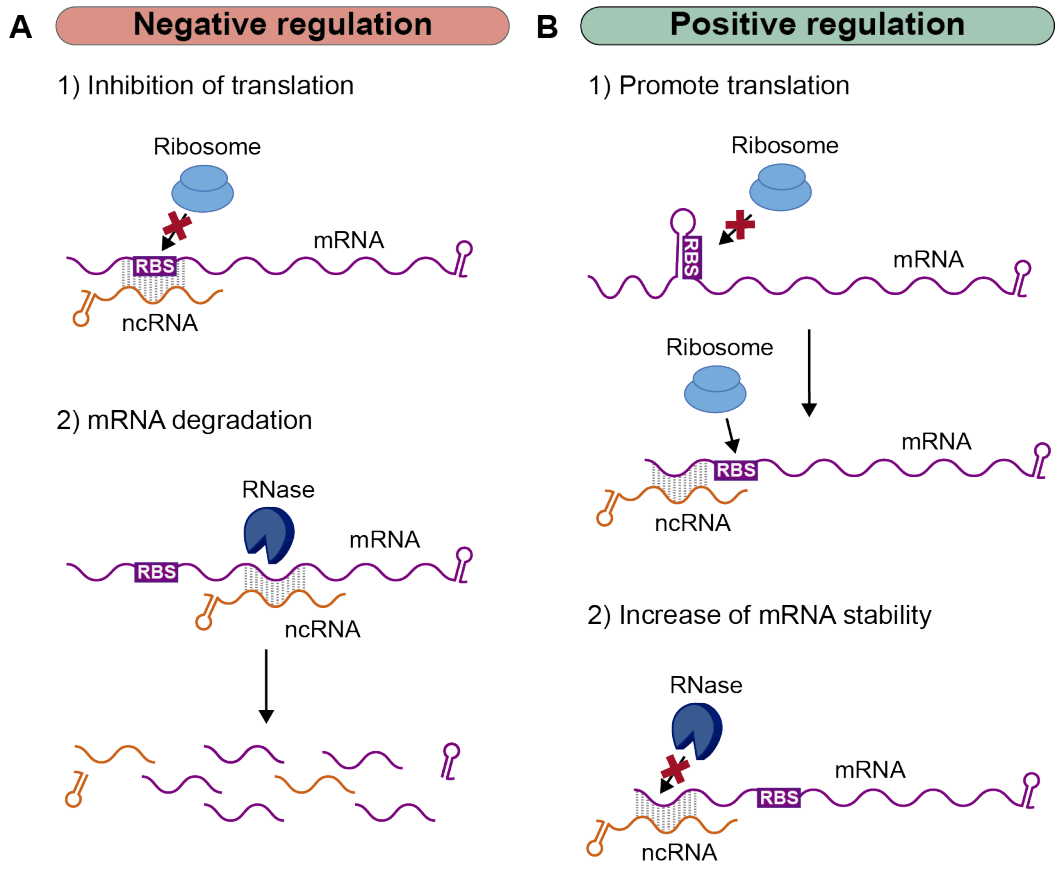


Figure 2. Main post-transcriptional regulatory mechanisms employed by *trans*-acting ncRNAs. (A) Negative regulation includes inhibition of translation and/or mRNA degradation. **(B)** Positive regulation can be promoted by facilitating translation and/or increasing mRNA stability.

To a lesser extent, ncRNAs can promote both translation and mRNA stability (Figure 2B). In the first case, the binding of the ncRNA to the 5'UTR of a target mRNA alleviates a secondary structure that is hindering the RBS. Thus, the ncRNA promotes ribosome accessibility and translation (Soper *et al.*, 2010). On the second case, the ncRNA binding to the mRNA avoids the access of an RNase and prevents its degradation.

For example, RydC from *Salmonella* stabilizes *cfa* mRNA through binding to its 5'UTR. RydC-*cfa* interaction interferes with RNase E processing, protecting *cfa* mRNA from degradation (Fröhlich *et al.*, 2013).

In addition, ncRNAs can cause early transcription termination as well as behave as sponge RNAs sequestering other ncRNAs or RNA-binding proteins (RBPs). In the latter case, sponge RNAs modify ncRNA or RBP activity by mimicking their targets. The interaction sponge RNA-ncRNA or sponge RNA-RBP prevents mRNA targeting by the ncRNA/RBP, avoiding control of mRNA translation (Bossi and Figueroa-Bossi, 2016).

Interestingly, the same regulatory RNA can modulate each of its mRNA targets in different ways. This is exemplified by RNAIII of *S. aureus*, which is activated by the Agr quorum sensing system according to bacterial cell density. RNAIII regulates several targets including surface proteins and virulence factors (Bronesky *et al.*, 2016). Binding of RNAIII to *spa, coa, sbi and rot* mRNAs causes inhibition of their translation and, in most of the cases, this is coupled to RNase III processing (Huntzinger *et al.*, 2005; Boisset *et al.*, 2007; Chevalier *et al.*, 2010; Chabelskaya *et al.*, 2014). On the other hand, RNAIII regulates positively the expression of *mgrA* and *hla* mRNAs, the former by stabilizing the mRNA while the latter by activating translation (Novick *et al.*, 1993; Gupta *et al.*, 2015). Overall, RNAIII manages to repress surface proteins and at the same time promotes secreted virulence factors (Bronesky *et al.*, 2016).

Cis-acting antisense RNAs

Antisense RNAs (asRNAs) are encoded in the opposite DNA strand of their target genes. They employ similar mechanisms as *trans*-acting ncRNAs including translation inhibition, mRNA degradation, mRNA stabilization and transcription attenuation (Brantl, 2007). They were first identified in plasmids, transposons and bacteriophages, in which they acted as regulators of the correct copy number of these mobile elements (Brantl, 2007; Waters and Storz, 2009; Gripenland *et al.*, 2010) (Figure 3A). Today it is known that asRNAs are widely distributed in all chromosomes and plasmids, generating a diverse variety of overlapping regions. For example, in *E. coli*, the asRNA GadY interacts with the *gadXW* mRNA at stationary growth phase leading to RNase cleavage of the operon mRNA and increased stabilization of *gadX* processed transcript (Opdyke *et al.*, 2004; Tramonti *et al.*, 2008). In *Vibrio anguillarum* the RNAα interacts with the *fatDCBAangRT* operon mRNA and entails transcription termination after *fatA*, reducing *angRT* expression (Stork *et al.*, 2007). Additionally, asRNAs are the main regulators of type I toxin-antitoxin systems, in which transcription of the asRNA represses translation of the toxic protein (Fozo *et al.*, 2008; Waters and Storz, 2009).

mRNAs carrying long 5' and 3'UTRs can also produce antisense regions. These occur when one or both mRNAs from divergent or convergent genes contain very long UTRs that pair with the one encoded in the opposite DNA strand, creating overlapping 5'UTRs and 3'UTRs, respectively (Lasa *et al.*, 2011; Lasa *et al.*, 2012) (Figure 3B-C).

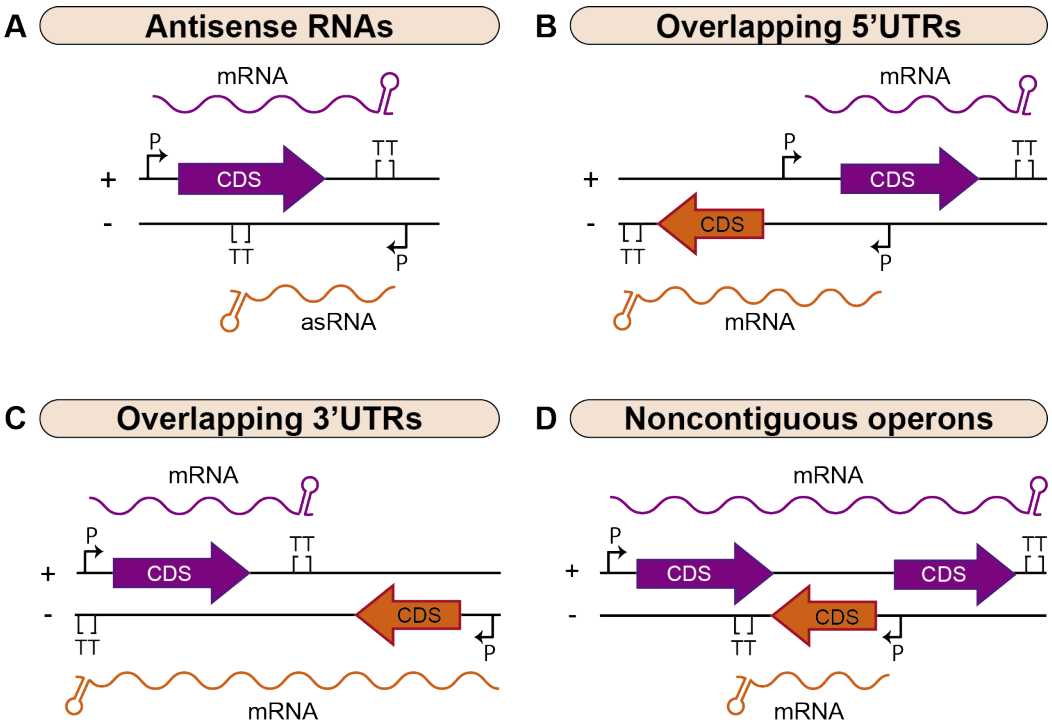


Figure 3. Types of antisense transcription in bacteria. (A) *Bona fide* antisense RNAs are generated by the presence of a promoter that produces a transcript that overlaps the mRNA encoded in the opposite DNA strand. **(B)** Overlapping 5'UTRs are produced when one (or both) promoters from divergent genes are encoded far away from the corresponding CDSs. **(C)** Overlapping 3'UTRs are generated when the transcriptional terminator of at least one convergent gene is far away from the corresponding CDS. **(D)** Noncontiguous operons consist in operons containing a gene(s) transcribed from the opposite direction to the rest of the operon. P, promoter; TT, transcriptional terminator; +/-, DNA strands. Adapted from (Lasa *et al.*, 2012).

Moreover, it is possible that transcription extends far away from the adjacent genes creating more complex transcriptional organizations, termed noncontiguous operons (Lasa *et al.*, 2011; Sáenz-Lahoya *et al.*, 2019) (Figure 3D). These consist of operons with a gene(s) transcribed in the opposite direction to the rest of the genes of the operon. The mRNA of the gene in the opposite direction has a perfect complementarity with the polycistronic transcript. This mRNA acts as both a coding RNA and an

antisense RNA (Sáenz-Lahoya *et al.*, 2019). This organization controls gene expression by at least two mechanisms. On the one hand, the gene configuration causes transcriptional interference, most probably due to collision between both RNA synthesis machineries. On the other hand, when transcription occurs, the overlapping RNA leads to a differential processing of the polycistronic transcript creating mRNA fragments with different stabilities (Sáenz-Lahoya *et al.*, 2019).

RNA-binding proteins

RNA binding proteins (RBPs) are a diverse group of proteins with the ability to regulate transcription termination, translation initiation and turnover of many mRNAs. RBPs can be functionally classified in several groups including RNA chaperones, RNA helicases, RNases and enzymes catalyzing RNA modifications. The latter are involved in the covalent modification of RNA nucleotides, while RNases are involved in the processing and turnover of RNA molecules. In contrast, RNA helicases and chaperones can modify secondary structures and promote or disrupt RNA-RNA or RNA-protein interactions. Unlike RNA helicases, RNA chaperones do not require ATP hydrolysis for their action (Rajkowitsch *et al.*, 2007; Babitzke *et al.*, 2009; Van Assche *et al.*, 2015).

RNA chaperones promoting binding of *trans*-acting ncRNAs

In Gram-negative bacteria, one of the most studied RNA chaperones is the Hfq protein (Host Factor I), which belongs to the Sm and Sm-like family. It promotes the interaction of specific ncRNAs with their

corresponding targets by increasing the local RNA concentrations and unfolding secondary structures that can hinder the ncRNA-mRNA binding (Møller *et al.*, 2002). The active form of Hfq is an hexameric complex that has a preference for A/U-rich regions (Valentin-Hansen *et al.*, 2004). In some cases, Hfq is able to recruit the degradosome by interacting with RNase E that promotes the clearance of the RNA duplex (Figure 4). Although the Hfq protein is also widely present in Gram-positive bacteria, its role has not yet been deciphered. In most of the ncRNA-mRNA interactions studied so far, with the exception of LhrA in *Listeria monocytogenes* (Nielsen *et al.*, 2009), Hfq was not required for the ncRNA interaction with their corresponding targets (Brennan and Link, 2007; Chao and Vogel, 2010).

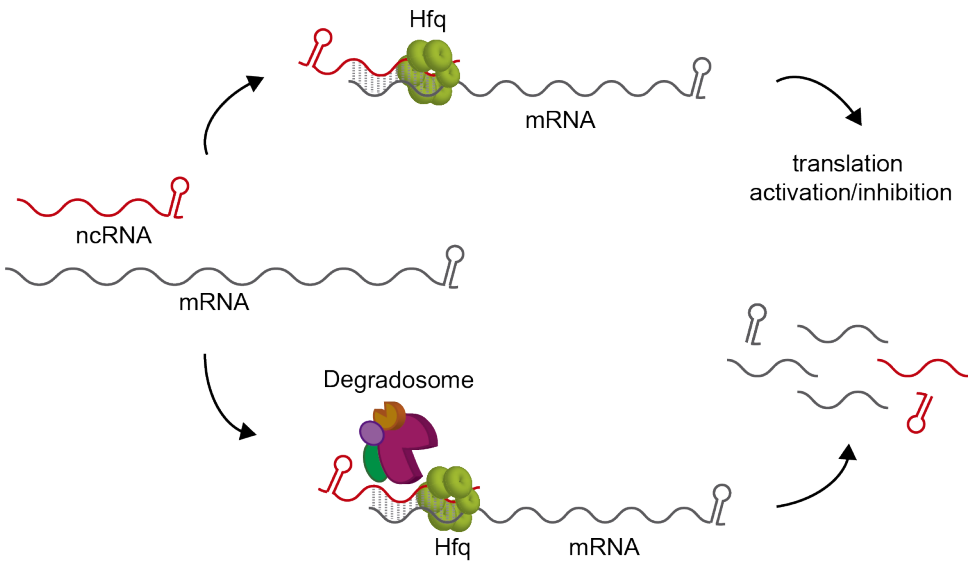


Figure 4. The RNA chaperone Hfq. In Gram-negative bacteria, Hfq helps binding of ncRNAs with their target mRNAs. In some cases, Hfq also interacts with the degradosome for rapid degradation of the ncRNA-mRNA complex.

In the last years, other RNA chaperones, which carry the ProQ/FinO domain, have also been found to participate in ncRNA-mediated regulation in bacteria, for example, RocC in *Legionella pneumophila*, ProQ in *S. enterica* and ProQ and FinO in *E. coli* (Attaiech *et al.*, 2017). *S. enterica* ProQ interacts and stabilizes more than 50 ncRNAs, whereas *L. pneumophila* RocC stabilizes RocR ncRNA to regulate genes involved in natural transformation. FinO is encoded in plasmids from the *IncF* family and it represses plasmid conjugation by stabilizing the ncRNA FinP (Attaiech *et al.*, 2017; Holmqvist and Vogel, 2018). Since the number of uncharacterized RNA chaperones is still huge, it is expected that more proteins will contribute to the activity of ncRNAs.

RNA chaperones modulating mRNA translation

The carbon storage regulator A (CsrA) is a small homodimeric chaperone that binds to GGA sequences usually located on the top of stem-loops at the RBS of mRNAs. As a consequence, in most of the cases CsrA hinders the RBS from the ribosomes inhibiting translation. This is the case of the *glcCAP* and *cstA* genes. To a lesser extent, CsrA can activate translation by destabilizing a hairpin that occludes the RBS or by blocking RNA degradation. CsrA is considered a global regulator of carbon metabolism, promoting glycolysis and suppressing gluconeogenesis. It is also involved in motility and biofilm formation, virulence, quorum sensing and oxidative stress response (Schubert *et al.*, 2007; Bhatt *et al.*, 2009; Timmermans and Van Melderen, 2010; Romeo and Babitzke, 2018). Interestingly, CsrA activity is regulated by the sponge ncRNAs CsrB and CsrC that contain 18

and 9 CsrA-binding sites, respectively. Depending on the CsrB/C concentration CsrA protein gets sequestered, inhibiting its activity (Liu *et al.*, 1997; Weilbacher *et al.*, 2003).

Cold shock proteins (CSPs) are a group of small RNA chaperones widespread among bacteria. Although all the members of this family contain a cold shock domain (CSD), not all of them are cold induced and some are required for adaptation to other stresses, such as oxidative and osmotic stresses. CSPs are able to reshape secondary structures to either modulate translation or inhibit transcription termination (Duval *et al.*, 2010; Phadtare and Severinov, 2010; Michaux *et al.*, 2017; Caballero *et al.*, 2018).

An additional group of RNA chaperones comprises some ribosomal proteins (r-proteins) that bind the 5'UTR of their mRNAs regulating their own expression (Nomura, 1970; Nomura, 1999). For example, the L20 r-protein, which is encoded in the *rpmL-rpmT* operon of *E. coli*, binds to the 5'UTR of the *rpmL-rpmT* transcript through two binding sites and competes with the ribosome for mRNA binding, affecting protein translation (Guillier *et al.*, 2005).

RNA helicases modifying RNA structures

RNA helicases use the energy of ATP hydrolysis to unfold RNA structures and displace RNA-bound proteins (Rajkowitsch *et al.*, 2007) (Figure 5). Together with DNA helicases, they can be classified based on their amino acid sequence in six superfamilies, among which proteins containing the

DEAD-box domain constitute the largest group. (Singleton *et al.*, 2007; Khemici and Linder, 2016). Helicases bind to double stranded RNAs to induce local strand separation, allowing the strands to be further used by ncRNAs and other RBPs. They are involved in diverse processes, some have a role in RNA turnover taking part in the degradosome, while others participate in ribosome biogenesis and translation initiation (Charollais *et al.*, 2003; Lehnik-Habrink *et al.*, 2010; Intile *et al.*, 2015). For instance, CsdA is required for the binding of the ncRNA DsrA to the 5'UTR of the *rpoS* mRNA at low temperatures. This binding de-sequesters the RBS allowing ribosome binding and protein translation (Resch *et al.*, 2010).

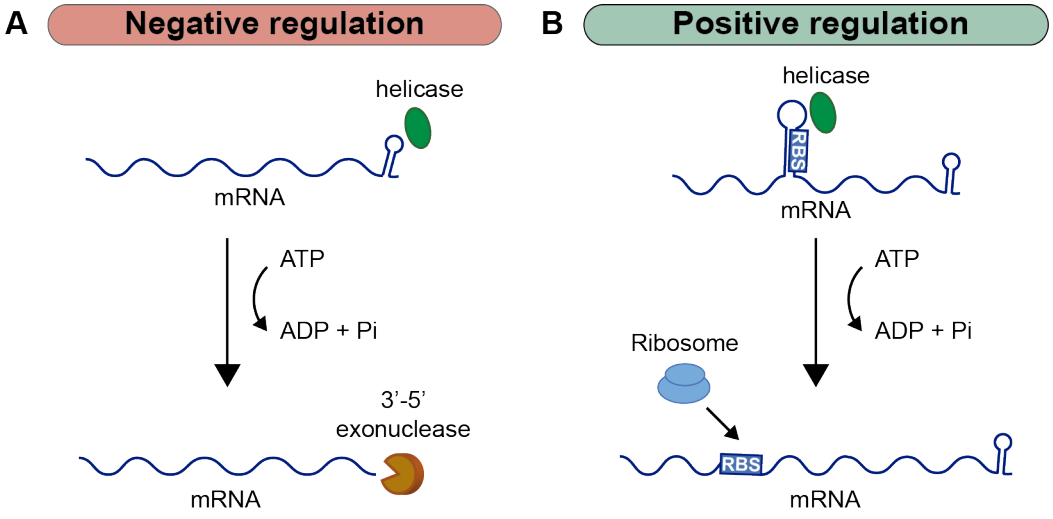


Figure 5. RNA helicases unfold secondary structures using ATP hydrolysis. Depending on the location of the RNA structures, RNA helicases can promote mRNA degradation **(A)** or translation **(B)**.

Ribonucleases involved in RNA processing and RNA degradation

Ribonucleases (RNases) are enzymes that control RNA processing and degradation, participate in the quality control of RNAs and are key players in the recycling of nucleotides. RNases can be classified in two broad groups: 1) endoribonucleases (endoRNases) which cleave within the RNA molecule and 2) exoribonucleases (exoRNases) that process the RNA from one of the two transcript ends (Arraiano *et al.*, 2010; Jester *et al.*, 2012; Mohanty and Kushner, 2016) (Figure 6).

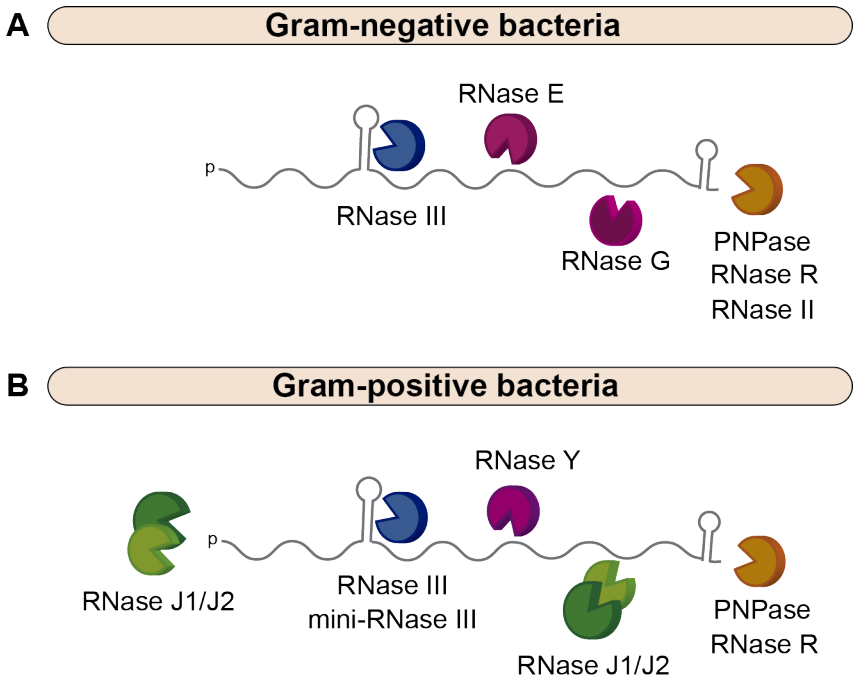


Figure 6. Main ribonucleases found in Gram-negative (A) and Gram-positive (B) bacteria. Single-stranded and double-stranded endoRNases are represented in pink and blue, respectively. 3'-5' exoRNases are colored in orange while RNases J1/J2 which possess both endoribonucleolytic and 5'-3' exoribonucleolytic activities are represented in green.

The best studied endoRNase is RNase E, an essential protein of Gram-negative bacteria that is involved in the processing of ribosomal RNAs (rRNAs), the maturation of tRNAs (tRNAs), and the ncRNA and mRNA decay (Li and Deutscher, 2002; Arraiano *et al.*, 2010). RNase E cleaves single stranded RNAs (ssRNAs) and it has a preference for A/U-rich sequences and 5' end monophosphorylated RNAs, although it can sometimes bypass this last requisite (Mackie, 1998). Paralogous and orthologous RNase E proteins can be found in both Gram-negative and Gram-positive bacteria. On the one hand, RNase G, a paralogous RNase E, can target a limited number of mRNAs, many of which are related to sugar metabolism (Ow *et al.*, 2003). On the other hand, RNase Y, which is the orthologous RNase E in Gram-positive bacteria, is also involved in mRNA decay (Shahbabian *et al.*, 2009; Lehnik-Habrink *et al.*, 2011).

RNase III is another well-studied endoRNase that is present in both Gram-negative and Gram-positive bacteria. In contrast to RNase E, RNase III targets double stranded RNAs (dsRNAs) using Mg^{2+} as a cofactor. It can attack dsRNAs that are formed either by intramolecular interactions, e.g. hairpin loops, or intermolecular interactions, e.g. imperfect or perfect RNA duplexes formed by ncRNA-mRNA or asRNA-mRNA pairing, respectively. RNase III is involved in mRNA decay as well as in the maturation of rRNAs (Wang and Bechhofer, 1997; Stead *et al.*, 2011). In addition, it has been shown that RNase III is the main endoRNase processing antisense RNAs (Lasa *et al.*, 2011; Lioliou *et al.*, 2012; Lasa *et al.*, 2012; Lybecker *et al.*, 2014). Lasa *et al.* discovered that RNase III was involved in the

processing of hundreds of overlapping regions in *S. aureus* that were produced by pervasive antisense transcription (Lasa *et al.*, 2011; Lasa *et al.*, 2012). Later on, a study, that precipitated antisense regions using antibodies against dsRNAs, confirmed that a genome-wide RNase III-mediated antisense RNA processing was also present in *E. coli* (Lybecker *et al.*, 2014).

An RNase III paralog has been identified in low GC-content Gram-positive bacteria. Due to its small size, it has been named mini-RNase III. In *B. subtilis*, mini-RNase III is involved in the maturation of the 23s rRNA (Redko *et al.*, 2008).

Regarding exoRNases, the polynucleotide phosphorylase (PNPase) is a 3'-5' exoRNase that requires Mg^{2+} to be functional. Interestingly, RNA hairpins inhibit PNPase activity, therefore, it requires the action of RNA helicases to continue with RNA processing (Luttinger *et al.*, 1996; Spickler and Mackie, 2000). Conversely, RNase R, a PNPase homolog present in many bacteria, can degrade RNA hairpins thanks to its intrinsic helicase activity. Nevertheless, it requires at least 10-12 nucleotide (nt) single stranded 3' overhang to process such secondary structures. RNase II is another 3'-5' Gram-negative exoRNase, which releases nucleotide monophosphates by a hydrolytic mechanism (Cheng and Deutscher, 2005; Awano *et al.*, 2010). Besides this, RNase J1 and RNase J2, only present in Gram-positive bacteria, are bifunctional ribonucleases with both ssRNA endoRNase and 5'-3' exoRNase activities. RNase J1 is a more

efficient exoRNase than J2, however it cannot degrade RNAs with 5' triphosphate or hairpin ends (Even *et al.*, 2005; Newman *et al.*, 2011).

Some RNases can be part of a large multiprotein complex, called the RNA degradosome that comprises RNA chaperones, RNA helicases and other accessory proteins. Specifically, in Gram-negative bacteria, the degradosome is formed by RNase E, PNPase, the helicase RhlB and the glycolytic enzyme enolase, while in Gram-positive bacteria, it consists of RNase Y, RNases J1/J2, PNPase, CshA helicase and the glycolytic enzymes phosphofructokinase and enolase. In *E. coli,* the C-terminal end of RNase E is the organizing scaffold of the degradosome (Miczak *et al.*, 1996; Callaghan *et al.*, 2004; Commichau *et al.*, 2009; Lehnik-Habrink *et al.*, 2010; Roux *et al.*, 2011). There are other minor components that bind to the degradosome and have the ability to modulate its activity. Therefore, the degradosome composition can change according to the accessory proteins that are loaded in relation to growth or stress conditions. For example, in *E. coli*, RhlB can be replaced by the CsdA helicase generating a "cold-shock" degradosome (Prud'homme-Généreux *et al.*, 2004).

Enzymes catalyzing RNA modifications

RNA molecules can be modified after transcription by the action of specific enzymes, which add different chemical groups to the RNAs. This can produce changes in RNA structure, base pairing and protein recognition, altering the original RNA function. Therefore, nucleotide modifications after

transcription add a new layer of regulation. RNA modifications are found in all three domains of life. This field, known as epitranscriptomics, has been widely studied in eukaryotes where modifications in mRNA molecules affect important cell processes such as stem cell differentiation (Wang *et al.*, 2014). In bacteria, since only few modifications have been described in tRNAs and rRNAs, little is known about the enzymes catalyzing the diverse modifications or the consequences of RNA modifications (Björk and Hagervall, 2005; Marbaniang and Vogel, 2016). However, a pioneer study carried out by Deng *et a*l., recently identified the N^6-methyladenosine (m^6A) transcriptome in several bacterial species and observed that this modification was present in many mRNAs of both Gram-negative and Gram-positive bacteria, indicating that mRNA modifications could be also extended in prokaryotes (Deng *et al.*, 2015).

The m^6A is the most common modification in eukaryotes and is enriched in 3'UTRs around the stop codon (Wang *et al.*, 2014). However, Deng *et al.* observed that in *E. coli* and *P. aeruginosa* m^6A modifications were more abundant inside CDSs with a consensus sequence of GCCAG. This also differs from the m^6A modification present in bacterial tRNAs and rRNAs suggesting different set of enzymes modifying the diverse RNA molecules (Deng *et al.*, 2015). The mRNAs carrying the m^6A modification were involved in different bacterial processes such as aerobic respiration and amino acid metabolism. Moreover, in *P. aeruginosa,* m^6A was temperature dependent, as this modification was almost abolished at high temperatures (45ºC) (Deng *et al.*, 2015).

In addition to RNA methylation, recent reports have identified that bacterial mRNAs can be capped at the 5' end like eukaryotic RNAs. Coenzyme A and nicotinamide adenine dinucleotide (NAD) were found covalently linked to the 5' end of RNAs (Kowtoniuk *et al.*, 2009; Chen *et al.*, 2009). In *E. coli*, ncRNAs and mRNAs were NAD-capped while no tRNAs and rRNAs carried this modification. It was proposed that NAD-capping stabilizes RNA molecules and protects them from RNase E degradation (Cahová *et al.*, 2015) (Figure 7).

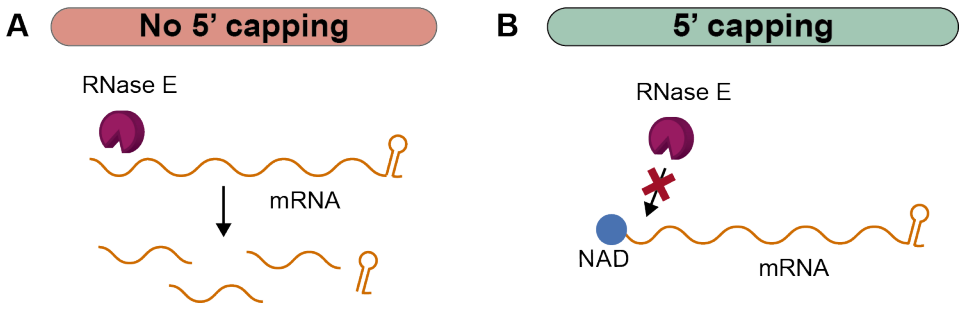


Figure 7. 5' capping of mRNAs. Some RNAs susceptible to RNase E degradation **(A)** can be modified at the 5' end with NAD molecules, protecting them from this RNase **(B)**.

Epitranscriptomics is a new field that will change our understanding about post-transcriptional regulation in bacteria. Further studies will be required to determine other existing modifications, the consequences of these modifications and the enzymes responsible for them (Marbaniang and Vogel, 2016).

Regulatory elements contained in mRNA molecules

A prototypical mRNA is comprised of a CDS flanked by two untranslated regions, the 5'UTR and the 3'UTR, respectively. Many regulatory elements have been found in 5'UTRs including riboswitches and thermosensors. In the last years, 3'UTRs have emerged as non-coding regions that also include diverse regulatory elements. As a result, many mRNAs are considered to possess a dual function: 1) they produce a specific protein and 2) they control the expression of their own protein or, in some cases, the expression of proteins encoded in other mRNAs.

5'UTRs with regulatory capacities

5'UTRs contain diverse regulatory elements that can affect transcription termination, RNA stability and translation of the mRNA in which they are contained. The most common elements are RNA thermosensors and riboswitches (Gripenland *et al.*, 2010).

RNA thermosensors or RNA thermometers (RNATs) are hairpin structures that are modified by temperature fluctuations affecting translation of the downstream genes (Figure 8A). An RNAT shifts between two conformations depending on temperature. It alternates from a closed one that sequesters the RBS impeding ribosome binding, to an open conformation that favors translation. This allows bacteria to quickly respond to temperature changes, an important skill for bacteria living in different niches. Three principal classes of genes are subjected to this type of post-transcriptional regulation: cold shock, heat shock and

virulence genes (Kortmann and Narberhaus, 2012; Loh *et al.*, 2018). In *Listeria monocytogenes*, the master regulator of virulence PrfA contains a 127 nt RNAT. While *L. monocytogenes* is living in the environment at temperatures lower than 30ºC, the *prfA* 5'UTR forms a secondary structure that inhibits PrfA translation. However, when *L. monocytogenes* enters into the human body through contaminated food, the temperature rises to 37ºC melting the RNAT structure and allowing PrfA translation. Then, PrfA activates several virulence factors required for intracellular replication and cell to cell dissemination (Johansson *et al.*, 2002; Freitag *et al.*, 2009). It is noteworthy that due to their simplicity, RNATs are widely distributed among pathogenic bacteria.

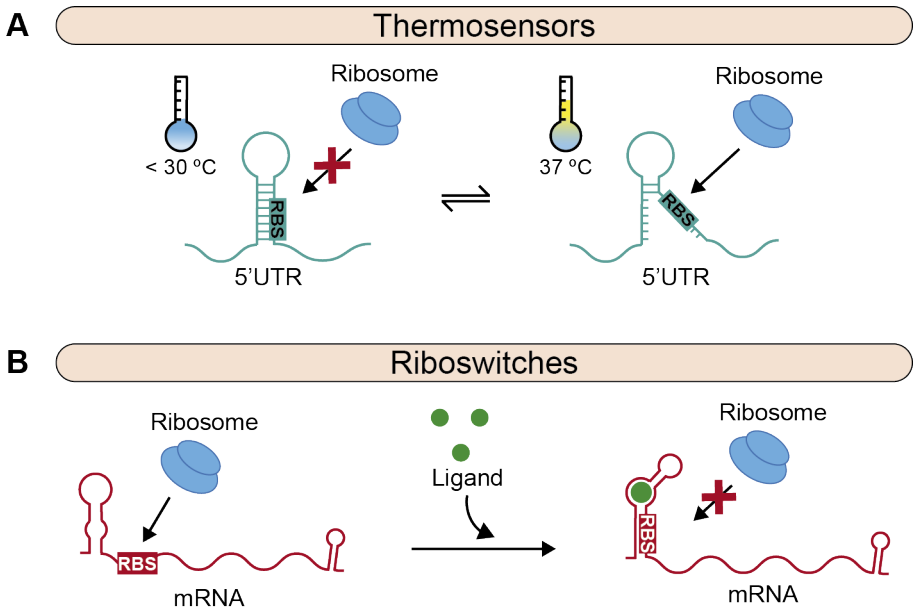


Figure 8. RNA thermosensors and riboswitches. Both RNA elements can shift between two RNA conformations in relation to temperature shifts **(A)** or upon binding of a specific ligand **(B)**. These conformational changes control gene expression.

Riboswitches are RNA elements that sense the concentration of different metabolites. A riboswitch binds to a specific molecule leading to a conformational change of its secondary structure (Figure 8B). This change in conformation is usually defined as an ON/OFF switch. The most common architecture of a riboswitch is composed of two separate domains, the aptamer domain that is used for ligand recognition and the expression platform, which carries out the regulatory response upon metabolite binding. Since these two domains share a mutual common region, the changes in the aptamer domain can be transmitted to the expression platform. Most of the ligands identified so far are molecules of different nature such as S-adenosylmethionine (SAM), nucleotides, amino acids, sugars, ions and tRNAs. Several mechanisms exist to ensure high selectivity of the ligand including shape, stacking interactions, hydrogen bonding and positive and negative interactions. For instance, purine riboswitches discriminate purines from pyrimidines using a single nucleotide that base pairs with adenine or guanine only (Peselis and Serganov, 2014; Sherwood and Henkin, 2016).

Depending on the riboswitch type, the conformational change upon ligand binding can produce different consequences, some affect transcription while others translation. Riboswitches can attenuate transcription by generating an intrinsic transcriptional terminator or exposing a Rho-binding site that leads to premature transcriptional termination. For example, T-box riboswitches bind to uncharged tRNAs to prevent transcription from early termination (Mellin and Cossart, 2015; Sherwood and Henkin, 2016).

In other riboswitches, once the ligand is bound, the riboswitch structure changes to liberate or occlude the RBS, promoting or inhibiting translation initiation, respectively (Sherwood and Henkin, 2016). In *E. coli*, the *btuB* mRNA translation is regulated by a coenzyme B_{12} riboswitch. Coenzyme B_{12} binding produces a structure rearrangement in the mRNA that leads to inhibition of translation, reducing the levels of the cobalamin transport protein BtuB (Nahvi *et al.*, 2002).

The expression of the same gene can be controlled by more than one riboswitch, for example, two located in tandem in the 5'UTR of a specific gene. These can be two copies of the same riboswitch, two riboswitches that sense the same ligand but act at different levels (transcription/translation) or two riboswitches that sense different ligands (Roßmanith and Narberhaus, 2017). Alternatively, two riboswitches can be found in opposite DNA strands creating antisense transcription to control the same gene depending on the concentration of two different metabolites. In *C. acetobutylicum,* the *ubiGmccBA* operon encodes a set of enzymes that convert SAM into cysteine. On the opposite DNA strand of the operon, a non-coding antisense RNA is controlled by a S-box riboswitch, which binds SAM. When methionine is low, the asRNA is transcribed and base pairs with the operon mRNA, inhibiting its translation. With high levels of methionine, the asRNA is early terminated. The *ubiGmccBA* operon is regulated by a T-box riboswitch. When high cysteine levels are found inside the cell, the charged $tRNA^{Cys}$ binds to the riboswitch causing transcriptional termination. In contrast, with low levels

of cysteine, the uncharged $tRNA^{Cys}$ binds to the riboswitch allowing transcription of the operon. As a result, only when there are low levels of cysteine and high levels of methionine the *ubiGmccBA* genes can be expressed (André *et al.*, 2008).

In addition to the expression control of the genes located downstream of the riboswitch, these regulatory elements can generate ncRNAs that can act in *trans*. In *L. monocytogenes*, high levels of SAM entail a conformational change in a S-box riboswitch located in the *lmo2419* gene that generates early transcription termination. The small RNA molecule generated (SreA) acts as an *trans*-acting ncRNA that binds to the 5'UTR of *prfA* mRNA, inhibiting its translation (Loh *et al.*, 2009).

A variation of thermosensors and riboswitches are pH sensors. In the 5'UTR of the *E. coli alx* gene is present an RNA element that responds to pH changes. Under neutral pH, the *alx* 5'UTR adopts a structure that hinders the RBS from ribosomes, inhibiting translation. Under alkaline conditions during transcription, the RNA polymerase (RNAP) pauses for a longer time at two different sites in the 5'UTR. This RNAP pausing allows an structural reorganization in the 5'UTR that liberates the RBS permitting translation initiation (Nechooshtan *et al.*, 2009).

In addition, RNA elements with self-cleaving capacities can also be found in 5'UTRs (Walter and Engelke, 2002). The *B. subtilis glmS* ribozyme is located in the 5'UTR of the *glmS* mRNA that encodes for the glucosamine-6-phosphate synthetase (GlmS), an enzyme that converts fructose-6-phosphate and glutamine into glutamate and glucosamine-6-phosphate

(GlcN6P). When GlcN6P accumulates, it binds to the *glmS* ribozyme inducing a self-cleavage at the 5'UTR that results in a processed mRNA with higher RNase J1 susceptibility, leading to mRNA degradation and a decrease of GlmS expression (Collins *et al.*, 2007).

Besides the particular RNA structures described so far, 5'UTRs possess specific sequences and structures that are recognized by ncRNAs, RNases and other RNA-binding proteins. This is the case of the 5'UTR of *cspA* mRNA in S. *aureus* where the same secondary structure is recognized by RNase III and the CspA chaperone (Figure 9).

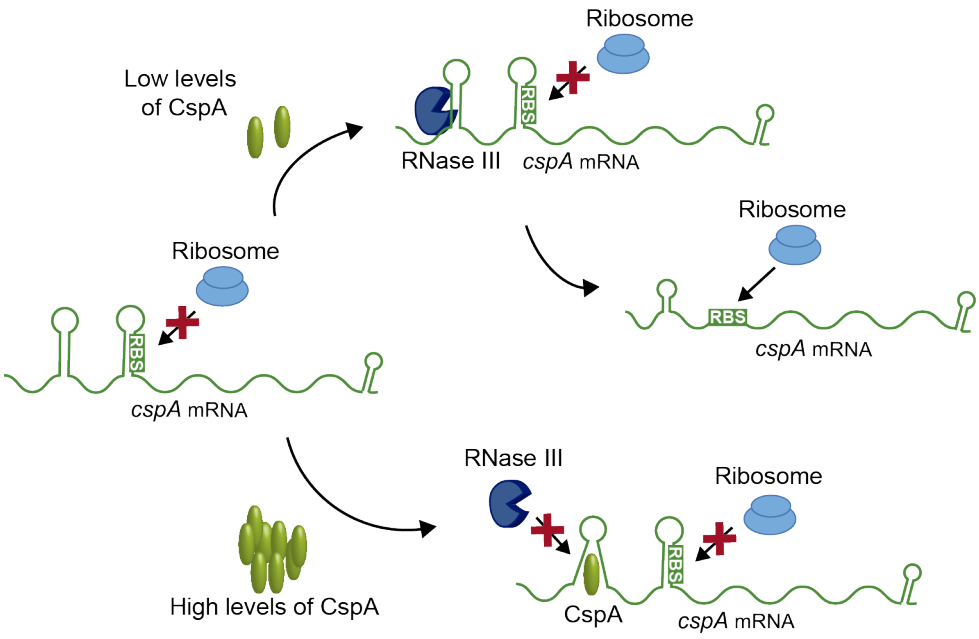


Figure 9. 5'UTRs can be targeted by RNases and RBPs. *cspA* mRNA possesses regulatory elements in its 5'UTR recognized by RNase III and its own protein, allowing the autoregulatory control of CspA expression. The 5'UTR of the *cspA* mRNA forms a hairpin structure, which is cleaved by RNase III to enhance CspA translation when CspA levels are low. If the CspA concentration rises, CspA is able to interact with this hairpin structure and unfold it. As a consequence, the *cspA* mRNA cannot be processed by RNase III and CspA translation is decreased. Adapted from (Menendez-Gil *et al.*, 2020b).

The *cspA* mRNA has a long 5'UTR with two stem loops, one of them occluding the RBS from the ribosomes. RNase III cleaves into the first stem loop generating a shorter mRNA that has the RBS accessible to the ribosomes. Therefore, the short *cspA* mRNA is more efficiently translated than the unprocessed transcript (Lioliou *et al.*, 2012). When high levels of CspA are reached in the cell, CspA binds to the first hairpin structure of its own 5'UTR impeding RNase III processing. Consequently, CspA auto-regulates its own expression by inhibiting translation (Caballero *et al.*, 2018).

3'UTRs: a new layer of post-transcriptional regulation in bacteria

In eukaryotes, 3'UTRs have been shown as key components of post-transcriptional regulatory mechanisms. Eukaryotic 3'UTRs possess regulatory motifs that are recognized by microRNAs (miRNAs) and RNA binding proteins affecting mRNA stability, localization and translation (Mazumder *et al.*, 2003; Wilkie *et al.*, 2003; Mayr, 2017). For instance, AU-rich (ARE) and GU-rich (GRE) elements, located in certain 3'UTRs, are recognized by proteins that favor mRNA degradation. The length of the polyA tail as well as the localization in which the mRNA is polyadenylated (nucleus or cytoplasm) can affect mRNA stability. In other cases, there are proteins that promote translation. For example, the PABP protein binds to the 3'UTR to then interact with the translation factor eiF4G, which is associated to the 5'UTR. This generates a circularization of the mRNA,

favoring translation efficiency due to the recycling of the ribosomes (Wilkie *et al.*, 2003; Halees *et al.*, 2008; Vlasova and Bohjanen, 2008).

The 3'UTR length varies according to the protein encoded in the mRNA. mRNAs encoding housekeeping proteins carry 3'UTRs shorter that mRNAs expressing regulatory proteins. Moreover, some genes can produce mRNAs with alternative 3'UTRs depending on the cell tissues where they are expressed. Deregulation or shortening 3'UTRs is associated with diverse diseases such as cancer or metabolic disorders (Khabar, 2010; Mayr, 2017).

It is noteworthy that the 3'UTR length has increased in relation to the complexity of the organism, being humans the organisms with longer 3'UTRs (> 500 nt), whereas 5'UTR length has remained constant during evolution. It is thought that 3'UTRs have contributed to organism complexity in eukaryotes (Mazumder *et al.*, 2003).

In contrast, bacterial 3'UTRs have been historically disregarded because studies were mainly focused on the identification of ncRNAs and regulatory elements in 5'UTRs, and the RNA-seq techniques did not initially mapped the 3' ends (Wagner and Romby, 2015). Additionally, it was believed that 3'UTRs were only needed for a correct transcription termination and/or protecting transcripts from exoribonucleases. However, several new regulatory elements have been described so far in 3'UTRs establishing them as a novel regulatory layer to be considered in bacteria. It is expected that this field will grow exponentially in the next years. In the

following sections, relevant aspects of post-transcriptional regulation mediated by 3'UTRs in bacteria are described.

Transcriptional termination determines the 3'UTR length

A prototypical 3'UTR comprises the mRNA region from the stop codon to the transcriptional termination site. Therefore, the transcriptional termination signal is usually included in this region. In bacteria, transcriptional termination can occur either by the presence of an intrinsic transcriptional terminator (TT), which consists of a hairpin structure followed by a U-tract (Rho-independent termination), or by the presence of a Rho utilization (*rut)* site that recruits the Rho protein (Rho-dependent termination). Rho binding to C-rich regions causes the activation of its ATPase activity and provides the energy to translocate through the mRNA and interact with the RNAP. This leads to the release of the transcript thanks to Rho helicase activity (Peters *et al.*, 2011).

Transcriptional termination systems are not always effective in dissociating the RNAP from the transcript. As a consequence, sometimes transcription continues far away from the transcriptional terminator signal. This phenomenon is defined as termination read-through and causes alternative 3'UTRs (Lasa *et al.*, 2011; Ruiz de los Mozos *et al.*, 2013).

The average length of an intrinsic transcriptional terminator was predicted to be around 30 nucleotides in *S. aureus*. Therefore, a 3'UTR of 40-50 nt long should be sufficient to allocate a functional transcriptional terminator (Ruiz de los Mozos *et al.*, 2013). However, with the expansion of high-

throughput RNA sequencing technologies, the real 3'UTR lengths could be determined in different bacteria revealing that 3'UTRs were longer than previously anticipated (Broeke-Smits *et al.*, 2010; Ruiz de los Mozos *et al.*, 2013; Dar *et al.*, 2016; Dar and Sorek, 2018). In *S. aureus*, more than 30% of the mapped 3'UTRs were longer than 100 nt. This was a strong evidence indicating that 3'UTRs could allocate additional regulatory elements (Ruiz de los Mozos *et al.*, 2013).

3'UTRs involved in the expression of their own mRNAs

The first example showing that a bacterial 3'UTR could play a role modulating the expression of the protein encoded in the same mRNA was described by Balaban and Novick in 1995 (Balaban and Novick, 1995). The RNAIII of *S. aureus* is the paradigm of a regulatory transcript, which can modulate the expression of several mRNAs by base paring (Bronesky *et al.*, 2016). In addition, this RNA encodes the δ–hemolysin. Balaban and Novick showed that translation of δ-hemolysin occurred one hour after transcription of RNAIII. However, deletion of the 3'UTR significantly diminished the delay, suggesting that the 3'UTR blocked δ–hemolysin translation (Balaban and Novick, 1995).

Several years passed until novel 3'UTR functions started to be characterized. Diverse 3'UTR-mediated mechanisms have been described so far, including 5'UTR-3'UTR interaction and 3'UTR targeting by RNases ncRNAs and RBPs (Figure 10).

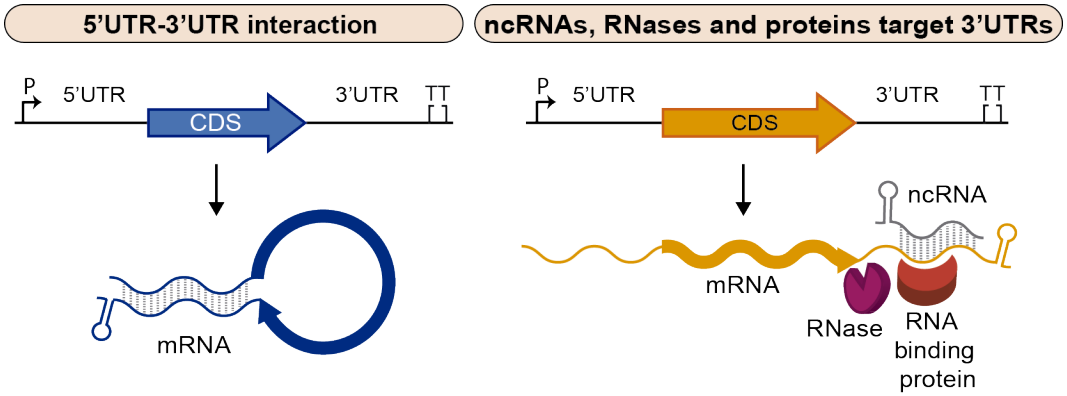


Figure 10. Regulatory mechanisms mediated by 3'UTRs. 3'UTRs can interact with the 5'UTRs from the same mRNA and they can also be targeted by ncRNAs, RNases, and other RBPs. P, promoter; TT, transcriptional terminator.

Our group unveiled that the 3'UTR of *icaR* mRNA in *S. aureus* interacts with the RBS of its own mRNA through a UCCCC sequence that acted as an anti-RBS motif. This 5'UTR-3'UTR interaction inhibits translation initiation of IcaR protein and produces a double stranded substrate that is cleaved by RNase III leading to mRNA degradation. IcaR is the repressor of the *ica* operon that encodes for the main exopolysaccharidic compound (PIA-PNAG) in *S. aureus* biofilms. Therefore, the *icaR* 5'UTR-3'UTR interaction favors biofilm formation (Ruiz de los Mozos *et al.*, 2013). Later, another example of 5'UTR-3'UTR interaction was described for the *hbs* mRNA in *B. subtilis.* A 15 nt long interaction protects *hbs* mRNA from RNase Y cleavage at the 5'UTR (Braun *et al.*, 2017). These mechanisms resembled the mRNA circularization occurring in eukaryotes to control protein expression (Mazumder *et al.*, 2003).

Other bacterial 3'UTRs are often used as direct entry points for ribonucleases to initiate mRNA degradation (Figure 10). This is the case

for the *hmsT* mRNA, which encodes a protein that modulates c-di-GMP synthesis, required for allosteric activation of polysaccharide production in *Yersinia pestis* biofilm formation. PNPase-dependent cleavage of the *hmsT* 3'UTR affects the turnover of the mRNA and, ultimately, HsmT expression (Zhu *et al.*, 2016; Zhao *et al.*, 2018). Additionally, several AU-rich 3'UTRs of *Y. pestis* are targeted by PNPase when the transcriptional termination is Rho-dependent (Zhao *et al.*, 2018). In *Salmonella*, HilD, one of the main transcriptional activator factors of the pathogenicity island 1 (SPI-1) is processed through its 3'UTR by RNase E and PNPase (López-Garrido *et al.*, 2014). Similarly, in *Corynebacterium glutamicum* the *aceA* mRNA, encoding the isocitrate lyase protein that is part of the glyoxylate cycle (Gerstmeir *et al.*, 2003), is processed by RNase E/G at the 3'UTR (Maeda and Wachi, 2012).

Bacterial 3'UTRs can also be targeted by other factors such as ncRNAs and RBPs to modulate mRNA expression (Figure 10). For example, the expression of the *E. coli* aconitase (*acnB*) mRNA is autoregulated by its own protein. AcnB is an enzyme of the TCA (tricarboxylic acid) cycle that uses iron as a cofactor, but it becomes an autoregulatory RBP when intracellular iron becomes scarce (apo-AcnB). The apo-AcnB binds to its own mRNA at a stem-loop located at the 3'UTR. This binding obstructs an RNase E cleavage mediated by the ncRNA RyhB together with Hfq (Figure 11). Therefore, during iron starvation apo-AcnB induces its own expression by avoiding *acnB* mRNA processing and maintaining iron homeostasis. This regulation creates an equilibrium dependent on the iron

pool between both states of AcnB: the enzymatic activity and the RNA-binding capacity (Benjamin and Massé, 2014).

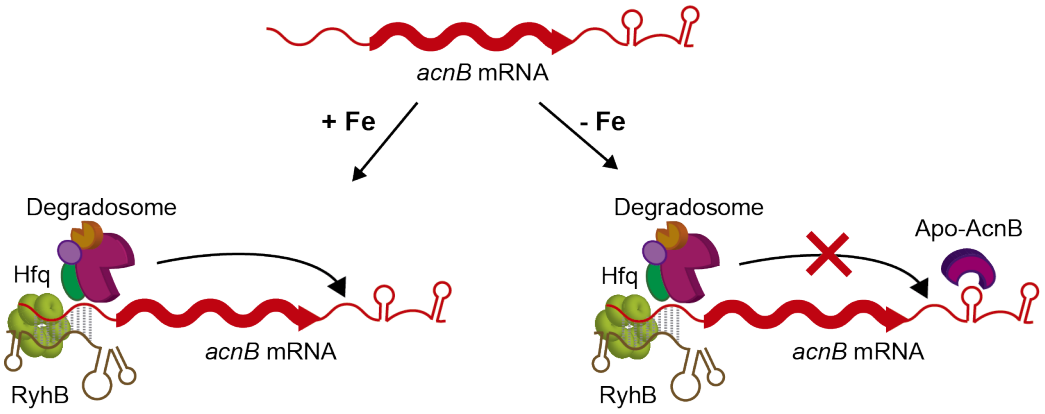


Figure 11. The *acnB* 3'UTR possesses a secondary structure recognized by its own protein. The Apo-AcnB can bind to the *acnB* 3'UTR to prevent the processing of the degradosome. The latter is recruited when RhyB ncRNA and Hfq interact with the *acnB* 5'UTR. Adapted from (Benjamin and Massé, 2014).

A recent report showed that the RNA chaperone ProQ targets 3'UTRs in *Salmonella* and *E. coli*. Deletion of ProQ revealed that many of the mRNAs targeted by this protein had a reduced stability in the mutant. It seems that ProQ recognizes secondary structures at the 3'UTR of several mRNAs stabilizing them. This is the case of *cspE* mRNA in which ProQ binding to the 3'UTR protects the mRNA from RNase II cleavage (Holmqvist *et al.*, 2018).

3'UTRs can also be targeted by *trans*-acting ncRNAs as occurred with microRNAs in eukaryotes (El-Mouali *et al.*, 2018; Bronesky *et al.*, 2019). The ncRNA RsaI of *S. aureus* binds to the 3'UTR of *icaR* mRNA. As mentioned before, the 3'UTR of *icaR* mRNA carries an anti-RBS motif (UCCCC sequence) that binds to the RBS of its own mRNA (Ruiz de los

Mozos *et al.*, 2013). The binding of RsaI occurs in a different location of the 3'UTR so it is thought that this ncRNA might help stabilizing the circularization of the mRNA, which inhibits IcaR translation (Bronesky *et al.*, 2019) Interestingly, RsaI expression is repressed by the catabolite control protein A (CcpA) when glucose is available in the medium. Therefore, if this carbon source is consumed, RsaI is activated promoting *ica* expression and biofilm formation (Bronesky *et al.*, 2019) (Figure 12).

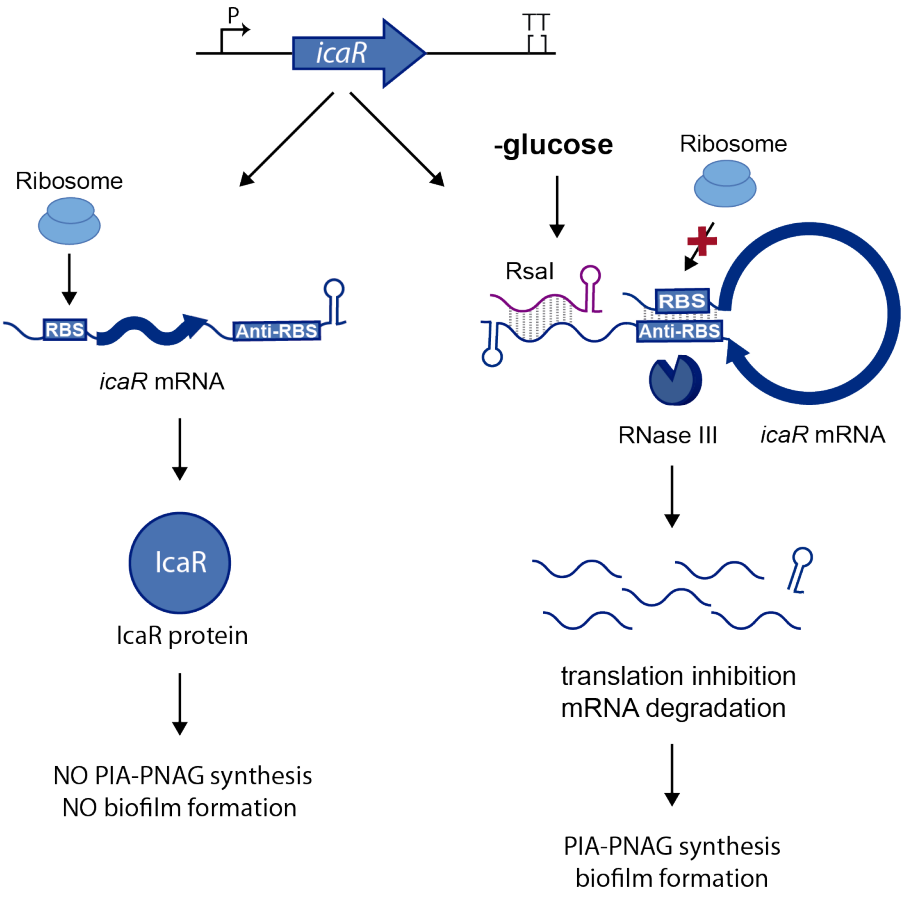


Figure 12. Regulatory elements of *icaR* 3'UTR. IcaR is the transcriptional repressor of the *icaADBC* operon, encoding the enzymes that synthesize PIA-PNAG, the main exopolysaccharide of *S. aureus* biofilms. The *icaR* 3'UTR carries an anti-RBS motif that interacts with the RBS of its own mRNA causing translation inhibition and RNase III-dependent degradation. The ncRNA RsaI binds to the 3'UTR of *icaR* in between the anti-RBS motif and the transcriptional terminator. Although the molecular mechanism of this ncRNA-mRNA paring is unknown, it has been shown that RsaI promotes PIA-PNAG synthesis. P, promoter; TT, transcriptional terminator.

In *Salmonella*, Spot42, a trans-acting ncRNA repressed by the catabolite repression protein (CRP) and cAMP (cyclic adenosine monophosphate), targets the long 3'UTR of *hilD* mRNA (El-Mouali *et al.*, 2018). The *hilD* 3'UTR has been shown essential to modulate HilD expression and HilD-dependent virulence genes (López-Garrido *et al.*, 2014). El Mouali *et al.*, showed that Spot42 binds to the 3'UTR of *hilD* mRNA in an Hfq-dependent manner and this interaction stabilizes and/or activates HilD expression (El-Mouali *et al.*, 2018) (Figure 13). These two examples illustrate how ncRNAs interactions can connect and control different pathways according to the availability of nutrient resources.

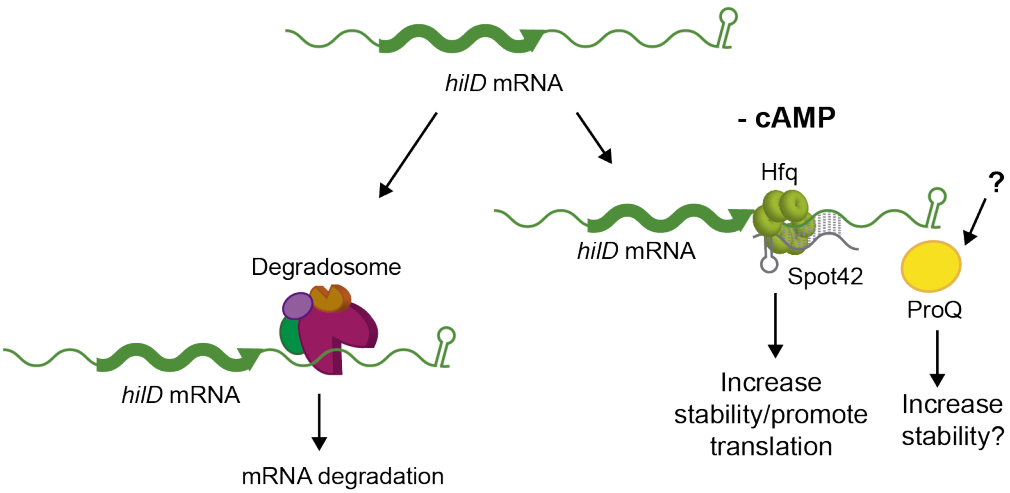


Figure 13. The regulatory 3'UTR of *hilD* mRNA. Spot42 together with Hfq binds to *hilD* 3'UTR promoting mRNA stability and/or translation. Additionally, *hilD* 3'UTR can be targeted by the degradosome and ProQ protein.

3'UTRs as reservoirs of *trans*-acting ncRNAs

3'UTRs have also emerged as reservoirs of *trans*-acting ncRNAs. 3'UTR-derived ncRNAs can be generated either by an internal promoter located inside or just downstream of the CDS (type I) or by processing of the

mRNA transcript at the 3'UTR (type II) (Figure 14). Therefore, type I ncRNAs carry a triphosphate at the 5' ends while a monophosphate is found at the 5' ends of type II ncRNAs (Kawano *et al.*, 2005; Chao *et al.*, 2012; Miyakoshi, Chao, and Vogel, 2015a).

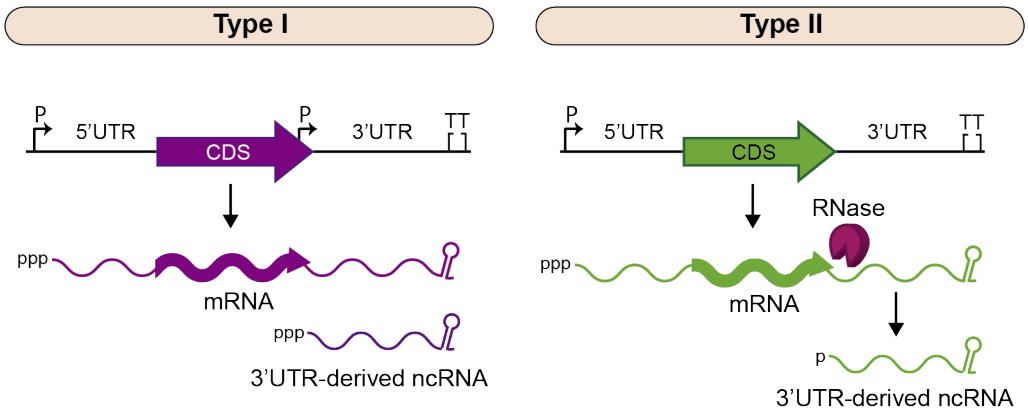


Figure 14. 3'UTR-derived ncRNAs. The type I 3'UTR-derived ncRNAs are generated by an internal promoter (P), which is inside or just downstream of the CDS. The type II 3'UTR-derived ncRNAs result from the processing of the mRNA by an specific RNase. Adapted from (Miyakoshi, Chao, and Vogel, 2015a).

Three type I ncRNAs have been characterized so far, DapZ, MicL and MicX. Type II ncRNAs are more abundant and have been described in several bacterial species. Table 1 summarizes the 3'UTR-derived ncRNAs characterized so far in bacteria. The genes targeted by these ncRNAs and the physiological processes controlled by them are also indicated. These examples suggest that 3'UTR-derived ncRNAs control a wide variety of processes and are widely distributed among bacteria. Note that the physiological process in which these ncRNAs participate is related to the function of the protein encoded in the mRNA from where they are originated.

Table 1. List of 3'UTR-derived ncRNAs and their functions

mRNA	ncRNA	Type	Targets	Relevant characteristics	Reference
Escherichia coli					
cutC	MicL	I	*lpp*	σ^E–dependent, involved in membrane stress	(Guo *et al.*, 2014)
sdhCDAB-sucABCD	ShdX	II	*ackA*	Coordinates the expression of the TCA cycle and the acetate metabolism	(De Mets *et al.*, 2019)
Salmonella					
dapB	DapZ	I	*oppA* *dppA*	HilD-dependent, represses two ABC transporters	(Chao *et al.*, 2012)
cpxP	CpxQ	II	*agp* *fimAICDHF* *nhaB* *skp-lpxD* *ydjN*	Hfq-dependent, targets extracytoplasmic proteins to alleviate inner membrane stress	(Chao and Vogel, 2016)
gltIJKL	SroC	II	GcvB	Sponge RNA, alleviates GcvB-repression of amino acid transport and metabolic genes	(Miyakoshi, Chao, and Vogel, 2015b)
Vibrio Cholerae					
vca0943	MicX	I	*vc0972*	Processed by RNase E in a Hfq-dependent manner, regulates outer membrane protein	(Davis and Waldor, 2007)
Rhodobacter sphaeroides					
RSP_0847	SorX	II	*potA*	Inhibits a polyamine transporter to counteract oxidative stress	(Peng *et al.*, 2016)
Streptomyces coelicolor					
sodF	s-SodF	II	*sodN*	Inhibits SodN, regulates superoxide dismutases expression in relation to nutrient availability	(Kim *et al.*, 2014)

***Staphylococcus aureus* as a bacterial model**

As indicated before, several novel regulatory mechanisms mediated by bacterial 3'UTRs have been described in the last decade. However, there are still many aspects that remain unknown. In this Thesis, aiming at extend our knowledge in the functionality of 3'UTRs in bacteria, we used as a model *Staphylococcus aureus*, one of the most relevant pathogens worldwide.

S. aureus is a facultative anaerobic Gram-positive bacterium that receives its name due to the production of a yellow pigment called staphyloxantin (Xue *et al.*, 2019). It is estimated that one third of the human population is a natural reservoir of *S. aureus*. It resides in the upper respiratory tracts and on the skin as a commensal. However, *S. aureus* is also an opportunistic pathogen that can cause a wide variety of diseases differing in severity, ranging from mild soft skin infections to life-threatening diseases such as meningitis, endocarditis and sepsis. *S. aureus* has a strong ability to develop resistance to antibiotics by horizontal transfer of resistance genes (Oliveira *et al.*, 2018; Horn *et al.*, 2018; Brockhurst *et al.*, 2019). In addition, it is a very versatile bacterium that can adapt and resist to a wide variety of conditions thanks to its capacity to form biofilms. This bacterial community is embedded by an extracellular matrix composed mainly of polysaccharides, proteins and extracellular DNA. The secreted matrix protects the bacterial community from antibiotics, the immune system, and extreme environmental conditions. On the one hand, this enables *S. aureus* to cause chronic infections by establishing biofilms on

tissues or medical-device surfaces such as catheters, artificial heart valves and joint prosthetics. On the other hand, it contributes to *S. aureus* transmission by increasing its survival on different niches (Archer *et al.*, 2011).

Moreover, *S. aureus* is able to produce a wide arsenal of exotoxins and virulence factors that are controlled by the quorum sensing system Agr (Novick, 2003; Queck *et al.*, 2008). The *agr* locus consists of two divergent transcriptional units originally named RNAII and RNAIII controlled by the P2 and P3 promoters, respectively (Figure 15).

RNAII transcript encodes a two-component system (TCS) composed of AgrA and ArgC and a cell density-sensing cassette formed by AgrB and AgrD. AgrD is a small peptide that is processed by AgrB, a membrane protease that secretes the processed AgrD into the medium. This processed version is called autoinducing peptide (AIP) because as cells start to replicate and accumulate, the amount of AIP increases inducing the autophosphorylation of the membrane kinase, AgrC. Then, AgrC transfers the phosphate to the response regulator AgrA to activate the expression of RNAIII (Novick, 2003; Queck *et al.*, 2008). RNAIII regulates several genes at the post-transcriptional level, promoting the expression of secreted virulence factors and inhibiting several adhesion proteins (Novick *et al.*, 1993; Bronesky *et al.*, 2016).

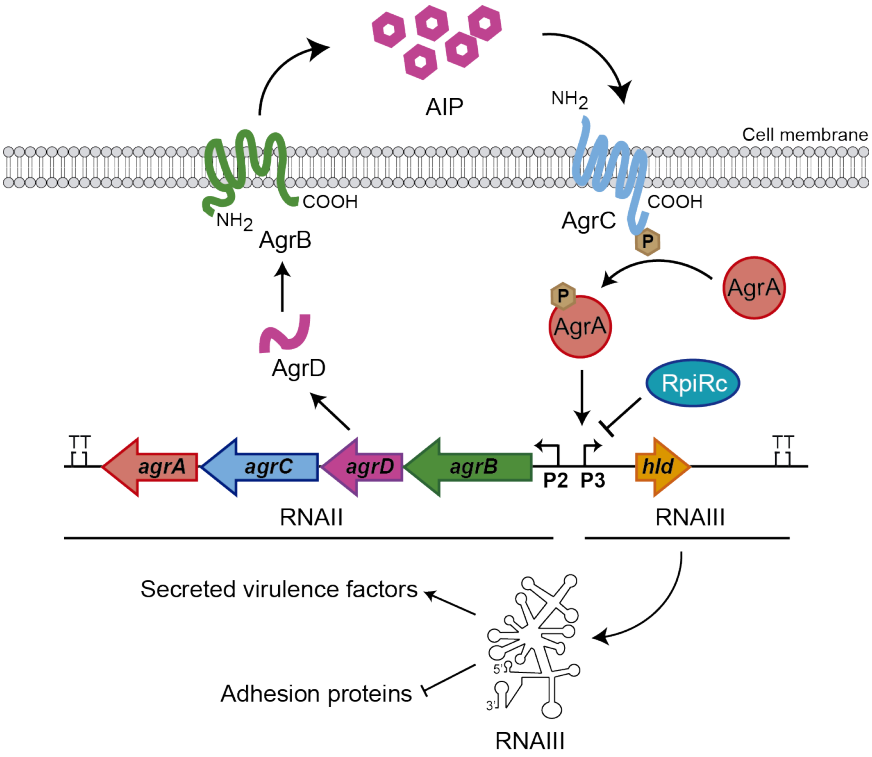


Figure 15. Agr quorum sensing system of *S. aureus*. It consists of two transcriptional units, RNAII and RNAIII, activated by promoters P2 and P3, respectively. RNAII encodes a cell density-sensing TCS that activates RNAIII. RNAIII is a dual-functional mRNA encoding the δ-hemolysin (*hld*), which promotes expression of secreted virulence factors and inhibits adhesion protein synthesis through mRNA-mRNA interactions. RpiRc is a transcriptional repressor that inhibits RNAIII depending on the metabolic status of the cell. Adapted from (Painter *et al.*, 2014).

Agr is tightly regulated by diverse environmental signals sensed by different regulatory systems (Bischoff *et al.*, 2001; Pragman *et al.*, 2007). For example, RpiRc (ribose phosphate regulator) is a transcriptional regulator involved in the control of the pentose phosphate pathway and the TCA cycle. At the same time, RpiRc inhibits the expression of RNAIII through repression of the P3 promoter. An *rpiR*c mutant produces higher levels of hemolysins and leukocidins enhancing virulence in a murine model (Zhu *et al.*, 2011; Balasubramanian *et al.*, 2016; Gaupp *et al.*,

2016). In this way, RpiRc is able to connect the cell metabolic status with an adequate expression of virulence genes.

Thanks to this wide palette of abilities, *S. aure*us is one of the most feared pathogens in the healthcare systems. Additionally, it is noteworthy that the proportion of infections caused by methicillin-resistant *S. aureus* strains (MRSA) is rising every year. The last weapon against MRSA infections is vancomycin but there are already some reports of vancomycin-resistant strains (Diep and Otto, 2008; Oliveira *et al.*, 2018; Horn *et al.*, 2018). For this reason, the World Health Organization (WHO) considers *S. aureus* as a high priority pathogen and strongly encourages the development of new antimicrobials against it (https://www.who.int).

Due to the obvious importance that *S. aureus* represents for the society worldwide, it has become a paradigm for the study of antimicrobial resistance, acute and chronic infections, bacterial clonal dispersion, evolution, etc. Since several key discoveries related to non-coding RNA regulation have been achieved in *S. aureus*, it is also considered an excellent model for studying post-transcriptional regulation. In this Thesis, we unveiled an evolutionary bias within 3'UTRs of *Staphylococcus* that resulted in sequence variations that created functional species-specific 3'UTRs; a phenomenon that seems to be widely distributed in bacteria.

OBJECTIVES

OBJECTIVES

The pioneering studies carried out so far in bacteria have shown the importance of 3'UTRs as a novel regulatory layer. These untranslated regions can contain diverse regulatory elements to post-transcriptionally control relevant physiological processes in bacteria. Previous work in our laboratory showed that the *icaR* 3'UTR of *S. aureus* presented sequence variability when compared to other staphylococcal species (Ruiz de los Mozos *et al.*, 2013). This preliminary observation opened several questions about the functionality and evolution of 3'UTRs in bacteria. For example, are 3'UTR sequences preserved within and between bacterial species? How often do conserved genes from closely related bacteria contain different 3'UTRs and how is 3'UTR variability originated? Do differences in 3'UTRs between orthologous genes have consequences in their expression at the protein level? In this Thesis, aiming to answer these questions, we established the following objectives:

1. Perform genome-wide comparative analyses to study the conservation and evolution of 3'UTRs among closely-related bacterial species.

2. If differences in 3'UTRs conservation are found, study how nucleotide variations in 3'UTRs may occur through evolution.

3. Analyze if 3'UTR sequence variations can create different expression patterns among mRNAs encoding orthologous proteins from closely-related staphylococcal species.

MATERIAL AND METHODS

MATERIAL AND METHODS

Strains, plasmids, oligonucleotides and growth conditions

Bacterial strains, plasmids and oligonucleotides used in this study are listed in Annex 1, Annex 2 and Annex 3, respectively. *Staphylococcus* strains were grown in Tryptic Soy Broth (Pronadisa) supplemented with 0.25% glucose (TSBg) or, when indicated, in Brain Heart Infusion (BHI) or chemically defined medium (Toledo-Arana *et al.*, 2005). *Escherichia coli* was grown in Luria-Bertani (LB) broth (Pronadisa). B2 (casein hydrolysate, 10 g l^{-1}; yeast extract, 25 g l^{-1}; NaCl, 25 g l^{-1}; K_2HPO_4, 1 g l^{-1}; glucose, 5 g l^{-1}; pH 7.5) and SuperBroth (tryptone, 30 g l^{-1}; yeast extract, 20 g l^{-1}; MOPS, 10 g l^{-1}; pH 7) media were used to prepare *S. aureus* and *E. coli* competent cells, respectively. For selective growth, media were supplemented with the appropriated antibiotics at the following concentrations: Ampicillin (Amp), 100 µg ml^{-1} for all plasmids transformed in *E. coli;* Erythromycin (Erm), 1,5 µg ml^{-1} or 10 µg ml^{-1} for pMAD or pCN plasmids, respectively, in *S. aureus*. When using the Pcad promoter, 1.5 µM of cadmium was added to activate the promoter.

Nucleotide conservation analysis

The conservation of UTRs among orthologous monocistronic mRNAs from phylogenetically-related bacterial species was determined by performing blastn comparisons using the Microbial Nucleotide BLAST tool (https://blast.ncbi.nlm.nih.gov/). Monocistronic mRNA boundaries were

manually annotated by visualizing the *S. aureus* transcriptomic maps from previous studies (Lasa *et al.*, 2011; Ruiz de los Mozos *et al.*, 2013; Koch *et al.*, 2014) in the Jbrowse application (Skinner *et al.*, 2009). For the batch comparison of *S. aureus* 3'UTRs, a file including the query sequences in FASTA format was used. Each sequence included the annotated 3'UTR and the last 200 nt of the CDS of a monocistronic gene (Figure 16). This restriction was applied to normalize the starting point for each of the UTRs from the blastn results, facilitating further analysis and plotting. We considered that 200 nt from the CDS were enough to properly identify their corresponding orthologous genes. Similarly, for batch comparison of *S. aureus* 5'UTRs, a FASTA file including the annotated 5'UTRs and the first 200 nt of the CDSs were used. In order to normalize the starting point of the CDS, we used reverse-complement query sequences for 5'UTR blastn comparisons (Figure 16). The UTR conservation length was registered as the last nucleotide from the query sequence showing a positive alignment with the region of the genome under comparison. For this purpose, the information contained in the blastn output TXT files was transformed into tables using a custom R script. Nucleotide alignments that started at values higher than 100, finished at values lower than 100, or were shorter than 80 nucleotides were not considered. The UTR conservation lengths of each *Staphylococcus* species were plotted against the reference UTR lengths of *S. aureus* (Figure 16). The square correlation coefficient (R^2) was calculated in R from the resulting fitted linear model.

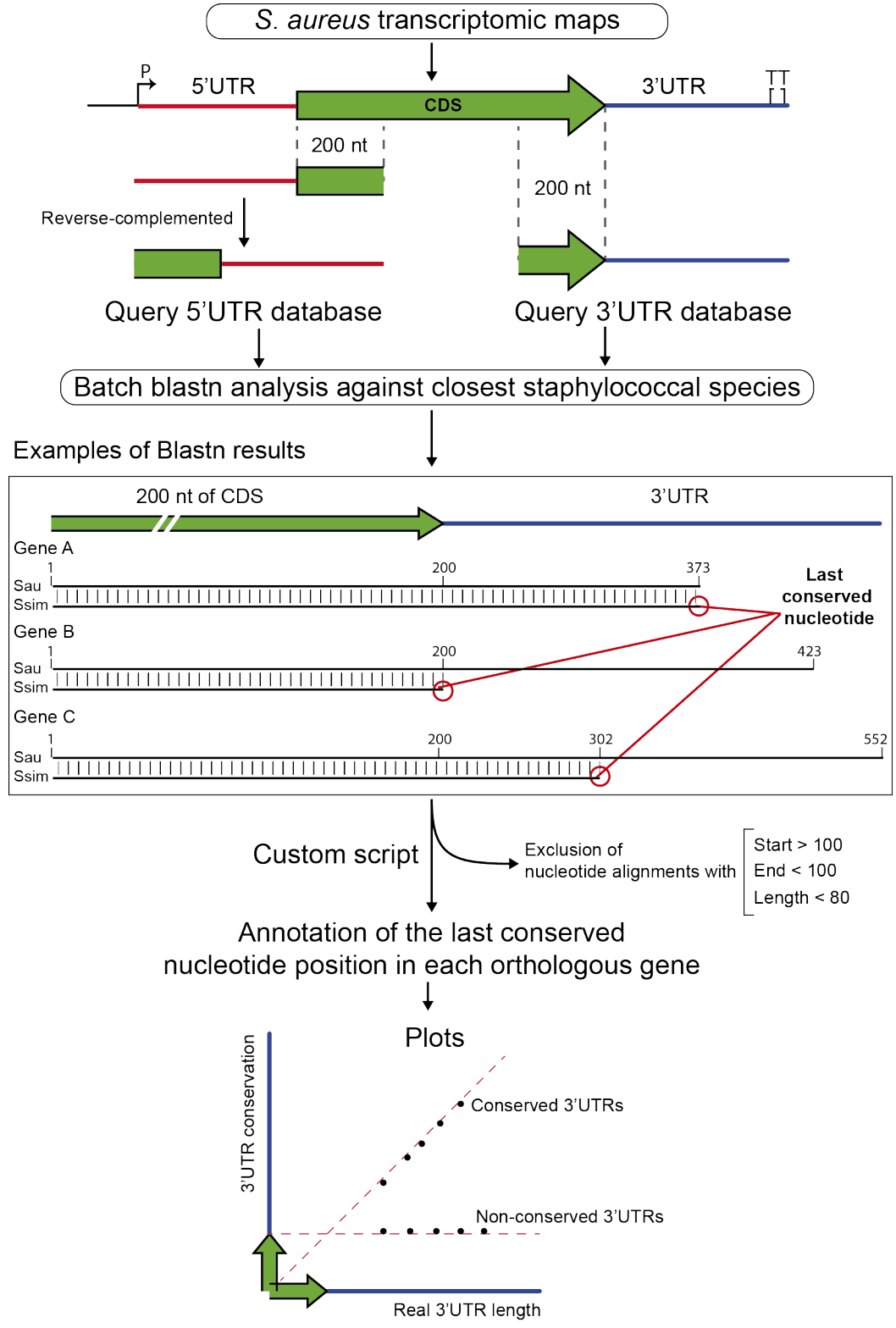


Figure 16. Schematic representation of the UTR conservation analyses performed among closely-related bacterial species. Query UTR databases were constructed based on available transcriptomic maps information and were used to perform batch blastn analyses to identify the last conserved nucleotides among orthologous genes. P, promoter; TT, transcriptional terminator; Sau, *S. aureus*; Ssim, *S. simiae*.

This analysis was also applied to compare the mapped 3'UTRs of *B. subtilis* (Dar *et al.*, 2016) and *E. coli* (Dar and Sorek, 2018) with phylogenetically-related species of the *Bacillus* genus and *Enterobacteriaceae* family, respectively. Note that in the case of *B. subtilis*, 3'UTRs longer than 150 nt were not annotated (Dar *et al.*, 2016).

RNA sequencing and data analysis

In order to perform the transcriptomic maps of *S. simiae* CCM 7213T, *S. capitis* SK14 and *S. epidermidis* RP62A, strains were grown in TSBg at 37ºC and 200 rpm until exponential phase was reached. Total RNAs were extracted as previously described (Toledo-Arana *et al.*, 2009). RNA sequencing and preliminary data analysis was carried out by the Stab Vida company. The RNA-seq reads from *S. simiae* CCM 7213T, *S. capitis* SK14 and *S. epidermidis* RP62A samples were aligned using the Rockhopper program (Tjaden, 2015) and the complete genome sequences from *S. simiae* NCTC13838 (NZ_LT906460.1), *S. capitis* AYP1020 (NZ_CP007601.1) and *S. epidermidis* RP62A (NC_002976.3) as references, respectively. The obtained read coverage files (.wig) from plus and minus strands were converted to BigWig (.bw) files with the wig2BigWig program. The generated .bw files and the gene annotation files, including the positions of transcriptional terminators predicted by TransTerm HP program (Kingsford *et al.*, 2007), were loaded into a public web server based on Jbrowse (Skinner *et al.*, 2009) (http://rnamaps.unavarra.es/). The 3'UTR lengths of the orthologous monocistronic mRNAs were manually annotated and plotted against the

lengths of previously described *S. aureus* 3'UTRs (Ruiz de los Mozos *et al.*, 2013).

Simultaneous mapping of 5' and 3' ends from RNA molecules

The mapping of 5' and 3' ends of mRNAs was performed using a modified version of the Rapid Amplification of cDNA Ends (mRACE) method (Britton *et al.*, 2007). Briefly, total RNA samples were treated with the Cap-Clip Acid Pyrophosphatase (Tebu-Bio) following the manufacturer's recommendations. After 1 hour of incubation, RNAs were extracted by phenol-chloroform and precipitated with sodium acetate and cold ethanol. Serial dilutions of Cap-Clip and non-Cap-Clip treated RNAs were ligated using the T4 RNA Ligase I (New England Biolabs) at 16ºC overnight (ON). RT-PCRs for all the ligated RNA dilutions were carried out using the SuperScript™ III One-Step RT-PCR System with the Platinum Taq DNA polymerase (Invitrogen) and the outward primers A and B (Annex 3). For the mapping of the 3'UTR of *rpiRc* from *S. epidermidis* and *S. capitis* primers RACE-3XFLAG-A and RACE-RpiRc-sau-B were used. PCR products were run in 2.5% agarose gels and bands of the expected size were purified and ligated into the pGEM-T Easy vector (Promega). The resulting reactions were used to transform *E. coli* XL1-Blue cells (Stratagene). Ten isolated white colonies were then analyzed by Sanger sequencing. Transcript boundaries were determined by blastn analysis and nucleotide frequencies representing the 5' and 3' ends, respectively, were registered as percentages (%).

Chromosomal mutagenesis

The mutants generated in this study (Annex 1) were obtained by a two-step homologous recombination that exchanges a specific chromosomic region by the mutant allele present in the pMAD plasmid (Arnaud *et al.*, 2004), as previously described (Valle *et al.*, 2003). The marker-less mutants were verified by PCR using the oligonucleotides E and F (Annex 3) and Sanger sequencing.

Plasmid construction

Most of the plasmids used in this study were engineered as previously described (Caballero *et al.*, 2018). Briefly, PCR fragments were amplified from chromosomic or plasmidic DNA with the DreamTaq DNA polymerase or Phusion High-Fidelity DNA Polymerase (Thermo Scientific), using the oligonucleotides listed in Annex 3. The PCR products were run and purified from agarose gels using the NucleoSpin® Gel and PCR Clean-up Macherey-Nagel kit, and ligated into the pJET 1.2 vector (Thermo Scientific). The resulting plasmids were used to transform *E. coli* XL1-Blue cells. These were then purified from ON cultures with the NucleoSpin® Plasmid Macherey-Nagel kit and verified by Sanger sequencing. When required, DNA fragments were excised using FastDigest restriction enzymes (Thermo Scientific) and ligated into the appropriate vector with the Rapid DNA ligation kit (Thermo Scientific). The final plasmids were introduced into *S. aureus* strains by electroporation, as previously described (Lee, 1995).

pMADs plasmids used for chromosomal mutations were constructed by amplifying flanking sequences (AB and CD) of the target regions using primers A/B and C/D (Annex 3). After cloning them into pJET they were digested and ligated into pMAD in a double-fragment ligation using BamHI, EcoRI and KpnI or NheI (Annex 2 and Annex 3).

The plasmids used for identification of regulatory 3'UTRs were generated by overlapping PCR to introduce the 3xFLAG tag sequence at the N-terminal of the protein. In particular, the tagged gene was generated by linking two partially overlapping PCR fragments that were amplified using oligonucleotides +1, 3XFLAG-izq, 3XFLAG-dcha and Term (Annex 3). In those genes were the 5'UTR was short, oligonucleotide A was used instead of +1. Then, a second PCR with oligonucleotides +1 and Term was performed (Annex 3). The PCR fragments were ligated into pJET and digested with restriction enzymes BamHI and EcoRI, except for *ftnA* in which BamHI and KpnI were used. They were inserted into pEW or pCN51, generating the wild-type (WT) gene plasmids (Annex 2). The pEW plasmid was constructed by subcloning the transcriptional terminator region from pCN47 to the pCN40 plasmid (Charpentier *et al.*, 2004). Plasmids expressing flagged mRNAs without 3'UTRs were generated from the WT plasmids with oligonucleotides +1 and D3UTR-term (Annex 3) and then inserted into pEW or pCN51 (Annex 2).

The plasmids expressing chimeric *icaR* mRNAs were constructed taking advantage of conserved natural restriction sites located at the beginning of the *icaR* 3'UTRs. Specifically, *icaR* mRNA regions were amplified by PCR

using chromosomic DNA from the corresponding staphylococcal species and specific oligonucleotides IcaR+1 and IcaR-Term (Annex 3). The resulting PCR products were inserted into pJET and digested with SpeI and EcoRI. In the case of *S. simiae*, restriction enzymes HincII and EcoRI were used. The digested fragments were ligated into the p^{3XF}IcaRm or pIcaRm plasmid (Ruiz de los Mozos *et al.*, 2013) producing the plasmids carrying the chimeric *icaR* mRNAs listed in Annex 2.

To generate the plasmids expressing the chimeric tagged *ftnA* and *rpiRc* mRNAs (Annex 2), the corresponding 3'UTRs from different staphylococcal species were fused to *S. aureus* FtnA or RpiRc CDSs by overlapping PCRs. First, PCR fragments including the 5'UTR and the 3xFLAG CDS from the *ftnA* or *rpiRc* genes were generated by using the p^{3xF}FtnA and p^{3xF}RpiRc plasmids as templates and oligonucleotide pairs +1-ftn/CDS-stop-ftn and +1-RpiRc/CDS-stop-RpiRc as primers, respectively. Second, the 3'UTR regions from different staphylococcal species were amplified using their corresponding fw and rvs oligonucleotides (Annex 3). Finally, the PCR fragments including the 3xFLAG tagged CDS and the 3'UTR were fused by PCR using oligonucleotides +1 and rvs (Annex 3). Next, they were ligated into pJET and digested with BamHI/KpnI and BamHI/EcoRI in the case of ^{3xF}FtnA and ^{3xF}RpiRc, respectively, and inserted into pEW.

The plasmid expressing ^{3XF}FtnAΔ3'UTR$^{57\text{-}93}$ was constructed using the p^{3xF}FtnA plasmid as template using oligonucleotides +1-ftn and 3UTR-ftn-

term-1/2 (Annex 3). The construct was inserted into pEW using BamHI and KpnI restriction enzymes (Annex 2).

The p^{3xF}RpiRc+3'UTR^{IS256} plasmid was generated by PCR simulating the 3'UTR of *rpiRc* from the *S. aureus* strain 2010-60-6511-5 (GenBank: JJCE00000000.1), which carries an IS256 insertion. Briefly, IS256 was amplified from the *S. aureus* strain 15981, using IS256-fw and IS256-3'UTR-RpiRc-rvs primers and the 3'UTR fragment was amplified from p^{3xF}RpiRc, using the RpiRc-CDS-SpeI-fw and 3UTR-RpiRc-42-IS256-rvs primers (Annex 3). Then, an overlapping PCR with RpiRc-CDS-SpeI-fw/IS256-3'UTR-RpiRc-rvs primers was used to generate the 3'UTR carrying the IS256 insertion. The amplified product was ligated into pJET. Natural restriction sites were used to digest p^{3xF}RpiRc with SpeI/EcoRI, pJET^{3xF}RpiRc with NspI/EcoRI and pJET-3'UTR+IS256 with SpeI/NspI, followed by a double-fragment ligation into the pEW plasmid in order to recreate an *rpiRc* gene with the IS256 insertion in its 3'UTR.

The plasmid carrying ^{3xF}RpiRc with the IS1181 insertion in the 3'UTR, p^{3xF}RpiRc+3'UTR^{IS1181}, was generated simulating the 3'UTR of *rpiRc* from the *S. aureus* strain DAR1183 (GenBank: KK099086.1) by PCR. This was achieved by amplifying the IS1181 from the *S. aureus* strain N315 and the 3'UTR from p^{3xF}RpiRc using primers 3'UTR-RpiRc-IS1181-fw/IS1181-rvs and IS1181-3UTR-RpiRc-fw/pCN-univ-AT, respectively (Annex 3). An overlapping PCR with the primer pair 3'UTR-RpiRc-IS1181-fw/pCN-univ-rv-AT was used to generate the 3'UTR carrying IS1181, which was then cloned into pJET. Natural restriction sites (SpeI/KasI or SpeI/HindIII) were

used to digest p^{3xF}RpiRc while HindIII/KasI were used to digest pJET-3'UTR+IS1181. Then, a double-fragment ligation into the pEW plasmid was performed to recreate an *rpiRc* gene carrying the IS1181 insertion in its 3'UTR.

The GFP reporter plasmids were constructed using the *Listeria monocytogenes* pAD-cGFP plasmid as a template (Balestrino *et al.*, 2010). The *hly* 5'UTR and GFP sequences were amplified with primers SalI-GFP-fw and BcuI-TT-BamHI-GFP-rvs (Annex 3). The resulting PCR fragment was cloned into the pEW plasmid using SalI and BamHI. The 3'UTR of *ftnA* was amplified using primers BamHI-EcoRI-3UTR-ftn-fw and SmaI-3UTR-ftn-rvs (Annex 3) and inserted downstream of the *gfp* gene using the restriction sites BamHI and SmaI. The pGFP-Δ3'UTR-*ftnA* was constructed using the pGFP-3'UTR-*ftnA* as a template and primers SalI-GFP-fw and KpnI-D3UTR-term-ftn (Annex 3). The amplification product was ligated into the pEW-GFP plasmid using SalI and KpnI. The 3'UTR of *rpiRc* was amplified using primers BamHI-EcoRI-3UTR-RpiRc-fw and SmaI-3UTR-RpiRc-rvs (Annex 3) and inserted downstream of the *gfp* gene using the restriction sites BamHI and SmaI. The 3'UTRs carrying the IS were amplified from p^{3xF}RpiRc+3'UTR^{IS256} and p^{3xF}RpiRc+3'UTR^{IS1181} using BamHI-EcoRI-3UTR-RpiRc-fw and KpnI-term-RpiRc and introduced into the pEW-GFP plasmid using BamHI and KpnI. Lastly, the pGFP-Δ3'UTR-*rpiRc* was constructed using the pGFP-3'UTR-*rpiRc* as a template and primers SalI-GFP-fw and KpnI-D3'UTR-term-RpiRc (Annex 3). The

amplification product was then ligated into the pEW-GFP using SalI and KpnI.

Total protein extraction and Western blotting

Preinocula were grown in 5 ml of TSBg supplemented with Erm (TSBg+Erm) and incubated ON at 37ºC and 200 rpm. Bacterial concentrations were estimated by measuring the optical density (OD_{600}) of the preinocula and then normalized to an OD_{600} of 0.02 in Erlermeyer flasks containing TSBg+Erm. Cultures were grown at 37ºC and 200 rpm until an OD_{600} of 0.5 (exponential phase) and/or OD_{600} 5-6 (stationary phase) were reached. Culture samples were harvested by centrifugation (10 min at 4,400 g and 4ºC) and bacterial pellets stored at -20ºC until needed. Pellets were thawed, washed and resuspended in 1 ml of phosphate-buffered saline (PBS). Next, bacterial suspensions were transferred to Fast Prep tubes containing acid-washed 100 µm glass beads (Sigma) and cells were lysed using a FastPrep-24 instrument (MP Biomedicals) at speed 6 for 45 s twice. Total cell extracts were obtained by centrifuging the Fast Prep tubes for 10 min at 21,000 g and 4ºC. The total protein concentration was quantified using the Bio-Rad protein assay kit. Samples were prepared at the desired concentration in Laemmli buffer and stored at -20ºC until needed.

Western blotting was performed as previously described (Caballero *et al.*, 2018). The 3xFLAG tagged protein samples were incubated with mouse monoclonal anti-FLAG M2-Peroxidase (HRP) antibodies (Sigma) diluted

1:1,000 whereas the GFP samples were incubated with mouse monoclonal anti-GFP antibodies 1:5,000 (Living Colors, Clontech) and peroxidase-conjugated goat anti-mouse immunoglobulin G and M antibodies 1:2,500 (Pierce-Thermo Scientific). Membranes were developed using the SuperSignal West Pico Chemiluminiscent Substrate kit (Thermo Scientific). Protein bands were quantified by densitometry of Western blot images using ImageJ (http://rsbweb.nih.gov/ij/). Each of the protein levels was normalized to the levels of the *S. aureus* (Sau) sample.

RNA extraction and Northern blotting

Bacteria were grown as described in the previous section, centrifuged for 3 min at 4,400 g and 4ºC and pellets stored at -80ºC. Total RNA extraction was performed as previously described (Toledo-Arana *et al.*, 2009). Northern blotting was performed as described by Caballero *et al.* (Caballero *et al.*, 2018) with the following modifications. Radiolabeled-RNA probes were synthesized from PCR products carrying the T7 promoter (Annex 3) using the MAXIscript T7 transcription kit (Ambion) and [α^{32}P]-UTP, following the manufacturer's recommendations. Probes were then purified with Illustra MicroSpin G-50 columns (GE Healthcare) and membranes were hybridized with the corresponding RNA probes at 68ºC ON rotating. mRNA levels were quantified by densitometry of Northern blot autoradiographies using ImageJ (http://rsbweb.nih.gov/ij/). Each of the mRNA levels was normalized to the levels of the *S. aureus* (Sau) sample.

RNA stability assay

Bacteria were grown as indicated in the previous sections. Once the exponential phase (OD_{600} = 0.5) was reached, six aliquots of 20 ml of the culture were transferred to 50 ml Falcon tubes containing 300 µg ml^{-1} of Rifampicin and incubated at 37 ºC for 0, 2, 4, 8, 15 and 30 min. Then, 5 ml of Stop solution (95% ethanol, 5% phenol) were added to the samples and centrifuged for 2 min at 4,400 g. Pellets were frozen in liquid nitrogen and stored at -80ºC. RNA extraction and Northern blot analysis were conducted as described before.

***In vitro* transcription**

PCR fragments containing the T7 promoter were used as templates for *in vitro* transcription. The T7 promoter was included in the corresponding Fw oligonucleotides (Annex 3). *In vitro* transcription was carried out using the T7 polymerase at 37ºC ON followed by the removal of the PCR templates with a DNase I treatment. Transcripts were run on a denaturing 6% polyacrylamide gel and visualized using UV shadowing. RNAs were excised from the gels and eluted by incubating the gel fractions with elution buffer (0,5 M ammonium acetate, 1 mM EDTA, 0,1% SDS) and phenol (pH 4.5) at 4ºC ON. RNAs were then purified using a phenol-chloroform extraction and precipitated with ethanol. Pellets were washed with 70% ethanol and resuspended in water. The RNA integrity was checked by polyacrylamide gel electrophoresis and their concentration measured by a NanoDrop Instrument (Agilent Technologies).

5' end labeling of synthetic RNAs

Prior to labeling, the RNA was dephosphorylated with FastAP (Thermo Scientific) at 37ºC for 15 min followed by a phenol-chloroform extraction. 7 µg of dephosphorylated RNA were incubated with [γ ^{32}P]-ATP and the T4 Polynucleotide Kinase (T4 PNK, Thermo Scientific) at 37ºC for 1h. Labeled nucleic acids were purified on a denaturing 6% polyacrylamide gel as described in the previous section. The efficiency of the labeling was measured using a liquid scintillation counter.

Electrophoretic mobility assays

Labeled and non-labeled RNAs were separately prepared at the desired concentrations by adding sterile milliQ water and 5X renaturing buffer (100 mM Tris-HCl pH 7.5, 300 mM KCl, 200 mM NH_4Cl, 15 mM DTT). Samples were denatured at 90ºC for 1 min and chilled on ice for an additional minute. Next, 10 mM of $MgCl_2$ were added and samples were incubated for 10 min at 22ºC. Subsequently, the labeled RNA was mixed with increasing concentrations of non-labeled samples in 1X reaction buffer (20 mM Tris-HCl pH 7.5, 60 mM KCl, 40 mM NH_4Cl, 3 mM DTT, 10 mM $MgCl_2$) in a total volume of 20 µl. The mixtures were then incubated at 37ºC for 15 min before running them in a non-denaturing 6% polyacrylamide gel and 1X Tris-borate buffer containing 1 mM $MgCl_2$ at 300V and 4ºC. The gel was developed by autoradiography.

PIA-PNAG quantification

PIA/PNAG exopolysaccharide was extracted and quantified as previously described (Cramton *et al.*, 1999; Ruiz de los Mozos *et al.*, 2013). Cell surface extracts (5 µl), obtained from bacteria grown ON in TSBg, were spotted onto a nitrocellulose membrane using a Bio-Dot microfiltration apparatus (Bio-Rad). The membranes were blocked ON with 5% skimmed milk in phosphate-buffered saline with 0.1% Tween 20 (PBS-Tw20). After washing several times with PBS-Tw20, the membranes were incubated for 2 h with specific anti-PNAG antibodies diluted 1:20,000 (Maira-Litrán *et al.*, 2005). Bound antibodies were detected by incubating the membranes for 1h hour with peroxidase-conjugated goat anti-rabbit immunoglobulin G antibodies (Jackson ImmunoResearch Laboratories, Inc., Westgrove, PA) diluted 1:10,000. Finally, membranes were washed with PBS-Tw20 and developed using the SuperSignal West Pico Chemiluminescent Substrate.

Growth under iron limiting conditions

Preinocula were grown in 5 ml of modified chemically defined medium without iron ON at 37ºC and 200 rpm (Toledo-Arana *et al.*, 2005). Since all glass material contains iron traces, bacteria are able to grow in this medium. Therefore, in order to eliminate the remaining free iron, we used 2'2-dipyridil (DIP) (Sigma) as an iron chelator. Preinocula were normalized to an OD_{600}= 0,1 and 5 µl of these aliquots were diluted in 195 µl of modified chemically defined medium containing different concentrations of DIP in 96-well microtiter plates. The growth curve was monitored using the

SpectraMax 340 PC Microplate Reader (Molecular Devices). Measurements at $0D_{650}$ were performed every 30 min at 37ºC during 20 h.

Identification of elements disrupting *rpiRc* mRNA sequence

All available *S. aureus* genome sequences were retrieved from ftp://ftp.ncbi.nlm.nih.gov/genomes/ through the wget command-line utility. The Fasta files were loaded into a local nucleotide database using the Geneious software package. Next, a Blastn algorithm was locally run using the *rpiRc* mRNA as a query. Genomes presenting pairwise alignment disruption located at the *rpiRc* 3'UTR were registered. When possible, sequences surrounding the position of alignment disruptions were used to look for insertion sequences in the IS database (https://isfinder.biotoul.fr/) (Siguier *et al.*, 2006). Note that in some cases, due to the assembly protocol, repeated sequences were eliminated from contig ends avoiding the identification of the IS responsible for the alignment disruption.

Hemolysis assay

Bacterial cultures were grown in 15 ml of BHI ON at 37ºC and 200 rpm. Cultures were centrifuged for 10 min 4,400 g at 4ºC to collect the supernatants containing the hemolysins. Supernatants were equalized using the wet weight of the pellets and the remaining cells were removed from by filtration employing 0.2 μm pore filters. Supernatants were concentrated 10 times using a SpeedVac Concentrator. 30 μl of the concentrated supernatant were transferred into 5 mm holes within

Columbia Sheep blood (5%) agar plates (Biomerieux). Plates were incubated at 37ºC ON and then kept at 4ºC for several days.

RESULTS

RESULTS

3'UTRs are highly variable among staphylococcal species

Next-generation RNA sequencing (NGS) technologies allow to accurately determine mRNA boundaries and, therefore, determine the lengths of the untranslated regions. To evaluate the evolutionary relationship between coding sequences and their corresponding 3'UTRs, we performed genome-wide comparative analyses. Using high-resolution transcriptome maps of *S. aureus* as a reference, we analyzed the 3'UTR sequence conservation of close phylogenetic members of the genus *Staphylococcus* (Figure 17).

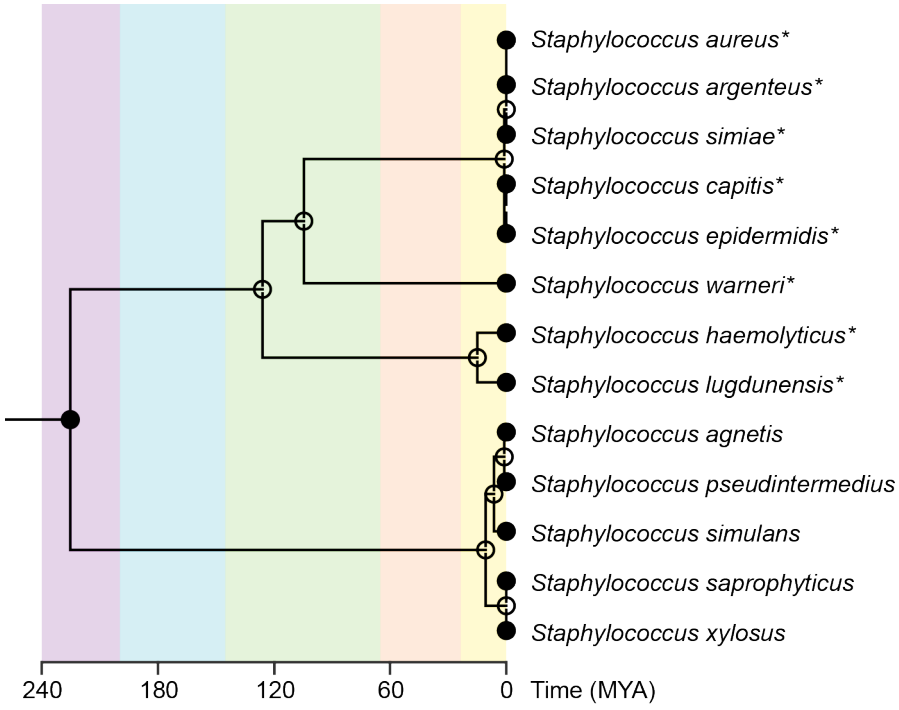


Figure 17. Phylogenetic tree including some relevant species of the *Staphylococcus* genus. Phylogenetic tree representation showing evolutionary time scales according to TimeTree knowledge-base (http://www.timetree.org) (Hedges *et al.*, 2015). Asterisk indicates those staphylococcal species included in the study. MYA, million years ago.

In order to simplify the study, we focused on monocistronic mRNAs. This ensured that each analyzed CDS was flanked by a 5'UTR and a 3'UTR. First, we generated a query database including the 3'UTR sequences plus the last 200 nucleotides of each monocistronic CDS (~590 sequences). This stretch of nucleotides was intended to serve as an indicator for the correct identification of orthologous genes among the staphylococcal species. Using blastn (Johnson *et al.*, 2008), we compared the selected sequences against 8 representative genomes of the most closely-related staphylococcal species according to the TimeTree knowledge-base (Hedges *et al.*, 2015) (Figure 17). The last conserved nucleotide position of each of the orthologous mRNAs was registered (Figure 16) and plotted against the length of their corresponding *S. aureus* mRNA (Figure 18).

The comparison of *S. aureus* NCTC 8325 and MW2 strains revealed that 97% of the analyzed mRNAs fell on the diagonal axis of the plot, indicating 3'UTR conservation (Figure 18A). Only 17 (3%) out of 589 monocistronic mRNAs showed 3'UTR sequence variations, which could be explained by the presence of unique insertion sequences (ISs) or *Staphylococcus aureus* repeat (STAR) elements (Cramton *et al.*, 2000).

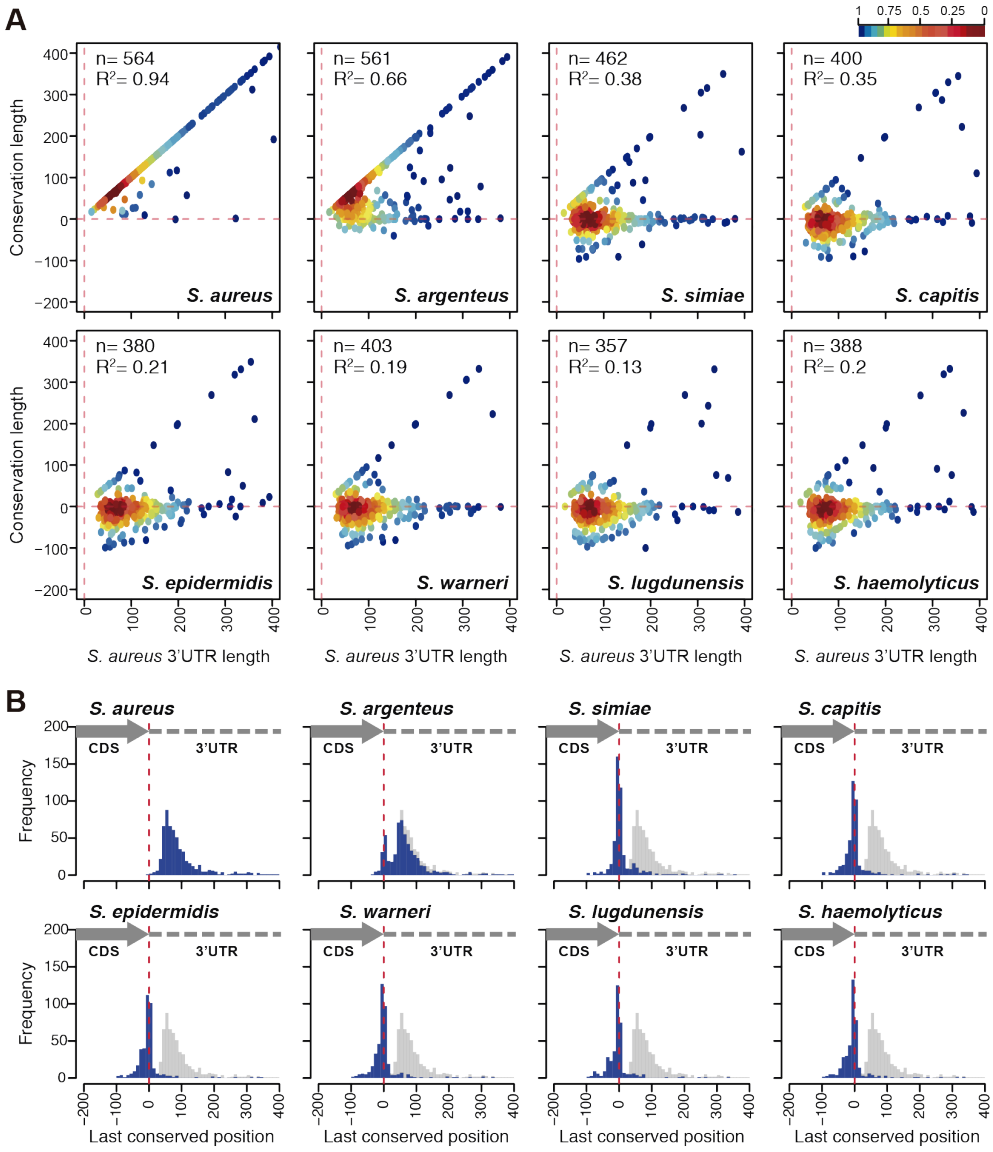


Figure 18. High-throughput conservation analysis of 3'UTRs from mRNAs encoding orthologous proteins among phylogenetically-related staphylococcal species. (A) Scatter plots representing the conservation of 3' end regions of *S. aureus* mRNAs compared to different staphylococcal species. The *S. aureus* NCTC 8325 3'UTR sequence query database was compared by blastn to *S. aureus* MW2, *S. argenteus* MSHR1132, *S. simiae* NCTC13838, *S. capitis* AYP1020, *S. epidermidis* RP62A, *S. warneri* SG1, *S. lugdunensis* HKU09-01 and *S. haemolyticus* JCSC1435 genome sequences. Each dot represents the last conserved nucleotide (y axis) of a specific species in function of the *S. aureus* 3'UTR length (x axis). The plot was colored by applying the Kernel density estimation, which indicates the proximity of the dots: blue, isolated dots; red, overlapping dots. The number (n) of plotted mRNAs and the square correlation coefficient (R^2) are indicated. **(B)** Histogram plots showing the distribution of the last conserved nucleotide position (blue bars) among the *S. aureus* mRNAs compared to the indicated species. Each blue bar represents the number of conserved 3'UTRs at a given position in windows of 10 nt (width of the bar). Grey bars represent the *S. aureus* distribution, which is included as a reference. The dashed red line indicates the position of the stop codon.

However, 3'UTR conservation among the analyzed mRNAs decreased as the phylogenetic distance between *S. aureus* and the other staphylococcal species increased. This phenomenon is reflected by the dots being positioned along the horizontal line that represents the stop codon in Figure 18A. The lack of conservation was already noticeable for several 3'UTRs of *S. argenteus*, which is the closest species to *S. aureus*. This increased considerably in *S. simiae* and left the remaining species of the genus with just a few conserved 3'UTRs (Figure 18).

As one would expect, among the conserved 3'UTRs (diagonal line) (Table 2), we identified RNAIII and several putative riboswitch-dependent 3'UTRs, meaning that the mRNA transcription ended at the TT of the downstream riboswitch when in the OFF configuration (Novick *et al.*, 1993; Toledo-Arana *et al.*, 2009; Ruiz de los Mozos *et al.*, 2013; Bronesky *et al.*, 2016). We also found three mRNAs, encoding SAOUHSC_00937, SAOUHSC_02781 and SAOUHSC_02702, which produce the 3'UTR-derived ncRNAs RsaE/F, RsaOT (i.e. srr43 or SAOUHSCs080) and Teg130 (i.e. SAOUHSCs100), respectively (Beaume *et al.*, 2010; Bohn *et al.*, 2010; Carroll *et al.*, 2016; Marincola *et al.*, 2019). The remaining mRNAs with conserved 3'UTRs encoded enolase, S1 RNA binding protein, ribosomal protein L32 (*rpmF*), and two hypothetical proteins.

Table 2. Genes with conserved 3'UTRs among the *Staphylococcus* genus

Gene ID	Name/description	3'UTR length (bp)	Relevant characteristics
SAOUHSC_02260	RNAIII/*hld*	353	Master regulator of virulence
SAOUHSC_01788	Threonyl-tRNA synthesase	332	Riboswitch-dependent 3'UTR
SAOUHSC_01091	tRNA/rRNA methyltransferase	319	Riboswitch-dependent 3'UTR
SAOUHSC_00008	Histidine ammonia lyase	306	Riboswitch-dependent 3'UTR
SAOUHSC_00826	Hypothetical protein	305	3'UTR antisense to SAOUHSC_00825
SAOUHSC_01876	Major facilitator superfamily MFS_1	270	Riboswitch-dependent 3'UTR
SAOUHSC_00937	Oligoendopeptidase F	199	It includes the 3'UTR-derived ncRNAs RsaE/F
SAOUHSC_01787	Lysine-specific permease	197	Riboswitch-dependent 3'UTR
SAOUHSC_02781	Hypothetical protein	171	It includes the 3'UTR-derived ncRNA RsaOT
SAOUHSC_02702	Hypothetical protein	155	It includes the 3'UTR-derived ncRNA Teg130
SAOUHSC_00799	Enolase	151	-
SAOUHSC_01055a	Hypothetical protein	148	-
SAOUHSC_00483	S1 RNA binding protein	120	It includes a 3'UTR-derived ncRNA
SAOUHSC_01329	30S ribosomal protein S14-2	112	Putative riboswitch-dependent 3'UTR
SAOUHSC_01078	Ribosomal protein L32	102	-

It is worth noting that, in most of the mRNAs, nucleotide conservation was lost downstream of the orthologous CDS (Figure 18A). This is illustrated in Figure 18B in which the distribution of the last conserved nucleotide position (quantified in windows of 10 nt) is shown. The top left panel of this figure shows the 3'UTR conserved length distribution in *S. aureus*, where the maximum of the peak is about 60 nt downstream of the stop codon. Meanwhile, in most of the other staphylococcal species, the maximum is located within the region comprised between –10/+10 nt from the stop codon (Figure 18B). The identification of some conserved 3'UTRs was indicative of a putative extended 3'UTR functionality. However, the finding that there was a significant lack of conservation among the 3'UTRs of orthologous genes was very intriguing and deserved further studies. The following sections of this Thesis present the main results obtained so far regarding the differential evolution of 3'UTRs in bacteria. Note that most of the results presented here have been published recently in Nucleic Acids Research (Menendez-Gil *et al.*, 2020a).

5'UTRs are more conserved than 3'UTRs

It is expected that non-coding sequences accumulate more changes throughout the course of evolution than their corresponding CDSs, which cannot significantly change without dramatically affecting protein functionality. In agreement with this, we reasoned that 5'UTRs (excluding the ribosome binding region) should suffer a similar degree of nucleotide variation than the 3'UTRs. Therefore, we performed the same blastn comparison, using a query database that included the 5'UTR sequences

plus the first 200 nt of the CDS. The scatter plots in Figure 19 show that evident variations among the 5'UTRs of orthologous genes existed. However, the number of conserved 5'UTRs was higher than their 3'UTR counterparts. This was significant when comparing the 5'UTR and 3'UTR scatter plots of specific staphylococcal species, as the number of dots closer to the diagonal line was always greater for the 5'UTRs (Figure 19). Although this analysis may have been biased by the presence of the RBS, it supported the idea of 3'UTRs being more prone to evolutionary changes than 5'UTRs.

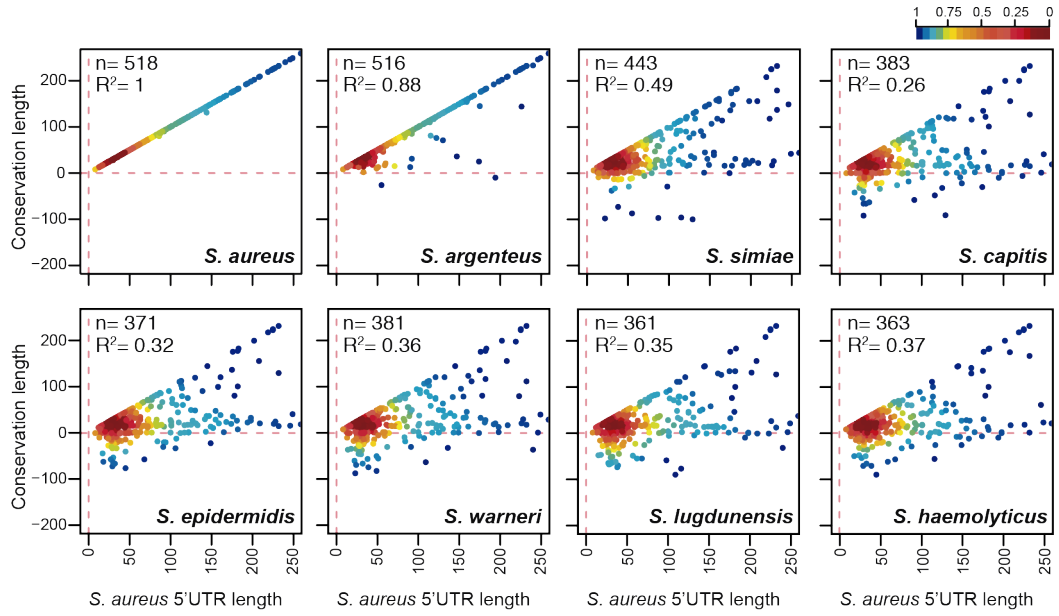


Figure 19. High-throughput conservation analysis of 5'UTRs from mRNAs encoding orthologous proteins among phylogenetically-related staphylococcal species. Scatter plots represent the conservation of 5' end regions of *S. aureus* mRNAs compared to different staphylococcal species. The *S. aureus* NCTC 8325 5'UTR sequence query database was compared by blastn to *S. aureus* MW2, *S. argenteus* MSHR1132, *S. simiae* NCTC13838, *S. capitis* AYP1020, *S. epidermidis* RP62A, *S. warneri* SG1, *S. lugdunensis* HKU09-01 and *S. haemolyticus* JCSC1435 genome sequences. Each dot represents the last conserved nucleotide (y axis) of a specific species in function of the *S. aureus* 5'UTR length (x axis). The plot is colored applying Kernel density estimation. The number (n) of plotted mRNAs encoding orthologous proteins and the square correlation coefficient (R^2) are indicated.

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

The fact that the conservation of several 3'UTRs was almost fully lost after the CDSs indicated that the mRNAs encoding orthologous proteins had different 3'UTRs. To determine and compare the real length of 3'UTRs among closely-related staphylococcal species, we performed RNA sequencing of whole transcripts from *S. simiae, S. epidermidis*, and *S. capitis* strains. The obtained transcriptomic maps were loaded into a Jbrowse-based server (Skinner *et al.*, 2009), which is publicly available at http://rnamaps.unavarra.es (some examples are represented in Figure 20). The lengths of the 3'UTRs from each *S. aureus* orthologous monocistronic gene were manually annotated by combining the RNA-seq data and the Rho-independent transcriptional terminator predictions (using TransTermHP algorithm) (Kingsford *et al.*, 2007).

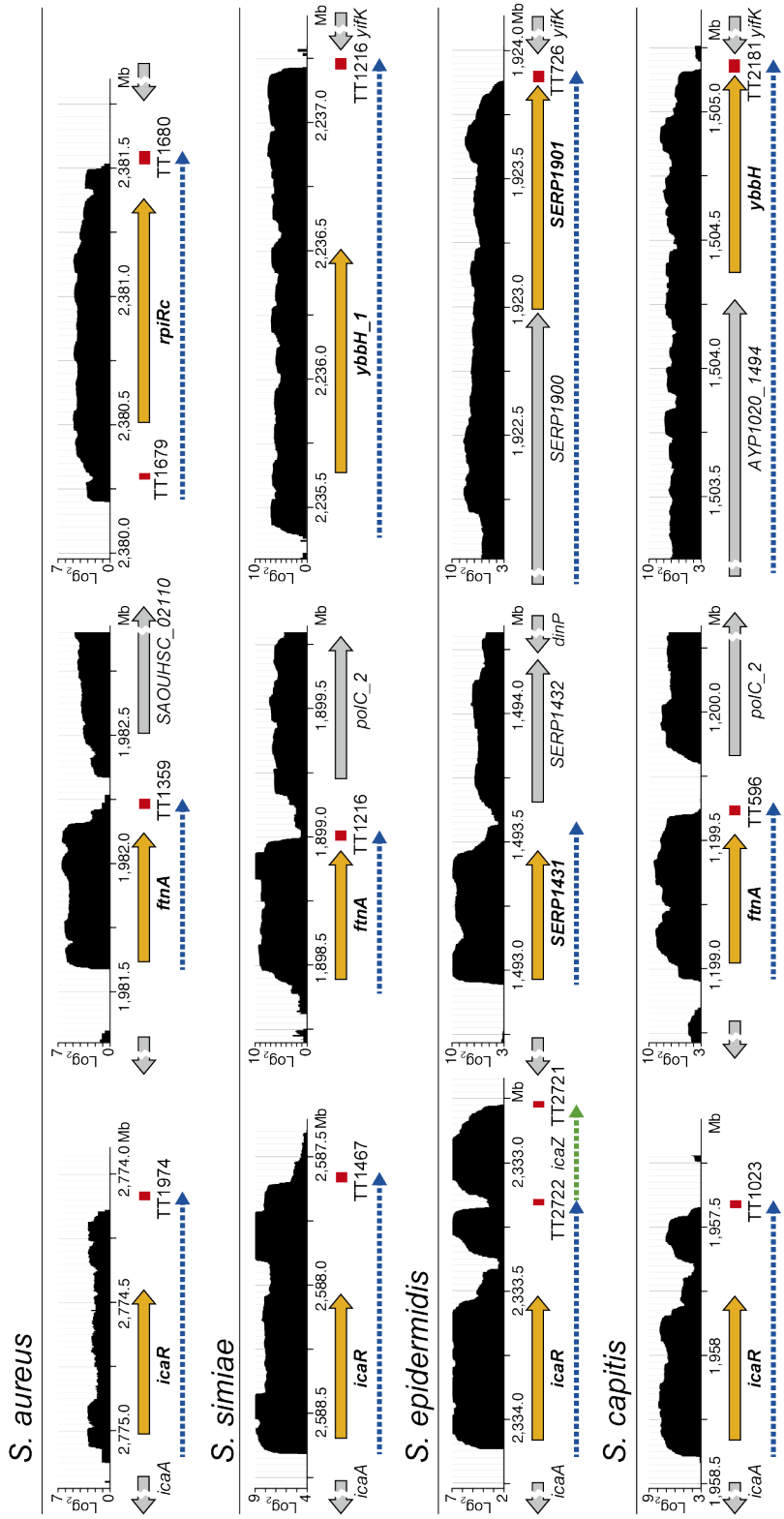


Figure 20. Browser images showing the RNA transcribed from the *icaR*, *ftnA*, and *rpiRc* chromosomic regions of the *S. aureus, S. simiae, S. epidermidis*, and *S. capitis* strains. Orange arrows, *icaR, ftnA* and *rpiRc* CDSs; grey arrows, upstream and downstream CDSs; red squares, transcriptional terminators; dashed blue arrows, *icaR, ftnA* and *rpiRc* mRNAs; dashed green arrow, icaZ ncRNA.

Subsequently, we plotted the real 3'UTR lengths from each staphylococcal species against the 3'UTR lengths of their corresponding *S. aureus* orthologous genes. Figure 21A shows that most of the 3'UTR lengths from *S. aureus* did not correlate with those of the three other analyzed staphylococcal species. This lack of correspondence was exemplified when selecting some biologically relevant genes (Figure 21B-C). As reflected in Figure 21B, in most cases, the length of the 3'UTR showed variability among all the analyzed species, while a similar 3'UTR length for a given gene was only preserved in a few examples. Overall, these data confirmed that most of the mRNAs encoding orthologous genes in staphylococcal species had 3'UTRs with different lengths in addition to sequence variation.

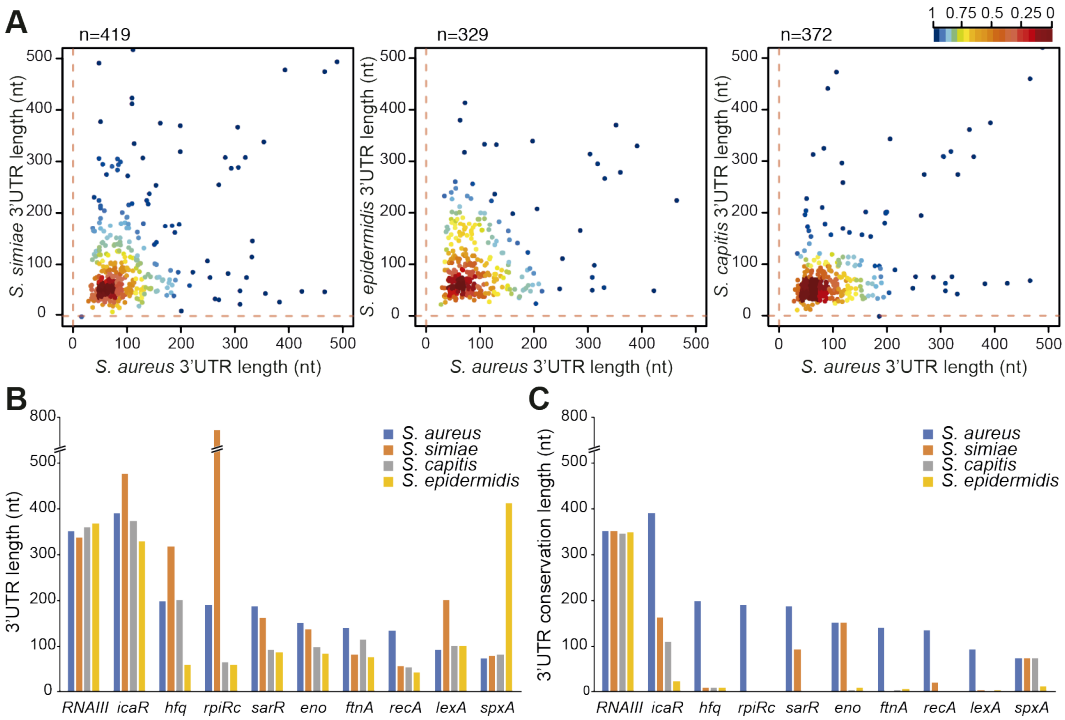


Figure 21. The lengths of the 3'UTRs do not correlate among staphylococcal species. (A) Scatter plots representing the real 3'UTR lengths of each staphylococcal species in function of the 3'UTR lengths of the corresponding orthologous *S. aureus* mRNAs. The length of the 3'UTRs from *S. aureus* orthologous monocistronic mRNAs were annotated by combining the transcriptomic data and the prediction of Rho-independent transcriptional terminators by TransTermHP (Kingsford *et al.*, 2007). The number (n) of plotted 3'UTRs is indicated; only 3'UTRs shorter than 500 nt are represented. The plot was colored by applying the Kernel density estimation. **(B)** Plot representing the 3'UTR length of relevant orthologous genes in the indicated staphylococcal species. **(C)** Plot representing the 3'UTR conservation length of the orthologous genes analyzed in B, which was determined by the blastn algorithm. RNAIII is included as an example of an mRNA with a highly conserved 3'UTR.

3'UTR sequence differences are originated by local nucleotide changes and gene rearrangements

In order to understand how 3'UTR differences originated among mRNAs encoding orthologous proteins, we compared the whole genomic sequence of *S. aureus* (Sau) against that of *S. simiae* (Ssim), *S. epidermidis* (Sepi), and *S. capitis* (Scap) using Mauve (Darling *et al.*, 2004). Note that we only focused on the monocistronic mRNAs, as stated above. We expected 3'UTR sequence variations to occur due to different genomic rearrangements, as previously described for some ncRNAs located at intergenic regions (IGRs) (Raghavan *et al.*, 2015). Therefore, we checked the sequence conservation in genomic regions downstream of the orthologous CDSs, including IGRs and adjacent CDSs, for each genomic pair (Sau vs Ssim, Sau vs Sepi, and Sau vs Scap). We assigned a value of 1 or 0 depending on whether the downstream IGR/CDS was conserved or not, respectively (Figure 22A). Confirming previous blastn analysis (Figure 18), we found that only around 6.8–10.6% of the *S. aureus* CDSs had a conserved downstream IGR/CDS when compared to the abovementioned species (Figure 22B). On the other hand, 28.3–35.2% of the orthologous CDSs lacked a conserved downstream IGR/CDS, indicating that gene rearrangements in these loci led to 3'UTR variations. The rest of CDSs showed a non-conserved downstream IGR but a preserved downstream CDS (between 58% to 62% depending on the analyzed species) (Figure 22B).

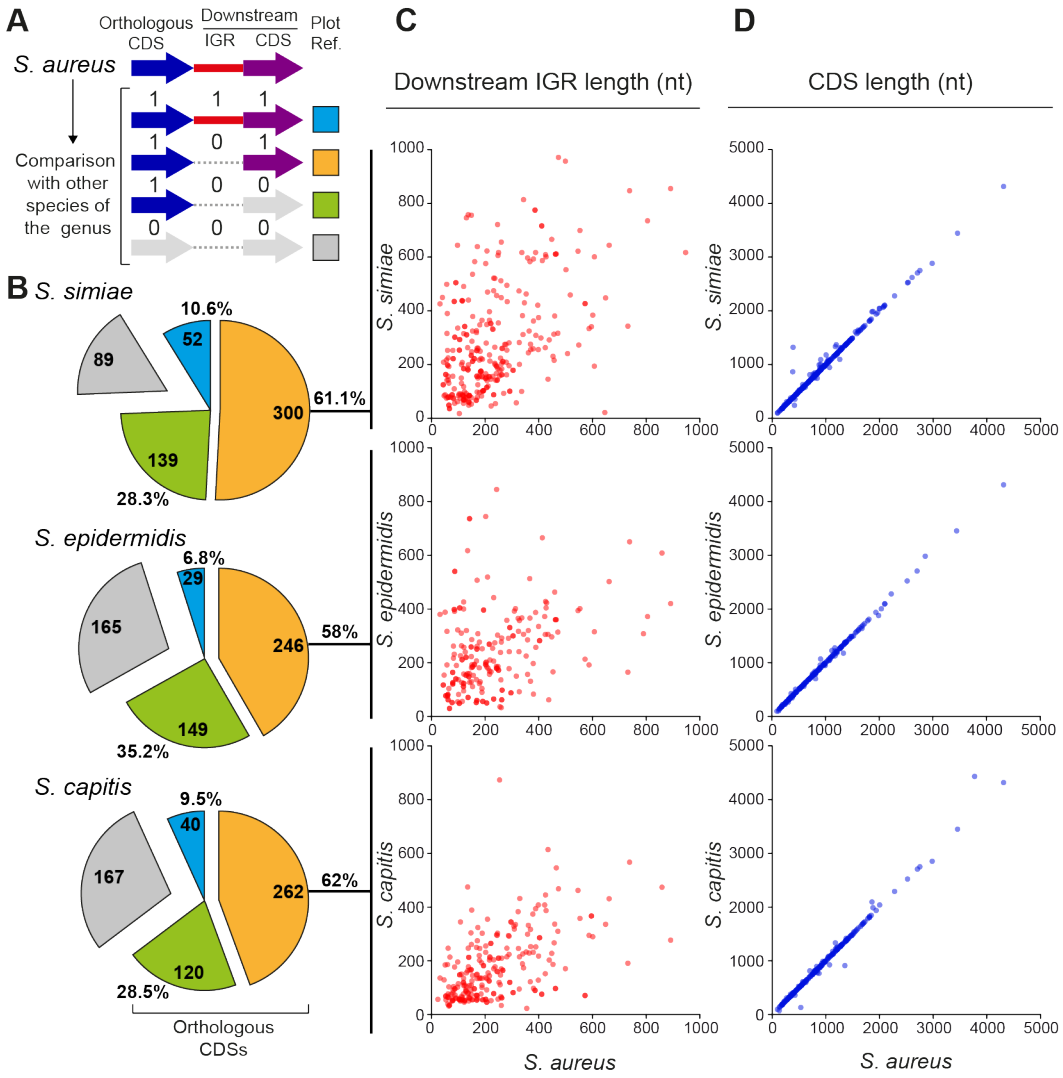


Figure 22. Nucleotide sequence variation occurring downstream of orthologous CDSs may explain 3'UTR diversity. (A) Schematic representation of the conservation analysis performed on IGRs and CDSs located downstream of an orthologous CDS. Whole-genome comparisons between *S. aureus* and its phylogenetically-related species were performed using Mauve (Darling *et al.*, 2004). Values 1 and 0 were assigned to conserved and non-conserved IGRs/CDSs, respectively, and a different color was attributed depending on the conservation configuration. Blue: the IGR and CDS downstream of the orthologous CDS are conserved; orange: the downstream CDS is conserved but not the IGR; green: both downstream regions are not conserved, and grey: no orthologous CDS was found in the analyzed species. **(B)** Pie chart quantifying the different categories represented in A. Percentages are calculated only with orthologous CDSs. **(C)** Plot showing the length of the IGR downstream of each orthologous CDS in the different species compared to that of *S. aureus*. **(D)** Plot showing the length of the orthologous CDSs in the different species compared to that of *S. aureus*. Note that only the IGRs and CDSs that fall under the orange category are plotted in C and D.

When performing this analysis, we realized that such variations substantially altered the IGR lengths (Figure 22C). Considering that the 3' end conservation was lost around the protein stop codon (Figure 18), IGR length variations may be attributed to shifts in the position of the stop codon, which may ultimately affect the CDS length. To address such a possibility, we compared the lengths of *S. aureus* CDSs to those of their corresponding orthologous genes for the three abovementioned staphylococcal species. As Figure 22D shows, most of the orthologous CDSs had similar lengths, indicating that variations in the lengths of the IGRs were not due to differences in the analyzed CDS lengths. Therefore, we concluded that the differences on the 3'UTR sequences occurred partly due to gene rearrangements and mostly because of sequence variations in the IGR sequences. The causes of local variations were unknown.

Deletion of 3'UTRs alters expression of relevant genes in *S. aureus*

3'UTR variations would be physiologically relevant if they led to differences in the expression of the orthologous genes of a particular bacterial species. To prove this hypothesis, it was necessary to identify mRNAs containing regulatory elements in their 3'UTRs. It has been previously shown that deletion of the 3'UTR from RNAIII and *icaR* mRNAs affected the expression of δ-hemolysin and IcaR, respectively (Balaban and Novick, 1995; Ruiz de los Mozos *et al.*, 2013). Our comparative analysis showed that RNAIII was among the few genes that contain a

conserved 3'UTR. This fact left *icaR* mRNA as the only model to study 3'UTR variation. Therefore, in order to identify additional functional 3'UTRs, we selected six monocistronic mRNAs that encoded relevant proteins for bacterial homeostasis and carried 3'UTRs longer than 100 nt. Table 3 lists the selected genes and the length of their corresponding 3'UTRs.

Table 3. Selection of mRNAs encoding relevant proteins and carrying long 3'UTRs

Gene ID	Name	Gene description	3'UTR length	Reference
SAOUHSC_02589	*rpiRc*	Phosphosugar-binding transcriptional regulator	189	(Gaupp *et al.*, 2016)
SAOUHSC_02566	*sarR*	Staphylococcal accessory regulator	189	(Manna and Cheung, 2001)
SAOUHSC_02108	*ftnA*	Iron storage protein	142	(Morrissey *et al.*, 2004)
SAOUHSC_01262	*recA*	SOS response regulator	135	(Cirz *et al.*, 2007)
SAOUHSC_00992	*atlR*	Autolysin regulator	117	(Houston *et al.*, 2011)
SAOUHSC_01997	*perR*	Peroxide-responsive regulator	105	(Horsburgh *et al.*, 2001)

To identify the mRNAs whose expression could be affected by putative regulatory elements located in their corresponding 3'UTRs, these genes were labeled by adding the 3xFLAG tag in the N-terminal of each protein. The whole tagged mRNA sequences (WT) were cloned into the pEW plasmid, which allowed mRNA expression under the control of the P*blaZ*

constitutive promoter (Charpentier *et al.*, 2004). Then, to compare protein expression in absence of 3'UTRs, a second set of plasmids was constructed by deleting the main 3'UTR sequences from the corresponding mRNAs (Δ3'UTRs). Note that these deletions were carried out keeping the corresponding transcriptional terminators (Figure 23A). Although the expression of all the genes was under the control of the same constitutive promoter, Western blot analyses revealed different expression protein levels for the selected genes (Figure 23). This indicated that protein expression was differentially affected at the post-transcriptional level. In fact, the plasmid including the *altR* gene did not produce any detectable protein bands. For this reason, the wild type and the Δ3'UTR mRNAs expressing ^{3xF}*altR* were cloned under the control of the P*cad* module, which is induced by cadmium (Charpentier *et al.*, 2004).

Figure 23 shows that deletion of 3'UTRs affected protein expression in different ways. On the one hand, no differences could be observed between the WT and Δ3'UTR mRNAs expressing ^{3xF}SarR and ^{3xF}AltR proteins, at least in the conditions tested. On the other hand, deletion of 3'UTRs from ^{3xF}*rpiRc,* ^{3xF}*ftnA* and ^{3xF}*perR* mRNAs increased protein expression in comparison to the WT mRNAs. In contrast, lower amounts of ^{3xF}RecA protein were expressed in absence of the 3'UTR. Note that changes in expression of ^{3xF}RpiRc protein were only observed in the stationary growth phase, suggesting the presence of growth-dependent external factors acting on this 3'UTR (Figure 23). Altogether, these results unveiled that several mRNAs carried functional 3'UTRs, highlighting the

importance of these non-coding regions in modulating protein expression in *S. aureus*.

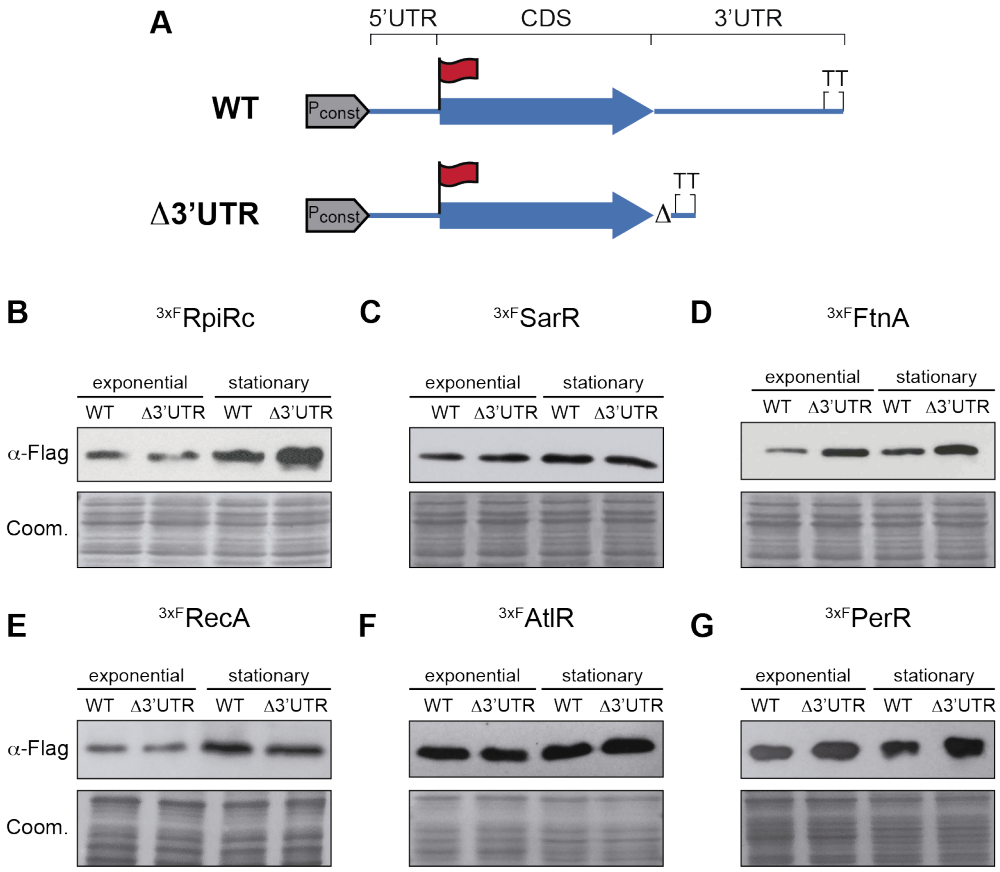


Figure 23. Identification of new regulatory 3'UTRs in *Staphylococcus aureus*. (A) Schematic representation of the constructs expressing the indicated proteins from wild type or 3'UTR-deleted mRNAs. Each protein was labeled with the 3xFLAG tag (red flag) at the N-terminal. The Δ3'UTRs constructs were created by deleting most of the 3'UTR sequence but keeping the TT functional. **(B-G)** Western blotting of ^{3xF}RpiRc **(B)**, ^{3xF}SarR **(C)**, ^{3xF}FtnA **(D)**, ^{3xF}RecA **(E)**, ^{3xF}AtlR **(F)** and ^{3xF}PerR **(G)** constructs showing the protein levels of each pair at exponential and stationary growth phases. Tagged proteins were detected with peroxidase conjugated anti-FLAG antibodies. Coomassie stained gel portions are shown as loading controls. Western images show the representative results from at least two independent replicates. Pconst, constitutive promoter; TT, transcriptional terminator; Coom, Coomassie.

Species-specific 3'UTR variations affect the expression of orthologous genes

To study if 3'UTR variations could lead to differences in the expression of the orthologous genes of a particular bacterial species, we selected the *icaR, ftnA*, and *rpiRc* genes as examples. According to the single nucleotide polymorphism (SNP) analysis carried out by Joseph and colleagues (Joseph *et al.*, 2016), the 3'UTR sequences of these mRNAs were highly conserved among different *S. aureus* strains (Figure 24A). At the same time, these 3'UTRs presented different levels of conservation when compared to their closest phylogenetically-related staphylococcal species (Figure 24B). In addition, transcript mapping by RNA-seq revealed variations in the 3'UTR lengths (Figures 20 and 21B).

According to synteny analysis, the 3'UTR of *icaR* showed significant variations in length and sequence due to gene rearrangements (Figure 25A), while the differences in the *ftnA* and *rpiRc* 3'UTRs were most likely caused by local nucleotide variations in the IGRs, as no gene rearrangements were found by synteny analysis (Figure 25B-C). Although a CDS insertion was found downstream of the *ftnA* in *S. argenteus*, it did not affect the mRNA sequence (Figure 25B).

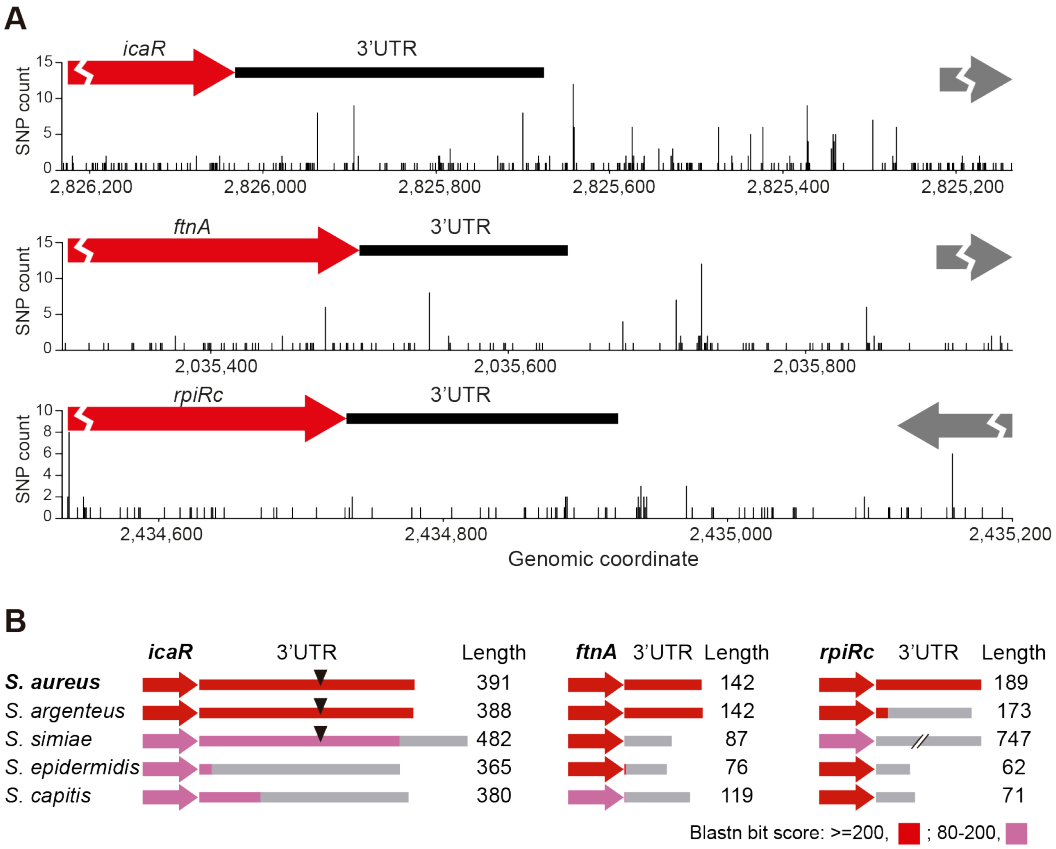
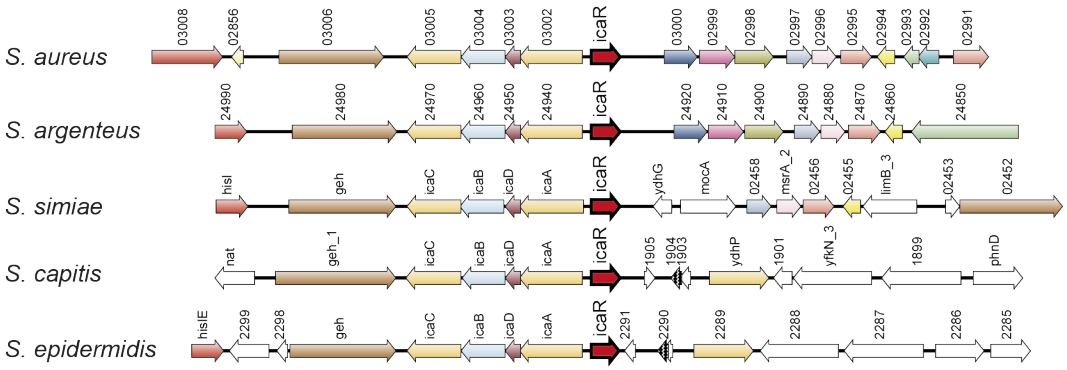
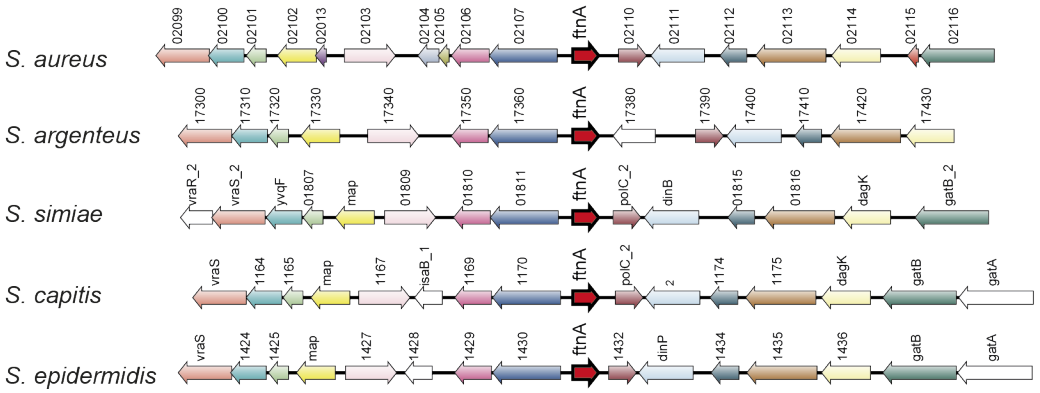


Figure 24. The *icaR*, *ftnA* and *rpiRc* 3'UTRs are conserved among *S. aureus* strains but not among staphylococcal species. (A) Plots representing the number and position of single nucleotide polymorphisms (SNPs) found in the *icaR, ftnA* and *rpiRc* mRNA regions of *S. aureus* according to the analysis carried out by Joseph *et al.* (Joseph *et al.*, 2016). **(B)** Schematic representation of the conservation analysis of the *icaR, ftnA*, and *rpiRc* mRNAs among phylogenetically-related species. The color code indicates the blastn bit score. The 3'UTR lengths in the corresponding species are indicated and represented as grey lines. Black triangles indicate the presence of the UCCCC motif (anti-RBS).

A SyntTax result: *icaR* region



B SyntTax result: *ftnA* region



C SyntTax result: *rpiRc* region

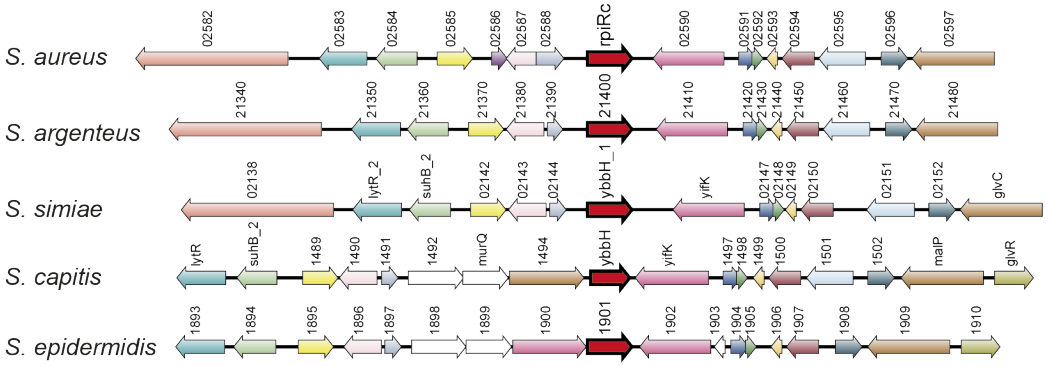


Figure 25. Comparison of *icaR*, *ftnA* and *rpiRc* genomic configurations. Synteny analyses of the chromosomic *icaR* **(A),** *ftnA* **(B)** and *rpiRc* **(C)** regions, performed using SyntTax, the Prokaryotic Synteny & Taxonomy Explorer (http://archaea.u-psud.fr/synttax/) (Oberto, 2013).

To confirm the transcript boundaries of the *icaR, ftnA*, and *rpiRc* mRNAs from *S. aureus*, *S. simiae*, *S. epidermidis*, and *S. capitis*, we performed a simultaneous mapping of their 5' and 3' mRNA ends using a modified version of the rapid amplification of cDNA ends technique (mRACE). Annexes 4 to 6 show the mRACE results that confirmed the mRNA mappings from the RNA-seq data in combination with the transcriptional terminator predictions. Note that *rpiRc* in *S. epidermidis* and *S. capitis* is the last gene of a polycistronic transcript. Therefore, we were unable to obtain mRACE mapping results from the native gene. We used Mfold to predict the putative transcriptional terminator structures of the mapped transcripts (Zuker, 2003). Interestingly, these stem-loops were different for the majority of the cases (Annex 4B, Annex 5B and Annex 6B).

Based on the information collected, we constructed chimeric mRNAs that combined the CDSs of the three selected *S. aureus* genes with the 3'UTR of their corresponding orthologous genes (from the same staphylococcal species analyzed in Figure 24B). We fused such 3'UTRs to their corresponding *S. aure*us CDS, which carried the 3xFLAG sequence at the N-terminal, to monitor their protein expression (Figure 26A). These plasmids were then transformed into the mutant strains that lacked the original genes to facilitate the detection of the mRNAs expressed from the plasmids. To confirm the mapping of the 3' end of the chimeric *rpiRc* mRNA carrying the 3'UTRs from *S. epidermidis* and *S. capitis*, we performed mRACE. The results showed that the transcripts ended at the

expected sites, which were downstream of the predicted stem-loops (Annex 6).

Expression analysis of the 3xFLAG-tagged proteins by Western blotting revealed that the deletion of the 3'UTRs produced an increase of the IcaR, FtnA, and RpiRc proteins (Figure 26B). Interestingly, the chimeric *icaR* mRNAs harboring the UCCCC motif (3'UTRs from *S. argenteus* and *S. simiae*), which was required for modulation of IcaR translation in *S. aureus* (Ruiz de los Mozos *et al.*, 2013), presented similar IcaR protein levels to those shown by the *S. aureus* WT mRNA. In contrast, the chimeric *icaR* mRNAs lacking the UCCCC motif (*S. epidermidis* and *S. capitis icaR* 3'UTRs) expressed a comparable amount of the IcaR protein to that of a *S. aureus* mRNA carrying the 3'UTR deletion (Figure 26B). Chimeric mRNAs expressing FtnA and RpiRc showed protein yields similar to those of the same mRNAs lacking the 3'UTR. The exception to such behavior was the chimeric construct carrying the *ftnA* mRNA in combination with the 3'UTR of *S. argenteus*, whose sequence was conserved when compared to *S. aureus* (Figure 24B).

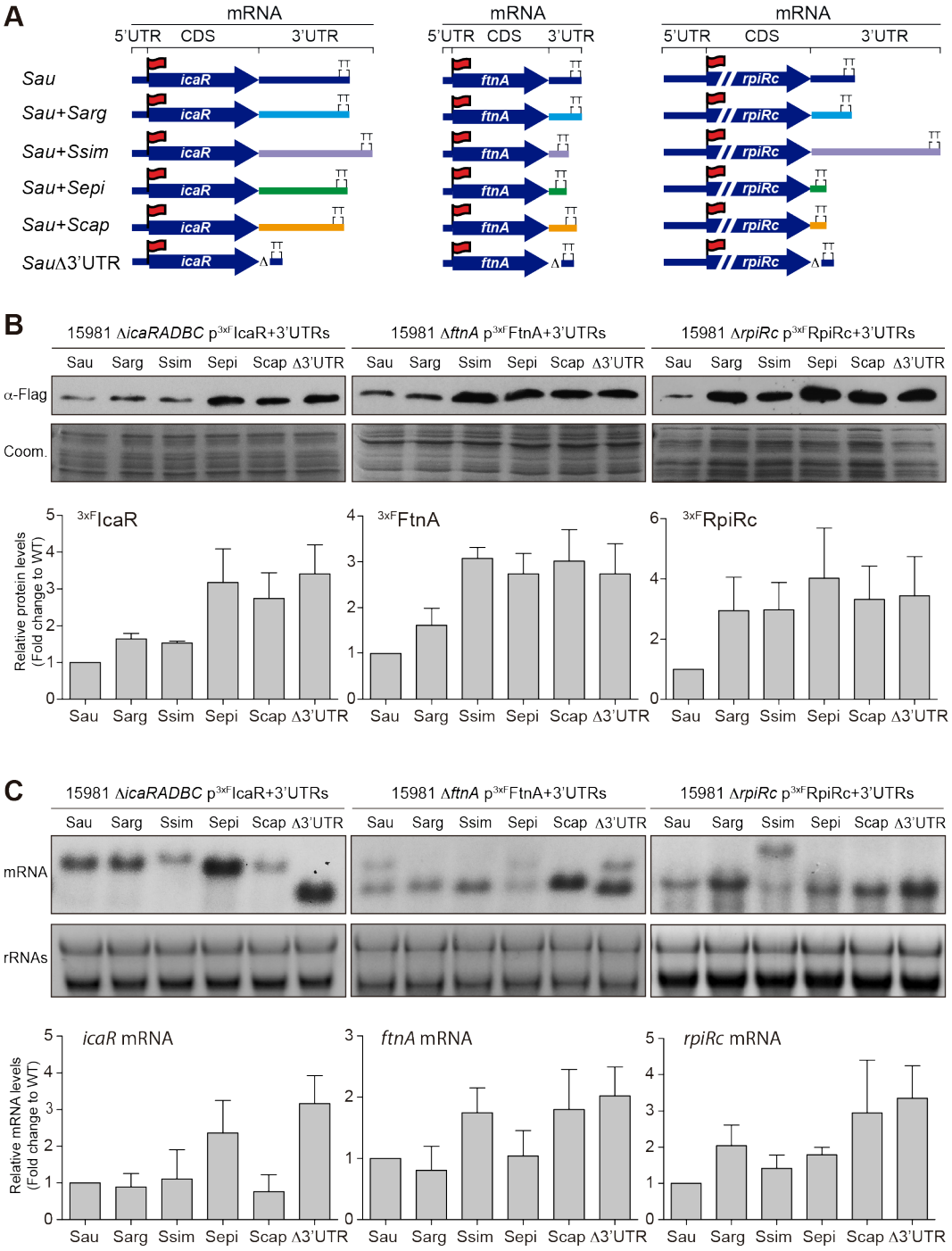


Figure 26. 3'UTRs variations affect protein expression. (A) Schema of constructed chimeric mRNAs combining the *S. aureus* CDS with a 3'UTR from the indicated species. Red flags indicate the 3xFLAG tag in the N-terminus. **(B)** Western blot (WB) showing the levels of proteins expressed from different chimeric mRNAs. **(C)** Northern blot (NB) showing the mRNA levels expressed from the constructs shown in B. Stained protein and RNA gel portions are shown as loading controls. WB and NB bands were quantified by densitometry of blot images using ImageJ from at least three biological replicates. TT, transcriptional terminator; Sau, *S. aureus*; Sarg, *S. argenteus*; Ssim, *S. simiae*; Sepi, *S. epidermidis*; Scap, *S. capitis;* Coom, Coomassie.

The same samples were then subjected to Northern blot analyses. The results showed that, in most cases, the steady-state mRNA levels correlated with the protein levels (Figure 26C). This suggested that the increase in protein levels, which occurred when the original 3'UTR sequence was substituted by the one present in its orthologous mRNA, was most likely due to an increase in mRNA stability, as previously described for the *icaR*Δ3'UTR mRNA (Ruiz de los Mozos *et al.*, 2013). To confirm this, we measured the half-life of the chimeric *icaR*+3'UTR^{Sepi} mRNA as an example. Figure 27 shows that this chimeric mRNA presented a higher half-life in comparison with the *icaR* WT mRNA.

Note that the *icaR*+3'UTR^{Scap} and *ftnA*+3'UTR^{Sepi} chimeras did not show increased mRNA levels, implying that protein expression could be affected in other ways (Figure 26C). Altogether, these results indicated that species-specific variations in the 3'UTR sequences could not reproduce the expression levels found in *S. aureus*, suggesting putative distinct functional roles for 3'UTRs in the different species.

The following sections present the studies carried out to demonstrate the biological relevance of the 3'UTR-mediated regulation and the consequences of nucleotide variations in the 3'UTRs of *icaR*, *ftnA* and *rpiRc* mRNAs.

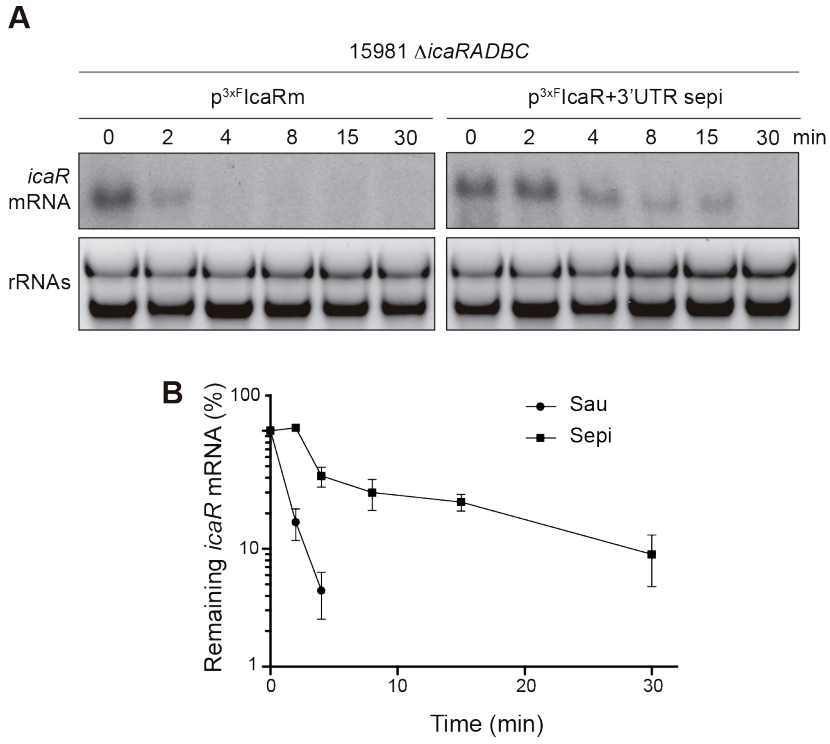


Figure 27. The chimeric *icaR* mRNA including the 3'UTR of *S. epidermidis* has a longer half-life than the *icaR* WT mRNA. (A) Half-life measurement of *icaR* wild type and *icaR*+3'UTR^{Sepi} mRNAs, which are constitutively expressed from the P*blaZ* promoter. The strains carrying these constructs were grown in TSBg at 37ºC until exponential phase and then Rifampicin (300 μg/ml) was added. Samples for RNA extraction were taken at the indicated time points (min). The experiment was repeated two times and representative images are shown. **(B)** Levels of mRNAs were quantified by densitometry of Northern blot autoradiographies using ImageJ (http://rsbweb.nih.gov/ij/). The mRNA levels were normalized to mRNA levels at time 0. The mean of mRNA levels was plotted as a function of time. Error bars indicate the standard deviation of mRNA levels from two independent experiments.

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

The expression analysis of chimeric *icaR* mRNAs indicated that the *icaR* 3'UTRs from *S. epidermidis* and *S. capitis* were unable to reproduce the 3'UTR-mediated regulation observed in *S. aureus* (Figure 26B). As previously described, the *icaR* 3'UTR interacts with the 5'UTR of the same mRNA molecule through a UCCCC motif to repress IcaR translation and,

in return, modulates PIA-PNAG synthesis and biofilm formation in *S. aureus* (Ruiz de los Mozos *et al.*, 2013). The absence of the UCCCC motif in the *S. epidermidis* and *S. capitis icaR* 3'UTRs may account for the differences in IcaR expression. This may be attributed to the lack of interaction between their 5' and 3'UTRs. To test this hypothesis, we carried out electrophoretic mobility shift assays (EMSAs). The *S. aureus icaR* 5'UTR was synthesized *in vitro*, radioactively labeled, and then incubated with increasing concentrations of cold *icaR* 3'UTRs from the five staphylococcal species analyzed in Figure 26. Figure 28A shows that the *S. aureus icaR* 3'UTR produced the expected gel-shifts when interacting with the 5'UTR, as previously described (Ruiz de los Mozos *et al.*, 2013). Consistently, a similar behavior was observed when using the 3'UTR of *S. simiae* (Figure 28B*)*. In the case of the *S. argenteus icaR* 3'UTR, despite sharing a relatively high sequence identity with its *S. aureus* counterpart (82% of identity), a much lower binding affinity was found (Figure 28B). This indicated that although *S. argenteus* preserved the UCCCC motif, the nucleotide differences might affect the binding affinity *in vitro*. *S. epidermidis* and *S. capitis* 3'UTRs, as one would expect, did not produce gel shifts when incubated with the 5'UTR due to the absence of the UCCCC motif in their sequences (Figure 28C). Moreover, recent results showed that the 5' and 3'UTRs from *S. epidermidis*, unlike what was observed for *S. aureus*, did not interact between them, excluding the existence of an orthologous mechanism (Lerch *et al.*, 2019). In contrast, it was found that the *S. epidermidis*-specific ncRNA IcaZ, downstream of the

icaR 3'UTR, targeted the *icaR* 5'UTR and prevented IcaR translation (Lerch *et al.*, 2019). Altogether, these results indicated that sequence variations around the 3'UTR determine the presence or absence of specific motifs and, ultimately, the generation of functional distinctions.

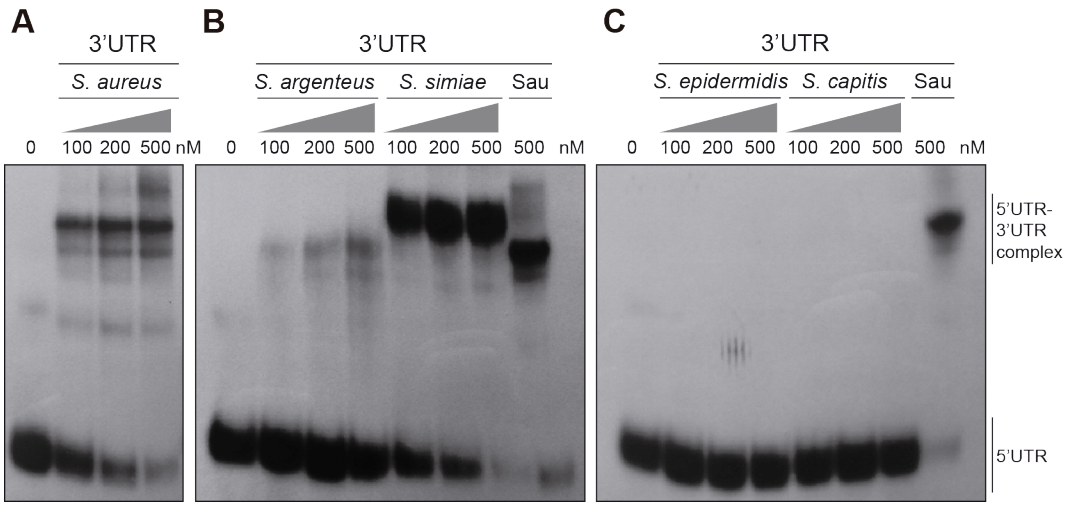


Figure 28. Only *icaR* 3'UTRs carrying the UCCCC motif bind to the *S. aureus icaR* 5'UTR *in vitro*. Electrophoretic mobility shift assays (EMSAs) between the synthetic *S. aureus icaR* 5'UTR and the synthetic *icaR* 3'UTR from: **(A)** *S. aureus* as a positive control; **(B)** *S. argenteus* and *S. simiae,* which carry the UCCCC motif necessary for the 5'UTR interaction, as previously described (Ruiz de los Mozos *et al.*, 2013); **(C)** *S. epidermidis* and *S. capitis*, which lack the UCCCC motif. The autoradiographies of the band-shifts result from incubating a ^{32}P-labeled synthetic *S. aureus* 5'UTR RNA fragment (40,000 cpm) with increasing amounts of the different synthetic 3'UTR RNAs (100 to 500 nM). The UTR complexes are indicated on the right side of the autoradiography. Images show representative results from at least two independent experiments. Sau, *S. aureus*.

In order to demonstrate that the differences observed in terms of IcaR expression by chimeric mRNAs were enough to alter PIA-PNAG synthesis and, thus, be of biological relevance, we introduced the plasmids expressing these chimeric mRNAs into the *S. aureus* 15981 strain, which produces high levels of PIA-PNAG exopolysaccharide. The quantification of PIA-PNAG levels by dot-blot using specific antibodies against the

exopolysaccharide showed that the strains expressing higher levels of IcaR (the chimeric Scap/Sepi and Δ3'UTRs strains, which lacked the UCCCC motifs) presented lower levels of PIA-PNAG in comparison with the strains producing IcaR from mRNAs containing the UCCCC motif (the WT and chimeric Sarg/Ssim strains) (Figure 29). These results confirmed that evolutionary variations in 3'UTRs, which occur among phylogenetically-related bacteria, are enough to create gene expression differences that affect important biological processes, such as exopolysaccharide production. This suggested that the synthesis of orthologous proteins might be differentially affected by 3'UTR sequence variations.

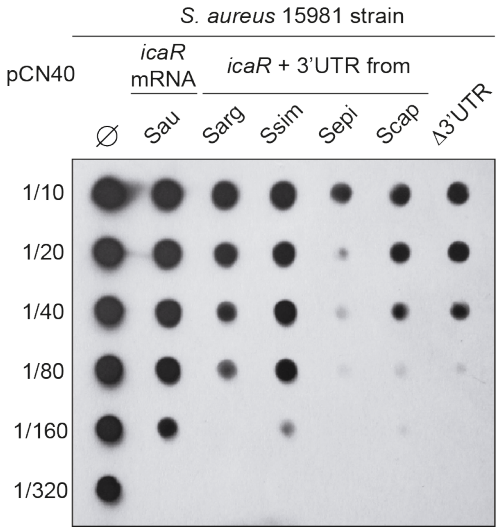


Figure 29. Species-specific variations in *icaR* 3'UTRs resulted in different PIA-PNAG levels. Quantification of PIA-PNAG exopolysaccharide production in the *S. aureus* 15981 strains expressing the different chimeric *icaR* mRNAs. Serial dilutions of the samples were spotted onto nitrocellulose membranes and PIA-PNAG expression was developed with specific anti-PIA-PNAG antibodies. Images show representative results from at least two independent experiments. ∅ indicates the presence of an empty pCN40 plasmid. Sau, *S. aureus*; Sarg, *S. argenteus*; Ssim, *S. simiae*; Sepi, *S. epidermidis*; Scap, *S. capitis*.

The *ftnA* 3'UTR is targeted by RNase III and PNPase to reduce mRNA levels

We previously showed that deletion of the *ftnA* 3'UTR and the inclusion of 3'UTRs from other staphylococcal species increased ^{3xF}FtnA protein levels (Figure 26B). However, the mechanism by which the *S. aureus ftnA* 3'UTR was capable of modulating FtnA expression was unknown. Therefore, we wondered whether the *ftnA* 3'UTR could play a role in translation initiation by interacting with the 5'UTR as *icaR* or could affect mRNA stability by leading a specific mRNA processing. On the one hand, we predicted the secondary structure of the *ftnA* mRNA using the Mfold application (Zuker, 2003). We could not find any evident interaction between the 3' and the 5'UTRs suggesting that the 3'UTR was not directly inhibiting translation as previously described for IcaR (Ruiz de los Mozos *et al.*, 2013). On the other hand, since the Northern blots comparing the *ftnA* mRNA levels showed that the strain carrying the *ftnA*Δ3'UTR plasmid produced higher *ftnA* mRNA levels than the one expressing the WT *ftnA* mRNA (Figure 26C), we analyzed if the *ftnA* 3'UTR could affect the mRNA stability by carrying a target site for a putative RNase. In order to identify the RNases that could potentially process the *ftnA* mRNA, we transformed the most relevant RNase mutants of *S. aureus*, including RNase III (Δ*rnc*), mini-RNase III (Δ*mrnc)*, PNPase (Δ*pnpA*), RNase R (Δ*rnr*), RNase Y (Δ*rny*) and RNase J1 (Δ*rnjA*), with the plasmids expressing the ^{3xF}*ftnA* WT and Δ3'UTR mRNAs. Western blots showed that, although the expression of ^{3xF}FtnA was variable among the RNase mutants compared to the WT

strain, the increased protein expression pattern due to the 3'UTR deletion was still observed in Δ*rnr,* Δ*mrnc* and Δ*rny* mutants (Figure 30A). Since we could not detect ^{3xF}FtnA in a Δ*rnjA* mutant, we repeated the Western blot but loading higher amounts of total protein extracts. Similarly, we could still observe protein differences when comparing the ^{3xF}*ftnA* WT and Δ3'UTR mRNAs (Figure 30B). In contrast, no differences in ^{3xF}FtnA levels could be detected between *ftnA* WT and Δ3'UTR mRNAs when they were expressed in the Δ*rnc* and Δ*pnpA* mutants. This result suggested that RNase III and PNPase could be targeting the *ftnA* mRNA through the 3'UTR. Note that ^{3xF}FtnA levels were higher in the Δ*rnc* than the WT strain, pointing out that this RNase could also be targeting other *ftnA* mRNA regions besides the 3'UTR (Figure 30A).

To confirm that the *ftnA* 3'UTR carried functional target sites that could be recognized by RNase III and PNPase, we fused this 3'UTR to a heterologous gene such as the green fluorescent protein (*gfp*), generating the pGFP-3'UTR-*ftnA* plasmid. As a negative control, the transcriptional terminator of the *ftnA* mRNA was cloned downstream of the *gfp* gene (pGFP-Δ3'UTR). Western blot analysis revealed that GFP levels were lower when this protein was expressed from the pGFP-3'UTR-*ftnA* plasmid in comparison to the one just including the *ftnA* TT (Figure 30C). This result indicated that *ftnA* 3'UTR alone was able to reduce protein expression. Interestingly, this effect could be abolished when the pGFP-3'UTR-*ftnA* plasmid was introduced into Δ*rnc* and Δ*pnpA* mutants (Figure 30D). This confirmed that these RNases targeted *ftnA* 3'UTR. Since the

action of PNPase is inhibited by transcriptional termination hairpin loops, it might be expected that PNPase acts after RNase III cleavage. Note that RNase III is a double stranded endoRNase. However, no secondary structures susceptible to be attacked by RNase III could be predicted in the *ftnA* 3'UTR. Therefore, it is plausible to think that a ncRNA could pair to the *ftnA* 3'UTR to create a double stranded substrate for this RNase.

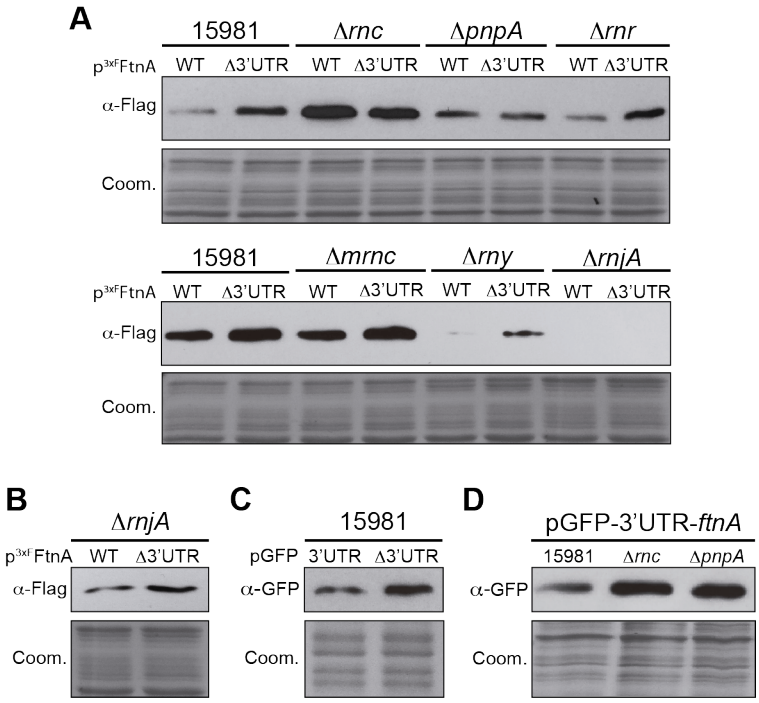


Figure 30. Role of RNases in *ftnA* 3'UTR-mediated regulation. (A) Western blot showing the ^{3xF}FtnA levels of WT and Δ3'UTR constructs in different RNase mutants. **(B)** Western blot of ^{3xF}FtnA in the Δ*rnjA* mutant with a protein load higher than in A. **(C)** Western blot showing GFP levels of 15981 strains carrying either pGFP-3'UTR-*ftnA* or pGFP-Δ3'UTR. **(D)** Western blot showing GFP levels of 15981, Δ*rnc* and Δ*pnpA* strains carrying pGFP-3'UTR-*ftnA*. The Western blots from A and B were developed using peroxidase conjugated anti-FLAG antibodies, while the Western blots from C and D were developed using monoclonal anti-GFP antibodies and peroxidase-conjugated goat anti-mouse immunoglobulin G and M antibodies. Coomassie (Coom) stained gel portions are shown as loading controls. Western blots images show the representative results from at least three independent replicates.

The putative ncRNA/RNase target site should be included in the 74-nucleotide region that has been deleted in the *ftnA* 3'UTR mutant, which is comprised between positions 20 to 93 downstream of the *ftnA* stop codon. To determine the putative functional region across these 74 nt, we created a second 3'UTR mutant by deleting the nucleotides from position 57 to 93 of the *ftnA* 3'UTR (*ftnA*Δ3'UTR$^{57\text{-}93}$) (Figure 31A). This plasmid was transformed into the WT strain and the protein expression was compared to the strains expressing the ^{3xF}*ftnA* WT and Δ3'UTR mRNAs, respectively. Western blot analysis revealed that the mRNA carrying the Δ3'UTR$^{57\text{-}93}$ deletion expressed similar ^{3xF}FtnA protein levels than the WT *ftnA* mRNA. This result indicated that the putative target site should be located in the region comprised between positions 20 and 56 of the *ftnA* 3'UTR (Figure 31B).

Our comparative analysis showed that the 3'UTR sequences from *S. simiae*, *S. capitis* and *S. epidermidis ftnA* mRNAs were completely different (Figure 24B). This meant that the putative target site was absent in these species and, therefore, it might explain the increased ^{3xF}FtnA protein levels found in these chimeric mRNAs (Figure 26B). Interestingly, nucleotide comparison analysis between *S. aureus* and *S. argenteus* 3'UTRs showed that the putative functional region was highly conserved in comparison to the surrounding nucleotides (Figure 31C). This was a strong evidence supporting that this region effectively contained a target site responsible for reducing *ftnA* mRNA levels in both *S. aureus* and *S. argenteus* species.

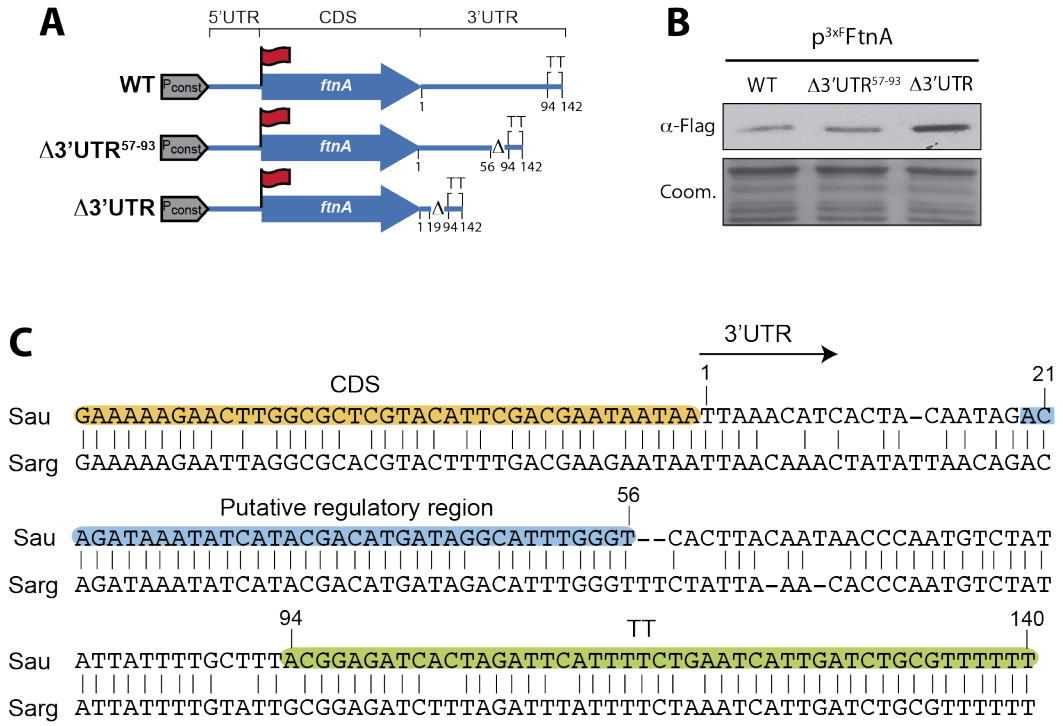


Figure 31. The putative regulatory region in the *ftnA* 3'UTR is located between positions 20 to 56 after the stop codon. (A) Schematic representation of the constructs generated to identify the functional region of the *ftnA* 3'UTR. **(B)** Western blot showing the levels of ^{3xF}FtnA when expressed from the constructs shown in section A. A Coomassie stained gel portion is shown as a loading control. Western blot images show the representative results from at least three independent replicates. **(C)** Blastn alignment of *ftnA* 3'UTR from *S. aureus* and *S. argenteus*. The nucleotides corresponding to the CDS, the putative regulatory region and the TT are highlighted in orange, blue and green, respectively. The arrow indicates the start of the 3'UTR. Pconst, constitutive promoter; TT, transcriptional terminator; Coom, Coomassie; Sau, *S. aureus*; Sarg, *S. argenteus*.

The *ftnA* 3'UTR deletion affects adaptation to iron starvation

The control of the iron pool inside the cells is essential to allow bacterial growth. On the one hand, there should be enough iron as a cofactor for a wide variety of enzymes while, on the other hand, iron excess should be avoided to prevent oxidative stress that could lead to cell damage (Andrews *et al.*, 2003; Hood and Skaar, 2012).

The *ftnA* gene encodes for the iron storage protein ferritin, a ubiquitous protein found in all cells. Ferritin function appears to be storage of iron, removing the intracellular iron to protect cells from its toxic effects. Transcription of *ftnA* mRNA is induced by iron and is highly expressed under anaerobiosis, most likely due to the increased solubility of iron in these conditions (Horsburgh *et al.*, 2001; Zühlke *et al.*, 2016). Under iron starvation conditions, the peroxide stress transcriptional regulator PerR, represses *ftnA* transcription (Horsburgh *et al.*, 2001; Morrissey *et al.*, 2004). This is required to avoid iron storage when it is scarce. In addition to this tight transcriptional regulation, our data showed that ferritin expression could also be modulated at the post-transcriptional level through the *ftnA* 3'UTR.

In order to evaluate the biological relevance of the *ftnA* 3'UTR-mediated regulation, we compared the capacities of the WT and the chromosomal *ftnA*Δ3'UTR mutant strains to grow under iron starvation conditions. To do that, these strains were inoculated into microtiter wells containing a modified chemically defined medium lacking iron (Toledo-Arana *et al.*, 2005). Since iron traces are still present in the medium, we added increasing concentrations of 2,2'-dipyridyl (DIP), an iron chelator. Microplates were incubated at 37ºC and bacterial growth was measured by registering the optical density every 30 minutes.

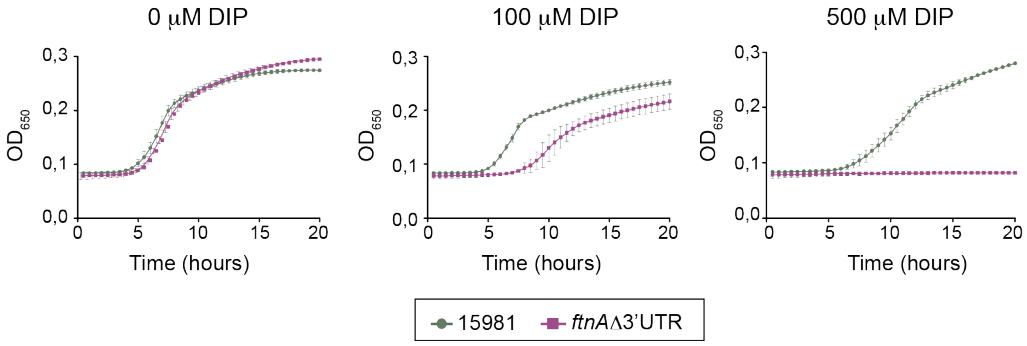


Figure 32. The 3'UTR of *ftnA* is essential under iron starvation conditions. 15981 WT and *ftnA*Δ3'UTR strains were grown for 20 hours in a modified chemically defined medium without iron (Toledo-Arana *et al.*, 2005) and increasing concentrations of the iron chelator 2'2-dipyridil (DIP). OD_{650} measurements were registered every half hour. Error bars represent the standard deviation of three independent replicates.

Figure 32 shows that the addition of the DIP chelator at a concentration of 100 μM significantly affected the growth of the *ftnA*Δ3'UTR mutant while 500 μM completely impaired it. Considering that the deletion of the *ftnA* 3'UTR increased ferritin concentration (Figure 26B), it could be speculated that higher ferritin levels would sequester the scarce iron available inside the cells. As result, the essential functions carried out by enzymes requiring iron as a cofactor would be affected leading to bacterial growth arrest. Altogether, these data demonstrated that the *ftnA* 3'UTR is essential to modulate ferritin expression and to maintain proper iron levels for *S. aureus* growth. Further studies are needed to decipher what are the molecular mechanisms by which RNase III and PNPase target this 3'UTR to reduce *ftnA* mRNA levels.

IS transposition to *rpiRc* 3'UTR modifies RpiRc expression and hemolysin production

When studying the *rpiRc* mRNA in the *S. simiae* CCM 7213T strain, we found that the *rpiRc* 3'UTR was considerably longer than the 3'UTRs from other staphylococcal species (Figure 24B). A deep sequence analysis revealed that the *S. simiae rpiRc* 3'UTR contains three copies of STAR elements (Cramton *et al.*, 2000), which are distributed among staphylococcal species (Purves *et al.*, 2012). Considering that the number and genomic locations of the STAR element copies are variable among species and strains (Purves *et al.*, 2012), we investigated whether insertions of STAR elements may also occur in the *rpiRc* 3'UTR of *S. aureus* and, therefore, alter protein expression as observed in the chimeric *rpiRc* construct carrying the 3'UTR of *S. simiae* (Figure 26). Since SNP analysis were not indicative of the presence of insertion/deletion sequences, we performed *rpiRc* mRNA sequence pairwise alignments to look for alignment disruptions in the *rpiRc* 3'UTR. We reasoned that alignment disruptions could help us identify repeated regions. Following this idea, we used the whole *rpiRc* mRNA sequence as the query to perform blastn against all draft and complete *S. aureus* genomes available on the NCBI database.

The results showed that 33 *S. aureus* genomes presented alignment disruptions within the *rpiRc* 3'UTR (Annex 7). Since many of these genomes were not fully closed, some alignment disruptions occurred due to the query sequence aligning to different contigs. By analyzing the

extremes of the aligned contigs, we identified that 16 out of the 33 genomes contained IS insertions in different locations of the *rpiRc* 3'UTR sequence. Specifically, we were able to map nine IS1181 (Chesneau *et al.*, 1999), five IS256 (Lyon *et al.*, 1987) and two ISs of the IS30 transposase family (Annex 7 and Figure 33A). The remaining genomes presented sequence deletions or duplications. Interestingly, we identified duplications of 8-12 nt in some of the strains, which were indicative of the presence/excision of ISs. The fact that no IS sequences were found in the related contigs may be explained by either the use of assembly algorithms that eliminated repeated sequences from the extremes of the contigs or the ISs being excised. However, the latter would not explain why these contigs were not assembled together, suggesting that an unknown IS may be inserted in the *rpiRc* 3'UTR of these genomes.

In order to analyze the consequences of such insertions, we constructed two plasmids containing the IS256 and IS1181 sequences 42 and 108 nt downstream of the RpiRc stop codon, respectively. These mimicked the chromosomic configurations found in the *rpiRc* region of the *S. aureus* 2010-60-6511-5 and DAR1183 strains, respectively (Figure 33A-B). In both cases, the insertion orientation of the IS placed the transposase gene in the DNA strand opposite to that of *rpiRc*. Northern blot analysis revealed different chimeric mRNA lengths (Figure 33C). On the one hand, IS1181 carried a stem-loop downstream of the transposase gene, in the same DNA strand as the *rpiRc* gene, which acted as a transcriptional terminator (Figure 33B). Therefore, the IS1181 insertion produced a

chimeric *rpiRc-IS1181* mRNA of ~1.4 Kb (3'UTR-IS1181 of 222 nt) that included a few nucleotides of the IS1181 sequence. This band migrated in a similar manner as the original *rpiRc* mRNA. The impact of this insertion was an increase in the mRNA levels (Figure 33B and C), which correlated with an increase in the ^{3xF}RpiRc protein levels (Figure 33D). On the other hand, since the IS256 lacked stem-loops able to prevent transcription from adjacent genes, a chimeric *rpiRc-IS256* mRNA of ~2.8 Kb (3'UTR-IS256 ~1.5 Kb) was produced. In contrast to the IS1181 insertion, the chimeric *rpiRc-IS256* mRNA included the whole IS256 sequence (Figure 33B and C). In this case, there were no considerable changes neither in the mRNA nor in the ^{3xF}RpiRc protein levels (Figure 33C and D). Overall, these examples indicated that IS insertions may produce different RpiRc protein expression outcomes depending on the type of IS, the insertion site, and the orientation of the IS.

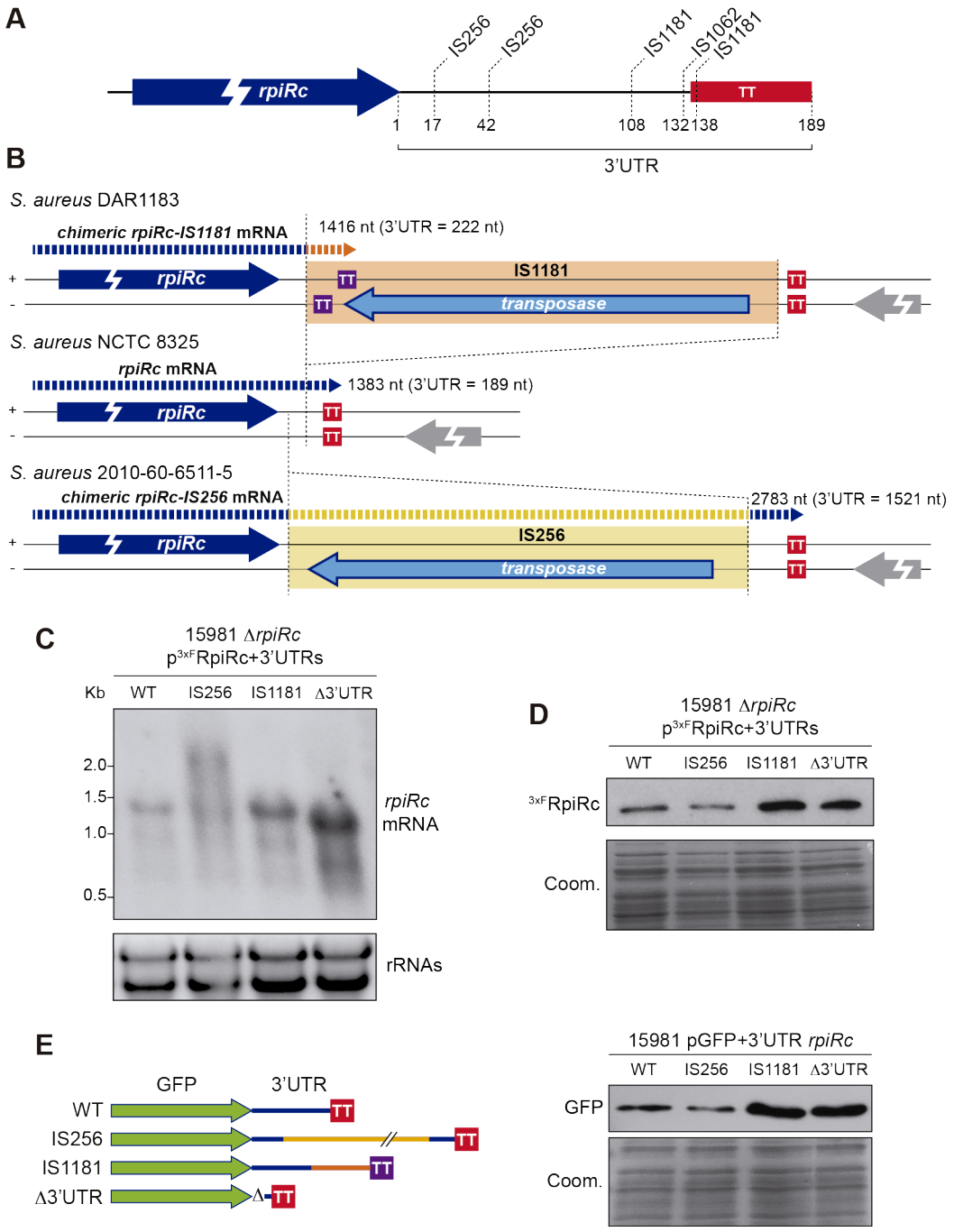


Figure 33. Disruption of the *rpiRc* 3'UTR sequence by ISs affects the RpiRc expression. (A) Schema showing the main IS insertion sites in the *rpiRc* 3'UTR of *S. aureus* pangenome. **(B)** Schema representing the putative chimeric mRNAs generated by IS1181 and IS256 insertions in the *rpiRc* 3'UTR of the *S. aureus* DAR1183 and 2010-60-6511-5 strains, respectively. **(C-D)** Northern and Western blots showing the mRNA or protein levels expressed from the WT *rpiRc* mRNA or carrying IS1181 and IS256 insertions or 3'UTR deletion, respectively. **(E)** Western blot showing the GFP levels from constructs carrying the *gfp* fused to the WT *rpiRc* 3'UTR, the 3'UTR+IS256, the 3'UTR+IS1181 and the *rpiRc* TT. Stained protein and RNA gel portions are shown as loading controls. Representative results from at least three independent experiments are shown. TT, transcriptional terminator; +/-, DNA strands; Coom, Coomassie.

To further confirm that the *rpiRc* 3'UTR had a functional role that could be affected by IS insertions, we analyzed the effect of the WT 3'UTR and the chimeric 3'UTR-ISs on the expression of a heterologous gene, the green fluorescent protein. To this end, we fused the *rpiRc* 3'UTR WT, the 3'UTR-IS1181, and 3'UTR-IS256 sequences downstream of the *gfp* gene. As a control, we included the Δ3'UTR plasmid that carried only the *rpiRc* TT fused downstream of *gfp* (Figure 33E). Western blot analyses showed that the *rpiRc* 3'UTR led to a lower GFP expression when compared to the Δ3'UTR version (Figure 33E). This indicated that the *rpiRc* 3'UTR could reduce the expression of a heterologous gene and, therefore, act as a functional module. However, the presence of IS1181 in the *rpiRc* 3'UTR increased GFP expression while IS256 only slightly decreased it. This confirmed that the transposition of ISs in the *rpiRc* 3'UTR could have different consequences on protein expression depending on the insertion events (Figure 33E).

Next, we decided to analyze the relevance of the *rpiRc* 3'UTR on the *S. aureus* physiology and the implications of the IS insertions in its role. Previous studies showed that RpiRc repressed the expression of RNAIII. RNAIII is the master regulator of *S. aureus* virulence, which activates many virulence factors, including hemolysins. A mutant lacking the *rpiRc* gene exhibited an increased hemolysis capacity due to the lack of RNAIII repression (Zhu *et al.*, 2011; Balasubramanian *et al.*, 2016; Gaupp *et al.*, 2016) (Figure 34A). To test whether alterations in the *rpiRc* 3'UTR could modify the hemolytic activity of *S. aureus*, we constructed chromosomal

mutants in *S. aureus* MW2, a community-acquired *agr*-positive methicillin-resistant (MRSA) strain. On the one hand, we performed a 115-bp chromosomal deletion that removed most of the 3'UTR of *rpiRc* while preserving its intrinsic TT (MW2 Δ3'UTR), as we did for the aforementioned construct. On the other hand, we inserted IS1181 in the same *rpiRc* region as it is naturally found in *S. aureus* DAR1183 (MW2 3'UTR+IS1181). As a control, we deleted the whole *rpiRc* gene (MW2 Δ*rpiRc*). These strains were grown overnight and the production of hemolysins in the culture supernatants was tested in sheep-blood agar plates (Figure 34B).

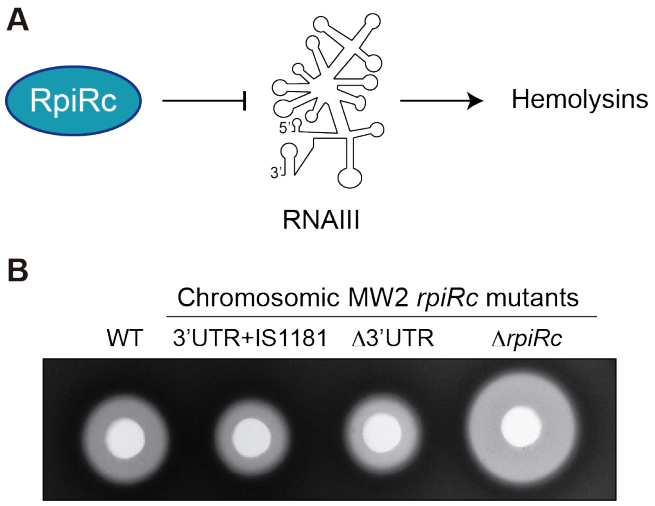


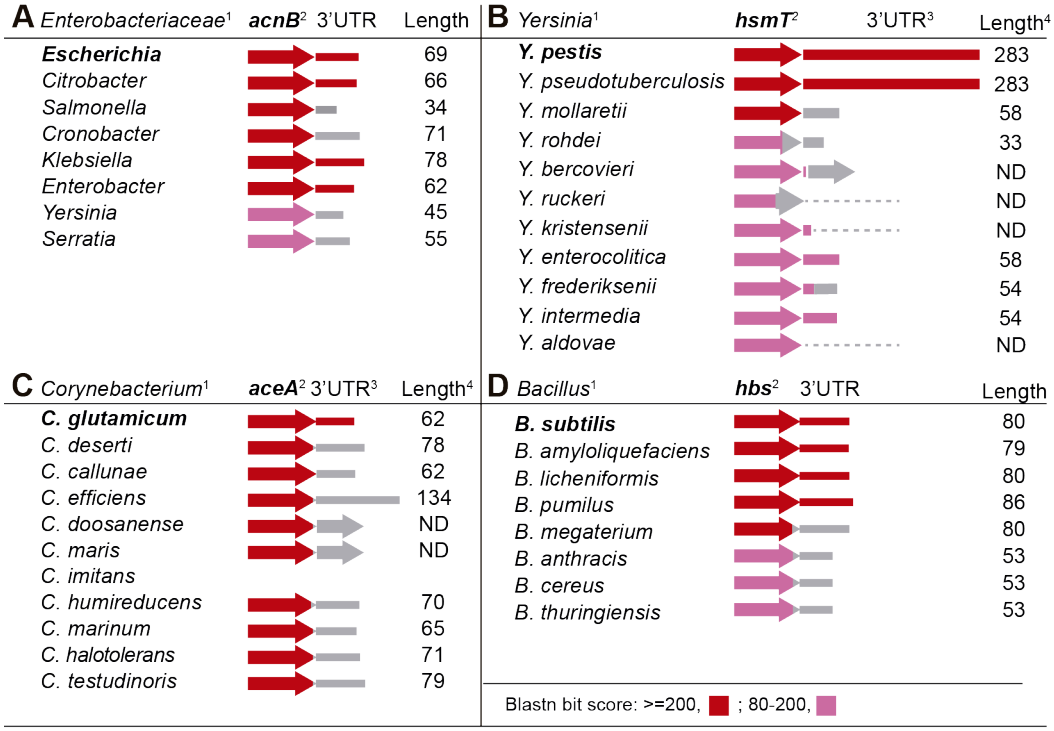
Figure 34. Deletion of *rpiRc* 3'UTR or insertion of IS1181 in this region affects hemolysin production in *S. aureus*. (A) Schema representing the RpiRc repression of RNAIII. **(B)** Hemolytic halos produced by the hemolysins contained in the supernatant of cultures of *S. aureus* MW2 wild-type (WT) and their isogenic chromosomic mutants Δ3'UTR, Δ*rpiRc* and 3'UTR+IS1181. The supernatants were concentrated 10 times and loaded into 5-mm holes made in Columbia Sheep blood (5%) agar plates. This figure shows representative results from at least three independent experiments.

As expected, the Δ*rpiRc* mutant showed an increase in hemolysis when compared to the WT strain. In contrast, the deletion of the *rpiRc* 3'UTR and the IS1181 insertion in the 3'UTR, which produced higher protein levels of the RpiRc repressor (Figure 33D) resulted in lower levels of hemolysis (Figure 34B). Overall, these results highlighted the relevance of the *rpiRc* 3'UTR in the physiology of *S. aureus* while demonstrating that the RpiRc function could be altered by the insertion of IS1181 in the *rpiRc* 3'UTR.

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

Having proved that the inter-species 3'UTR variability had consequences on *S. aureus* gene expression, we decided to extend the 3'UTR conservation analysis to other bacterial genera. First, we focused on several previously described functional 3'UTRs (Maeda and Wachi, 2012; Benjamin and Massé, 2014; Zhu *et al.*, 2016; Braun *et al.*, 2017). We compared the last 200 nt of the CDSs plus the entire 3'UTR sequences from the *E. coli acnB*, *Y. pestis hmsT, C. glutamicum aceA*, and *B. subtilis hbs* mRNAs with those of their corresponding phylogenetically-related species (Annexes 8-11). Although the CDSs were highly conserved across all closely-related species, there were 3'UTR sequence variations in all four gene examples (Figure 35). The *E. coli acnB* 3'UTR was conserved in the *Citrobacter*, *Enterobacter*, and *Klebsiella* species with only a few nucleotide variations. In contrast, and according to synteny analysis by SyntTax (Oberto, 2013), the *Salmonella*, *Cronobacter*, and *Serratia acnB*

3'UTRs sequences varied due to gene rearrangements (Annex 8B). For example, in *S. enterica* serovar Typhimurium SL1344, an insertion of the SL1344_0160 gene (which encodes a putative HNH restriction endonuclease) deleted the sequence downstream of the *acnB* CDS and included a different TT. As a result, the *Salmonella acnB* 3'UTR was only 34 nt long and it lacked the stem-loop recognized by the apo-AcnB protein (Benjamin and Massé, 2014).



[1] Representative bacterial species according to the phylogenetic tree built at TimeTree knowledge-base (www.timetree.org).
[2] Color code indicates blastn bit score. Gray lines represent the 3'UTR length in the correspoding species, which is non-conserved.
[3] Dashed line indicates absence of a putative TT, gray arrow indicates the presence of a CDS instead of a 3'UTR.
[4] ND, 3'UTR length not determined.

Figure 35. Schematic representation of the conservation analysis of regulatory 3'UTRs among phylogenetically-related species. Blastn analyses were performed using the *E. coli acnB* **(A)**, *Y. pestis hmsT* **(B)**, *C. glutamicum aceA* **(C)**, and *B. subtilis hbs* **(D)** mRNAs as queries, which carried 3'UTRs with proven regulatory capacities (Maeda and Wachi, 2012; Benjamin and Massé, 2014; Zhu *et al.*, 2016; Braun *et al.*, 2017).

Sequence variations were also present in the *C. glutamicum aceA*, *B. subtilis hbs*, and *Yersinia hmsT* 3'UTRs. Although the *hbs* and *hmsT* CDSs are widely distributed across the *Bacillus* and *Yersinia* genera, their 3'UTRs were only conserved in the closest species: *B. amyloliquefaciens*, *B. licheniformis*, and *B. pumilus* in the former and *Y. pestis* biovars and *Y. pseudotuberculosis* in the latter (Figure 35). Finally, none of the 10 compared species of the genus *Corynebacterium* carrying the *aceA* CDS presented a similar sequence to that of the *C. glutamicum aceA* 3'UTR (Figure 35). We carried out the prediction of Rho-independent transcriptional terminator structures in the non-conserved 3'UTRs from *Escherichia*, *Corynebacterium*, and *Bacillus* species to determine the length variability (Figure 35, Annexes 8-11). Most of these 3'UTRs variations were explained by gene rearrangements occurring downstream of CDSs, with the exception of the *hbs* 3'UTR, in which variations may have occurred locally (Annexes 8-11).

Finally, to further investigate whether 3'UTR variations could also be distributed among other genes of *E. coli* and *B. subtilis*, as observed in *S. aureus*, we performed similar genome-wide 3'UTR conservation analyses. Since the 3'-end transcript boundaries were recently released for *E. coli* and *B. subtilis* (Dar *et al.*, 2016; Dar and Sorek, 2018), we generated a query database including the sequences of all mapped 3'UTRs plus the last 200 nt of their corresponding CDSs. Using blastn, we compared the *E. coli* and *B. subtilis* databases against representative genomes of species from *Enterobacteriaceae* and *Bacillus* genus, respectively (Figure 36).

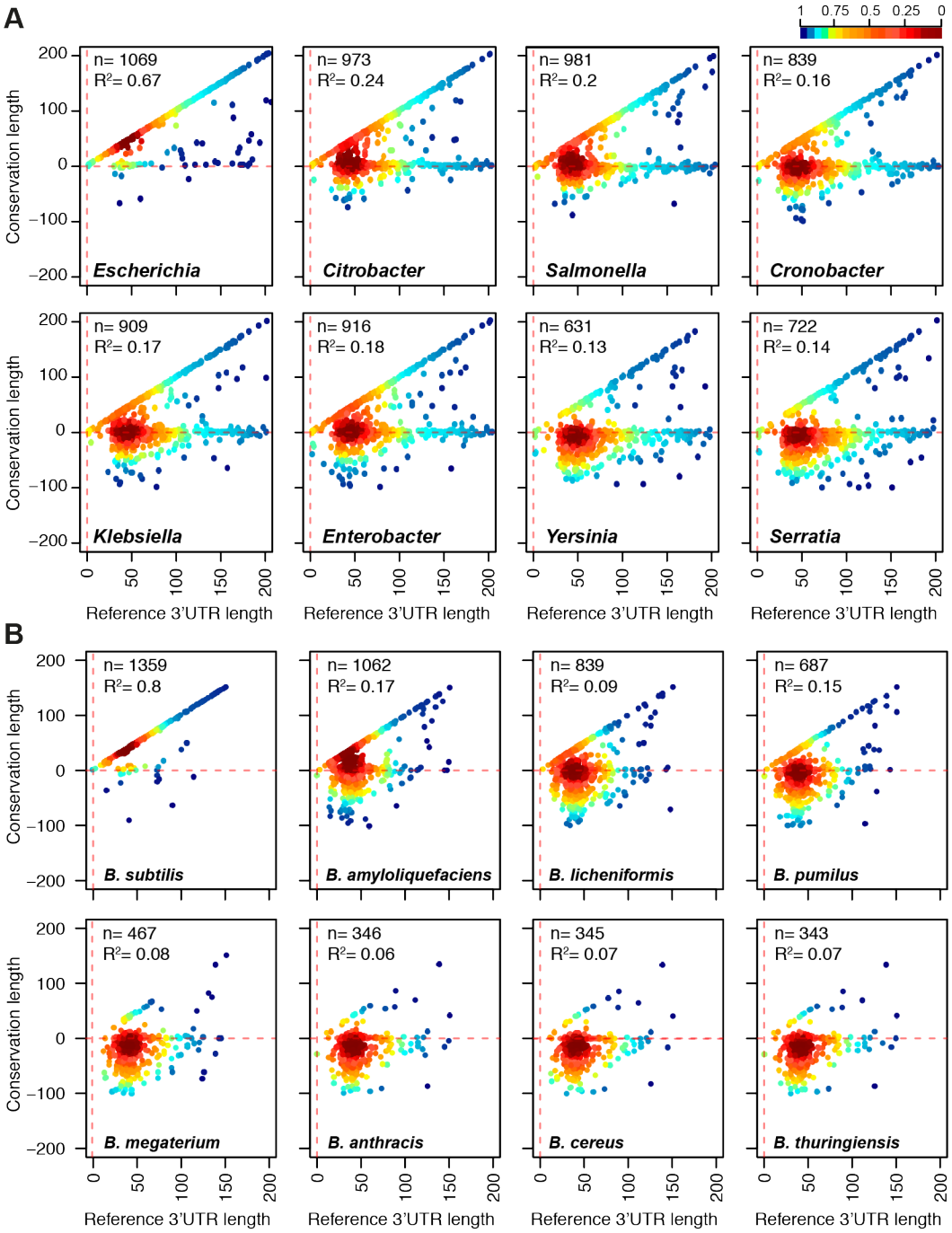


Figure 36. 3'UTR sequence variations are widely distributed in bacteria. Scatter plots representing the conservation of the 3' end regions of *E. coli* BW25113 **(A)** and *B. subtilis 168* **(B)** mRNAs compared to their corresponding 3' end regions in phylogenetically-related species**:** *E. coli* O157-H7 str EC4115, *Citrobacter koseri* ATCC BAA-895, *Salmonella typhimurium* SL1344, *Enterobacter aerogenes* KCTC 2190, *Klebsiella pneumoniae* DMC1097, *Cronobacter sakazakii* ATCC BAA-894, *Serratia marcescens* Db11 and *Yersinia pestis* CO92 strains; *B. subtilis* OH 131.1, *B. amyloliquefaciens* DSM7, *B. licheniformis* ATCC 14580, *B. pumilus* SH-B9, *B. megaterium* NBRC 15308, *B. anthracis* Ames, *B. cereus* ATCC 14579 and *B. thuringiensis* serovar konkukian strains. Data are plotted as indicated in Figure 19.

We found a similar outcome to the one observed for the genus *Staphylococcus*, meaning that 3'UTR sequences were also variable among mRNAs encoding orthologous proteins (Figure 36). The bacterial species of the *Enterobacteriaceae* family showed a higher number of conserved 3'UTRs in comparison with the *Bacillus* species. This was probably due to the phylogenetic distances between the species in the former being shorter than the ones in the latter. However, both bacterial groups presented a very high variation rate of their 3'UTR sequences, as illustrated in Figure 36, where most of the dots fell onto the horizontal line, which represented the protein stop codon. This indicated that nucleotide conservation was also lost downstream of most CDSs, as represented in Figure 37. On the one hand, these comparative analyses confirmed that the nucleotide conservation analysis might only be useful for detecting a small percentage of inter-species conserved 3'UTRs. On the other hand, the results indicated that the 3'UTRs from orthologous genes underwent different evolutionary events, a phenomenon that seems to be widespread in bacteria.

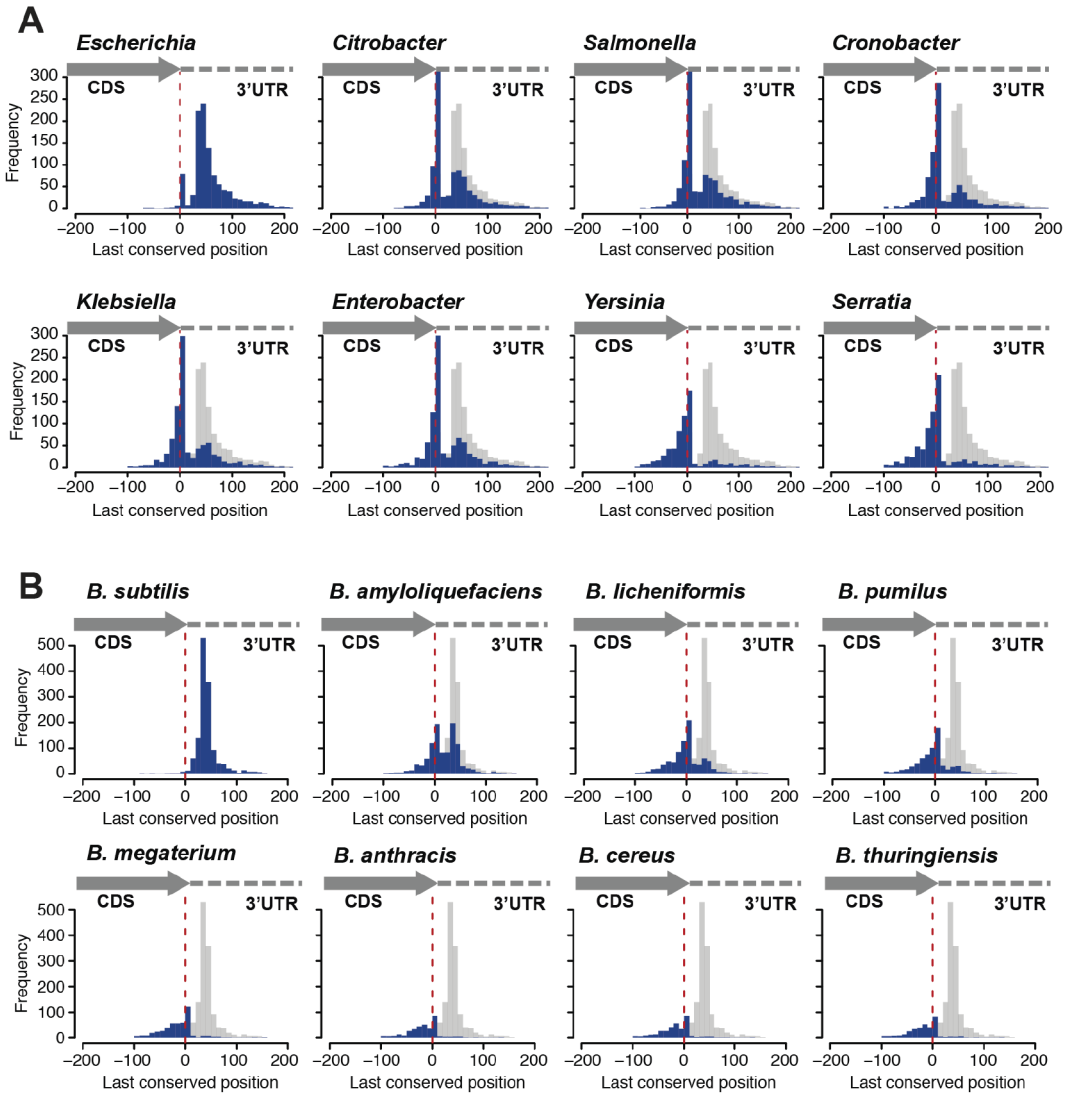


Figure 37. Distribution of the 3' end conservation lengths of *E. coli* (A) and *B. subtilis* (B) mRNAs compared to their corresponding 3' end regions in phylogenetically-related species. Each blue bar represents the number of conserved 3'UTRs at a given position in windows of 10 nt. Grey bars represent the distribution of *E. coli* and *B. subtilis* that are included as a reference, respectively. The dashed red line indicates the position of the stop codon.

DISCUSSION

DISCUSSION

Bacterial IGRs, which have been traditionally defined as the genomic regions between two CDSs, are known for being poorly conserved among phylogenetically-related bacteria (Molina and van Nimwegen, 2008; Luo *et al.*, 2011; Sridhar *et al.*, 2011; Thorpe *et al.*, 2017). For this reason, they have often been regarded as "junk" genetic material lacking relevant biological functions. However, recent RNA-seq technologies revealed that IGRs may be more complex than initially anticipated, accommodating a plethora of regulatory elements that are either transcribed independently as ncRNAs or fully functional from within the UTRs of mRNAs. In recent years, several studies showed the biological relevance of 3'UTRs in both bacteria and eukaryotic organisms (Miyakoshi, Chao, and Vogel, 2015a; Ren *et al.*, 2017; Mayr, 2017). Since the 3'UTR-regulatory elements are physically- and functionally-related to the adjacent CDSs through mRNAs, an evolutionary connection between them might be expected in phylogenetically-related species. This indicates that the 3'UTR sequences would be preserved to the same degree as their corresponding CDSs. However, in this Thesis, we observed this to be true for just a few examples, since most of the 3'UTRs from mRNAs encoding orthologous proteins, in closely-related bacterial species, displayed different lengths and sequences.

Conserved 3'UTRs in the genus *Staphylococcus*

Among the few mRNAs that carried conserved 3'UTRs in the genus *Staphylococcus* RNAIII stood out (Table 2). This RNA molecule was one of the first regulatory RNAs described in the literature (Novick *et al.*, 1993). Since it also encodes the δ–hemolysin (*hld*), it is the paradigm for dual-functional mRNAs. RNAIII is the master regulator of *S. aureus* virulence targeting several mRNAs encoding Protein A, Sbi, LytM, Coa, Rot, MgrA and Hla. RNAIII pairs these targets mainly through its 3'UTR to either inhibit translation initiation and/or promote RNase III-mediated degradation or promote mRNA stability and/or translation initiation (Novick *et al.*, 1993; Huntzinger *et al.*, 2005; Boisset *et al.*, 2007; Chevalier *et al.*, 2010; Chabelskaya *et al.*, 2014; Gupta *et al.*, 2015; Bronesky *et al.*, 2016). RNAIII is also able to regulate its own translation through a 5'UTR/3'UTR interaction but the molecular mechanism is unknown (Balaban and Novick, 1995).

Since the RNAIII 3'UTR, which is the main regulatory region, is highly conserved, it might be expected that the RNAIII-regulated targets would be also conserved among closely-related staphylococcal species. However, analyzing the principal targets we found that only *mgrA* is present in other *Staphylococcus* species. The rest of the targets were only present in a few species. For example, *rot* and SAOUHSC_01100 were found in *S. argenteus, S epidermidis* and *S. haemolyticus*, whereas *spa and sbi* were present in *S. argenteus* and *coa* and *hla* in *S. argenteus* and *S. haemolyticus*. Considering that the Agr system is conserved in all

staphylococcal species analyzed, it suggested that the RNAIII-regulatory mechanism is preserved over the targets, which might vary according to the cell niche. This phenomenon has been observed in eukaryotes with some transcriptional factors that are highly conserved but regulate different genes depending on the cell tissue or the organism (Mercurio *et al.*, 2019; Trefflich *et al.*, 2019). Recently, Connolly *et al.* observed for different pathogenic *E. coli*, that the conserved transcriptional factor YhaJ had different targets in each pathotype. They believe that each organism is capable of recycling conserved regulators to create tailor-made regulatory circuits that will be beneficial in a particular lifestyle, a situation that could be occurring with the RNAIII-mediated regulation (Connolly *et al.*, 2019).

Regarding other conserved 3'UTRs in *Staphylococcaceae*, it is noteworthy that several of them corresponded to riboswitch-dependent 3'UTRs (Toledo-Arana *et al.*, 2009; Ruiz de los Mozos *et al.*, 2013). Riboswitches are regulatory elements that are usually located in the 5'UTRs to control the expression of the downstream genes. However, when the gene located upstream of the riboswitch lacks a transcriptional terminator, the transcription of this gene will terminate at the transcriptional terminator generated by an OFF-riboswitch configuration. As a result, a 3'UTR containing the riboswitch is generated (Figure 38). In contrast, if an ON-riboswitch configuration exists, a long transcript including the upstream and downstream genes relative to the riboswitch will be created (Figure 38). This transcriptomic configuration was initially described in *L.*

monocytogenes but it is present in several bacterial species (Toledo-Arana *et al.*, 2009; Ruiz de los Mozos *et al.*, 2013). The fact that the roles of the upstream genes and the riboswitches are related, could explain why these intergenic regions are well conserved among related species. It has not yet been elucidated the consequences that this transcriptomic architecture can cause on the expression and function of the genes involved. Interestingly, most of the riboswitch-dependent 3'UTRs comprised T-box riboswitches, which binds specific tRNAs. These riboswitches normally control the expression of enzymes related to the amino acid metabolism or tRNA synthesis in relation to specific amino acid availability (Mellin and Cossart, 2015; Sherwood and Henkin, 2016).

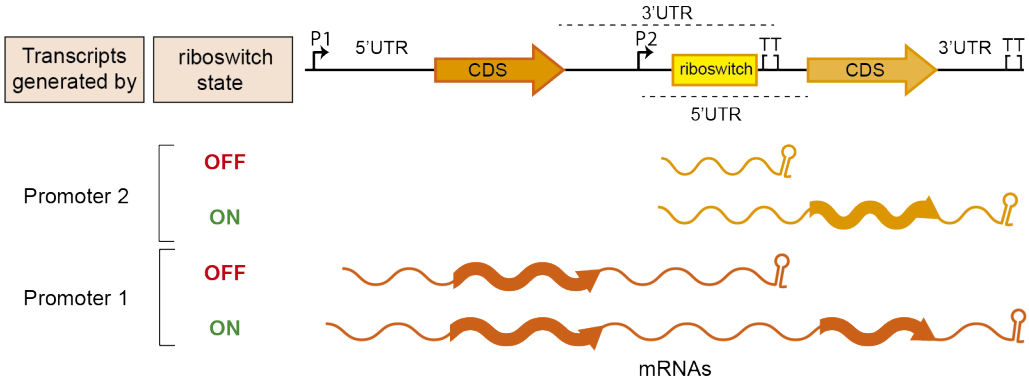


Figure 38. **Riboswitch-dependent 3'UTRs**. Depending on the configuration of the riboswitch (ON/OFF) and the promoters' activation up to 4 different mRNAs can be transcribed from this chromosomal organization. P, promoter; TT, transcriptional terminator.

It is also noticeable that at least four of the conserved 3'UTRs are able to produce type I non-coding RNAs (Table 2). For instance, the 3'UTR of the *pepF* (SAOUHSC_00937) gene generates the RsaE and RsaF ncRNAs.

RsaE is conserved in bacteria from the order *Bacillales* and participates in the regulation of the central carbon flux and amino acid metabolism (Marincola *et al.*, 2019). There are also evidences of species-specific regulatory nodes that might be linked to the bacterial niche. In *S. epidermidis* RsaE promotes biofilm formation by alleviating repression of the *ica* genes through binding to the *icaR* 5'UTR and by promoting the release of extracellular DNA. Interestingly, RsaE in *S. epidermidis* is heterogeneously expressed within the bacterial cells, suggesting that this heterogeneity might favor the creation of bacterial populations with distinct roles within the biofilm community (Schoenfelder *et al.*, 2019).

3'UTR variability is widespread among bacteria

Despite the few mRNAs carrying conserved 3'UTRs, the vast majority of mRNAs presented a huge variation in their 3'UTR sequence when compared among their close-relatives. The idea of synthetizing hundreds of nucleotides at the 3' ends without a relevant function seems contradictory as bacteria tend to efficiently manage energy and clones with a lower fitness are rapidly eliminated from populations. Therefore, an explanation for the relatively high 3'UTR variation rate should be sought. First, it is important to highlight that 3'UTR sequences are more highly conserved in comparison with non-transcribed regions of strains from the same bacterial species (Ruiz de los Mozos *et al.*, 2013). This implies that an evolutionary pressure to preserve 3'UTR sequences in a particular species exists. In addition, our data indicates that specific inter-species variations in 3'UTR sequences often occur. Therefore, these variations

could differentially affect the expression of orthologous proteins from different species. This could be a strategy used by bacteria to create diversity without changing the CDS, an idea supported by previous studies that show how bacterial genes can shift rapidly between multiple regulatory nodes. Oren *et al.* demonstrated the existence of promoter-regulatory regions that changed among orthologous genes, contributing to expression divergence and conferring distinct fitness advantages (Oren *et al.*, 2014). Analogously, the finding that 3'UTR sequence variations can affect protein expression depending on the species may be an evolutionarily-selected trait to create species-specific post-transcriptional regulatory processes (Miyakoshi, Chao, and Vogel, 2015a; Updegrove *et al.*, 2015; Miyakoshi *et al.*, 2019). Since the 5'UTRs are closely related with the initiation of translation, any significant sequence variations could impair protein synthesis and, thus, explain the bias in variability within 3'UTRs. This hypothesis is in agreement with the role of 3'UTRs in eukaryotes, which have been specifically targeted through the course of evolution to contribute to the divergence of species by accumulating regulatory elements in their sequences (Mazumder *et al.*, 2003; Mayr and Bartel, 2009; Mayr, 2017). It is worth noting that longer 3'UTRs (including multiple miRNA target sites) are preferred in functionally critical eukaryotic genes that are spatially or temporally expressed. In contrast, housekeeping genes are selected to have shorter 3'UTRs (Cheng *et al.*, 2009). Strikingly, the 3'UTRs of *Salmonella hilD* and *S. aureus icaR* mRNAs (which encode the main transcriptional regulators of the SPI-1

pathogenicity island and biofilm formation, respectively) are among the longest 3'UTRs found in bacteria.

Resembling the eukaryotic process in which mRNAs are targeted by miRNAs, it has been recently shown that *hilD* and *icaR* 3'UTRs are targeted by ncRNAs (El-Mouali *et al.*, 2018; Bronesky *et al.*, 2019). In the former, the Spot 42 ncRNA (transcriptionally repressed by CRP-cAMP) targets the *hilD* 3'UTR, exerting a positive effect on HilD expression (El-Mouali *et al.*, 2018; El-Mouali and Balsalobre, 2018). In the latter, the RsaI ncRNA modulates PIA-PNAG synthesis by interacting with the 3'UTR of the *icaR* mRNA (Bronesky *et al.*, 2019). The RsaI interacting region is located between the UCCCC motif, which is required for the *icaR* 3'-5'UTR interaction, and the transcriptional terminator of the *icaR* mRNA (Bronesky *et al.*, 2019). Interestingly, while the UCCCC motif is conserved across the *S. aureus, S. argenteus*, and *S. simiae icaR* mRNAs, the RsaI paring region is only present in the *S. aureus icaR* mRNA. Note that the RsaI ncRNA is conserved in most species of the genus *Staphylococcus*. The Spot 42 ncRNA is present in several bacterial genera, while the *hilD* gene is exclusive to *Salmonella* (Bækkedal and Haugen, 2015). Overall, this shows that the 3'UTRs can accumulate more than one regulatory motif, which may be differentially selected during evolution.

Causes of 3'UTR sequence variability

We have observed that 3'UTR variability can be caused by different phenomena. For instance, the acquisition of *icaR* 3'UTR variations may be

explained by gene rearrangements that created mRNA chimeras among staphylococcal species. The *S. aureus icaR* CDS is only present in 5 out of 9 staphylococcal species that encode the *icaADBC* operon. It is interesting to note that in the four species lacking the *icaR* CDS, either DspB (hexosaminidase), a TetR-like regulator, or proteins of a two-component system are encoded in place of the *icaR* gene. These different genomic organizations indicate that the regulatory locus evolved independently from the *icaADBC* operon and may also explain the appearance of 3'UTR variations in the *icaR* mRNA when such a locus was acquired or mobilized. Similarly, gene rearrangements could also be responsible for the disparity of ncRNA content in phylogenetically-related bacterial species and may also justify some of the 3'UTR variations (Raghavan *et al.*, 2015).

Furthermore, there may be other phenomena beyond gene rearrangements behind 3'UTR variations, since only a minor percentage of them (around 28.3-35.2%) fell under this category (Figure 22). Between 58-62% of 3'UTR variations were originated from local variations inside IGRs. Several mechanisms have been described that could explain these variations. For instance, SNPs and small nucleotide variations can be caused by errors in the replication machinery or by exposure to a mutagenic agent such as UV light. In this last case, the SOS response of each species is key in determining if these nucleotide variations are or not corrected (Arber, 2000). Bigger changes can be originated by repeated sequences that apart from promoting gene rearrangements due to

homologous recombination, they can cause duplications and deletions of certain regions by replication machinery errors (Hardy *et al.*, 2004; Delihas, 2011). Additionally, mobile elements such as ISs can disrupt intergenic regions when a transposition event occurs. Even when ISs are no longer present in the 3'UTR, the excision process in some ISs still alters the original sequence by producing an excision mark of 8-12 nt duplication (Siguier *et al.*, 2014). If this phenomenon happens during a long period of time the IGR sequence could significantly change.

We hypothesized that mobile insertion sequences and/or repeated sequences could be an evolutionary motor changing the functionality of the 3'UTRs of orthologous mRNAs (Hardy *et al.*, 2004; Purves *et al.*, 2012). Interestingly, we showed that the presence of STAR elements and ISs that interrupt 3'UTR sequences affect the expression of RpiRc in *S. aureus* (Figure 33). This suggests that alternative IS transposition/excision processes may generate local 3'UTR variations that can be evolutionarily selected if they produce an advantage for a specific clone in comparison to the rest of the bacterial population.

Alternatively, ISs might be also responsible for phase variation mechanisms (Ziebuhr *et al.*, 1999; Vandecraen *et al.*, 2017). Here, we showed that the presence of IS1181 in the *rpiRc* 3'UTR of several *S. aureus* strains reduced hemolysin production by increasing the levels of the RNAIII-repressor RpiRc (Figure 34). Further studies would be required to determine whether the transposition of ISs into the *rpiRc* 3'UTR provide

an additional way to regulate *S. aureus* virulence through the *agr* locus, as seen for the cases of *agr* phase variation (Gor *et al.*, 2019).

Although ISs are able to disrupt 3'UTR sequences, the consequences in protein expression might be difficult to predict. Here, we observed that the insertion of IS1181 or IS256 in the *rpiR*c 3'UTR produced different outputs. On the one hand, the IS1181 contained a hairpin structure that acted as a transcriptional terminator, creating a chimeric *rpiRc/IS1181* mRNA of similar size to the *rpiRc* WT mRNA. The insertion increased both mRNA and protein levels affecting hemolysin production. On the other hand, IS256 insertion produced a longer chimeric *rpiRc/IS256* that did not show any significant effect on RpiRc expression, at least under the tested conditions. Therefore, the consequences of an IS insertion into a 3'UTR cannot be predicted. Several variables should be taken into account such as the type of IS, the insertion point and the orientation of the transposase. Nevertheless, our results demonstrated that 3'UTR function could be affected by ISs. It would be interesting to create genomic maps compiling the entire set of specific IS sites throughout the genome in a specific species. This information could give insights on how ISs might participate in the creation of 3'UTR variability.

Alternative 3'UTR-mediated regulatory mechanisms for the same orthologous gene

In this Thesis, we demonstrated that 3'UTRs vary among bacteria, affecting protein expression. Variability in 3'UTRs can cause the emergence or vanishing of regulatory elements creating different

regulation patterns for the same orthologous proteins in closely-related bacteria. These changes can affect the sequences required for interactions with 5'UTRs, ncRNAs or RNA-binding proteins. This is exemplified by the *icaR* 3'UTR. The UCCCC motif needed for the 5'UTR interaction is only present in *S. argenteus* and *S simiae* (Figure 24). Interestingly, *S. epidermidis* has developed an alternative mechanism to regulate IcaR protein levels instead of the 5'UTR-3'UTR interaction. The ncRNA IcaZ that is located just downstream of the *icaR* mRNA base pairs with *icaR* 5'UTR to inhibit translation initiation (Lerch *et al.*, 2019). Similarly, the presence of specific target sites for RNA-binding proteins in 3'UTRs could also contribute to regulatory differences among bacterial species (Benjamin and Massé, 2014). Our comparative analysis revealed that gene rearrangements have eliminated the Apo-AcnB binding site from *acnB* 3'UTR from *Salmonella, Cronobacter* and *Serratia.* Although further experiments are needed, it might be speculated that *acnB* regulation would be different from the one found in Escherichia, *Citrobacter, Enterobacter* and *Klebsiella* (Figure 35A and Annex 8).

Furthermore, we have shown that changes in the 3'UTR sequence could affect mRNA turnover by modifying the 3'UTR accessibility to exoRNAses. This indicates that the accumulation/elimination of ribonuclease target sites could be an alternative way to specifically modulate the protein expression at the post-transcriptional level in a particular species. In agreement with this, the deletion of 3'UTRs (e.g. *hilD*, *aceA*, *hmsT*, and *icaR*), and with them, the putative ribonuclease target sites, increases

protein expression (Maeda and Wachi, 2012; Ruiz de los Mozos *et al.*, 2013; López-Garrido *et al.*, 2014; Zhu *et al.*, 2016; Zhao *et al.*, 2018). This could also be the case of the mechanism mediated by the *ftnA* 3'UTR in *S. aureus*, which is targeted by RNase III and PNPase (Figure 30). Note that the putative target region that we identified is only shared by *S. aureus* and *S. argenteus.* This suggested that the 3'UTR mediated post-transcriptional regulation of FtnA expression would be different in the other staphylococcal species.

In addition to a differential effect on orthologous genes, it should also be considered that variations in 3'UTR sequences might affect the expression of downstream genes. For example, if variations in the 3'UTR sequence alter the strength of the corresponding TT or even eliminate it (e.g. Figure 26C shows that transcriptional read-through events occur in *ftnA* of some staphylococcal species), transcription would continue to increase the expression of downstream genes or create antisense RNAs if the downstream gene is encoded in the opposite direction (Lasa *et al.*, 2012). If a downstream gene is close enough, variations in the 3'UTR sequence could also alter promoter regulatory elements of the next transcription unit, thereby affecting their expression. In the case of convergent genes, variations in the 3'UTRs can create overlapping 3'UTRs that can produce transcriptional interference due to collision between the RNA synthesis machineries and/or antisense regions targeted by RNase III. This could generate a reciprocal expression regulation of both genes depending on the promoters activation (Sesto *et al.*, 2013).

Further studies are needed to improve our knowledge of how sequence variations in 3'UTRs may specifically affect gene expression in different bacterial species. The development of RNA sequencing techniques, mapping the precise 3'-ends of transcripts, as well as ribonuclease cleavage regions and RNA-protein binding sites will considerably contribute to this (Dar *et al.*, 2016; Chao *et al.*, 2017; Holmqvist and Vogel, 2018; Altuvia *et al.*, 2018). Combining such information with studies like the one presented here will help improving our understanding of the evolutionary features responsible for the generation of species-specific sequences that may, ultimately, lead to bacterial species differentiation.

CONCLUSIONS

CONCLUSIONS

1. Genome-wide comparative analyses of mRNA sequences encoding orthologous proteins revealed that only few mRNAs contained conserved 3'UTRs, including RNAIII and mRNAs carrying riboswitch-dependent 3'UTRs or type I ncRNAs.

2. In contrast, most of the 3'UTRs were not conserved among phylogenetically-related species of the genus *Staphylococcus*. The mRNA sequence conservation was lost around the stop codon.

3. Additional comparative analyses showed that 5'UTRs were more conserved than 3'UTRs in *Staphylococcus,* supporting the idea of 3'UTRs being more prone to sequence variations than 5'UTRs.

4. RNA sequencing analyses of diverse *Staphylococcus* species revealed that most of the mRNAs encoding orthologous proteins carried 3'UTRs with different lengths in addition to sequence variation.

5. Sequence conservation analyses of the genomic regions downstream of the orthologous CDSs, including IGRs and adjacent CDSs, revealed that differences in the 3'UTR sequences occurred partly due to gene rearrangements but mostly because of local variations in the IGR sequences.

6. The construction of chimeric mRNAs carrying the 3'UTR of the orthologous *icaR, ftnA* and *rpiRc* genes from five staphylococcal species demonstrated that 3'UTR sequence variations produced changes in both mRNA and protein levels. This indicated that 3'UTRs might provide distinct

functional roles to differentially control the expression of the same orthologous protein in closely-related species.

7. The IcaR protein is the main repressor of the polysaccharidic biofilm in *S. aureus*. The expression of IcaR is modulated by the interaction of the 5' and 3'UTRs from the same mRNA molecule through base pairing of the RBS sequence and a UCCCC motif. The latter is present in the 3'UTR of *S. aureus*, *S. argenteus* and *S. simiae,* but absent in *S. epidermidis* and *S. capitis*. Only the former species could repress IcaR translation through the 3'UTR and, in return, modulate the polysaccharide PIA-PNAG synthesis.

8. The *ftnA* mRNA encodes the ferritin protein required for iron storage. The *ftnA* 3'UTR of *S. aureus* and *S. argenteus* carries a putative regulatory region located between nucleotides 20 and 56 after the stop codon, which is absent in other staphylococcal species. This region is targeted by RNase III and PNPase to decrease mRNA levels and consequently FtnA expression.

9. The chromosomal deletion of the *ftnA* 3'UTR inhibited *S. aureus* growth when iron was scarce, indicating that the 3'UTR-mediated regulation is essential for the control of FtnA levels under iron deprivation conditions.

10. The finding of specific regulatory elements in *icaR* and *ftnA* mRNAs demonstrated that sequence variations around the 3'UTR determine the presence or absence of alternative regulatory motifs and, ultimately, the generation of functional distinctions among closely-related species.

11. The RpiRc protein is a transcriptional regulator repressing the synthesis of RNAIII, the main post-transcriptional regulator of several virulence factors including hemolysins. The chromosomal deletion of the *rpiRc* 3'UTR increased RpiRc expression and, consequently, reduced hemolysin production. This and the other examples presented in this Thesis highlight the importance of 3'UTR-mediated regulation to tightly modulate relevant biological processes such as biofilm formation, iron metabolism and virulence in a clinical important pathogen such as *S. aureus.*

12. Several *S. aureus* strains carried different insertion sequences (IS1181, IS256, IS1082, etc) within the *rpiRc* 3'UTR. The consequences of these insertions depended on their location, orientation and type of IS element. Specifically, the IS1181 insertion created a chimeric *rpiRc* mRNA that disrupted the *rpiRc* 3'UTR function by increasing RpiRc expression and reducing hemolysin production. This demonstrated that the 3'UTR-mediated regulatory mechanisms can be affected by transposition events.

13. The conservation analyses of several previously described functional 3'UTRs, as well as the entire set of *Escherichia coli* and *Bacillus subtilis* mRNAs, suggested that the 3'UTR sequence variability is widespread in bacteria.

14. 3'UTR variability might be responsible for creating different functional regulatory roles and, ultimately, bacterial diversity through the course of evolution, in a similar way to what happened during the diversification of eukaryotic species.

CONCLUSIONES

CONCLUSIONES

1. Los análisis comparativos globales de las secuencias de los mRNAs que codifican para proteínas ortólogas revelaron que sólo unos pocos mRNAs contenían 3'UTRs conservadas. Entre estos destacaban RNAIII y los mRNAs que incluían 3'UTRs dependientes de riboswitches o ncRNAs de tipo I.

2. Por el contrario, la gran mayoría de las 3'UTRs no estaban conservadas entre especies filogenéticamente relacionadas del género *Staphylococcus*. La conservación de las secuencias de los mRNAs se perdía al a la altura del último codón de la traducción.

3. Análisis comparativos adicionales mostraron que las 5'UTRs se encontraban más conservadas que las 3'UTRs en *Staphylococcus*. Estos resultados apoyan la idea de que las 3'UTRs son más propensas a variaciones nucleotídicas que las 5'UTRs.

4. Experimentos de secuenciación masiva de RNAs de diversas especies de *Staphylococcus* mostraron que la mayoría de las 3'UTRs de mRNAs que codifican para proteínas ortólogas no sólo variaban en su secuencia sino también en su longitud.

5. Los análisis de conservación de secuencia de las regiones genómicas posteriores a las CDS ortólogas, incluidas las IGRs y las CDSs adyacentes, revelaron que las variaciones nucleotídicas observadas en las 3'UTRs se produjeron principalmente debido a variaciones locales en las secuencias y en menor medida por reordenamientos genómicos.

6. La construcción de mRNAs quiméricos que contenían las CDSs de *icaR, ftnA y rpiRc* de *S. aureus* fusionadas a las 3'UTRs provenientes de cinco especies de *Staphylococcus* demostró que las variaciones nucleotídicas en las 3'UTRs producen cambios en los niveles de mRNAs y proteínas. Esto sugería que las 3'UTRs podrían proporcionar funciones distintas para controlar de manera diferencial la expresión de la misma proteína ortóloga en especies estrechamente relacionadas.

7. La proteína IcaR es el principal represor del biofilm polisacarídico en *S. aureus*. La expresión de IcaR está modulada por la interacción de la 5'UTR y la 3'UTR del mismo mRNA por medio del apareamiento de bases de la RBS con un motivo UCCCC que está presente en las 3'UTRs de *S. aureus, S. argenteus* y *S. simiae,* pero ausente en las de *S. epidermidis y S. capitis*. Únicamente las primeras especies eran capaces de reprimir la traducción de IcaR por medio de la 3'UTR y, así, controlar la síntesis del exopolisacárido PIA-PNAG.

8. El mRNA de *ftnA* codifica a la proteína ferritina necesaria para el almacenamiento de hierro. La 3'UTR de *ftnA* de *S. aureus* y *S. argenteus* contiene una región con una posible capacidad reguladora ubicada entre los nucleótidos 20 y 56 después del codón stop, la cual está ausente en otras especies de *Staphylococcus*. Esta región es procesada por la RNasa III y la PNPasa para disminuir los niveles de mRNA y, en consecuencia, la expresión de FtnA.

9. La deleción cromosómica de la 3'UTR de *ftnA* inhibió el crecimiento de *S. aureus* cuando el hierro era escaso en el medio de cultivo, lo que

indicaba que la regulación mediada por la 3'UTR era esencial para el control de los niveles de FtnA en condiciones de privación de hierro.

10. El descubrimiento de elementos reguladores específicos en los mRNAs de *icaR* y *ftnA* demostró que las variaciones de secuencia alrededor de las 3'UTRs determinan la presencia o ausencia de motivos alternativos con capacidad reguladora, y, en última instancia, la generación de distinciones funcionales entre especies estrechamente relacionadas.

11. La proteína RpiRc es un regulador transcripcional que reprime la síntesis de RNAIII, el principal regulador post-transcripcional de varios factores de virulencia, incluidas las hemolisinas. La deleción cromosómica de la 3'UTR de *rpiRc* produjo un aumento en la expresión de RpiRc y, en consecuencia, redujo la producción de hemolisinas. Éste y los otros ejemplos presentados en esta Tesis destacan la importancia de la regulación mediada por las 3'UTRs para modular procesos biológicos relevantes, como la formación de biofilms, el metabolismo del hierro y la virulencia en un patógeno de relevancia clínica como *S. aureus.*

12. Varias cepas de *S. aureus* contenían diferentes secuencias de inserción (IS1181, IS256, IS1082, etc.) localizadas en la 3'UTR de *rpiRc*. Las consecuencias de estas inserciones dependían del sitio, la orientación y el tipo de inserción. Específicamente, la inserción IS1181 creaba un mRNA de *rpiRc* quimérico que interrumpía la función de la 3'UTR. En consecuencia, se producía un aumento en la expresión de RpiRc y una reducción de la producción de hemolisinas. Esto demuestra

que los mecanismos reguladores mediados por 3'UTRs pueden verse afectados por los eventos de transposición.

13. Los estudios de conservación de varias 3'UTR funcionales descritas anteriormente, así como en el conjunto de los mRNAs de *E. coli* y *B. subtilis*, demostraron que la variabilidad de las secuencias de las 3'UTRs está muy extendida entre las bacterias.

14. La variabilidad en las 3'UTRs podría ser responsable de crear distintas funciones reguladoras y, en última instancia, contribuir en la diversidad bacteriana a lo largo de la evolución de manera similar a lo que sucedió durante la diversificación de las especies eucariotas.

BIBLIOGRAPHY

BIBLIOGRAPHY

Altuvia, Y., Bar, A., Reiss, N., Karavani, E., Argaman, L., and Margalit, H. (2018) *In vivo* cleavage rules and target repertoire of RNase III in *Escherichia coli. Nucleic Acids Res* **46**: 10380–10394.

Andrews, S.C., Robinson, A.K., and Rodríguez-Quiñones, F. (2003) Bacterial iron homeostasis. *FEMS Microbiol Rev* **27**: 215–237.

André, G., Even, S., Putzer, H., Burguière, P., Croux, C., Danchin, A., *et al.* (2008) S-box and T-box riboswitches and antisense RNA control a sulfur metabolic operon of *Clostridium acetobutylicum. Nucleic Acids Res* **36**: 5955–5969.

Arber, W. (2000) Genetic variation: molecular mechanisms and impact on microbial evolution. *FEMS Microbiol Rev* **24**: 1–7.

Archer, N.K., Mazaitis, M.J., Costerton, J.W., Leid, J.G., Powers, M.E., and Shirtliff, M.E. (2011) *Staphylococcus aureus* biofilms: properties, regulation, and roles in human disease. *Virulence* **2**: 445–459.

Arnaud, M., Chastanet, A., and Debarbouille, M. (2004) New vector for efficient allelic replacement in naturally nontransformable, low-GC-lontent, gram-positive bacteria. *Appl Environ Microbiol* **70**: 6887–6891.

Arraiano, C.M., Andrade, J.M., Domingues, S., Guinote, I.B., Malecki, M., Matos, R.G., *et al.* (2010) The critical role of RNA processing and degradation in the control of gene expression. *FEMS Microbiol Rev* **34**: 883–923.

Attaiech, L., Glover, J.N.M., and Charpentier, X. (2017) RNA chaperones step out of Hfq's shadow. *Trends Microbiol* **25**: 247–249.

Awano, N., Rajagopal, V., Arbing, M., Patel, S., Hunt, J., Inouye, M., and Phadtare, S. (2010) *Escherichia coli* RNase R has dual activities, helicase and RNase. *J Bacteriol* **192**: 1344–1352.

Baba, T., Takeuchi, F., Kuroda, M., Yuzawa, H., Aoki, K.-I., Oguchi, A., *et al.* (2002) Genome and virulence determinants of high virulence community-acquired MRSA. *Lancet* **359**: 1819–1827.

Babitzke, P., Baker, C.S., and Romeo, T. (2009) Regulation of translation initiation by RNA binding proteins. *Annu Rev Microbiol* **63**: 27–44.

Balaban, N., and Novick, R.P. (1995) Translation of RNAIII, the *Staphylococcus aureus agr* regulatory RNA molecule, can be activated by a 3'-end deletion. *FEMS Microbiol Lett* **133**: 155–161.

Balasubramanian, D., Ohneck, E.A., Chapman, J., Weiss, A., Kim, M.K., Reyes-Robles, T., *et al.* (2016) *Staphylococcus aureus* coordinates leukocidin expression and pathogenesis by sensing metabolic fluxes via RpiRc. *MBio* **7**: e00818–16.

Balestrino, D., Hamon, M.A., Dortet, L., Nahori, M.-A., Pizarro-Cerda, J., Alignani, D., *et al.* (2010) Single-cell techniques using chromosomally tagged fluorescent bacteria to study *Listeria monocytogenes* infection processes. *Appl Environ Microbiol* **76**: 3625–3636.

Beaume, M., Hernandez, D., Farinelli, L., Deluen, C., Linder, P., Gaspin, C., *et al.* (2010) Cartography of methicillin-resistant S. *aureus* transcripts: detection, orientation and temporal expression during growth phase and stress conditions. *PLoS ONE* **5**: e10725.

Benjamin, J.-A.M., and Massé, E. (2014) The iron-sensing aconitase B binds its own mRNA to prevent sRNA-induced mRNA cleavage. *Nucleic Acids Res* **42**: 10023–10036.

Bhatt, S., Edwards, A.N., Nguyen, H.T.T., Merlin, D., Romeo, T., and Kalman, D. (2009) The RNA binding protein CsrA is a pleiotropic regulator of the locus of enterocyte effacement pathogenicity island of enteropathogenic *Escherichia coli. Infect Immun* **77**: 3552–3568.

Bischoff, M., Entenza, J.M., and Giachino, P. (2001) Influence of a functional *sigB* operon on the global regulators *sar* and *agr* in *Staphylococcus aureus*. *J Bacteriol* **183**: 5171–5179.

Björk, G.R., and Hagervall, T.G. (2005) Transfer RNA Modification. *EcoSal Plus* **1**.

Bohn, C., Rigoulay, C., Chabelskaya, S., Sharma, C.M., Marchais, A., Skorski, P., *et al.* (2010) Experimental discovery of small RNAs in *Staphylococcus aureus* reveals a riboregulator of central metabolism. *Nucleic Acids Res* **38**: 6620–6636.

Boisset, S., Geissmann, T., Huntzinger, E., Fechter, P., Bendridi, N., Possedko, M., *et al.* (2007) *Staphylococcus aureus* RNAIII coordinately represses the synthesis of virulence factors and the transcription regulator Rot by an antisense mechanism. *Genes & Development* **21**: 1353–1366.

Bossi, L., and Figueroa-Bossi, N. (2016) Competing endogenous RNAs: a target-centric view of small RNA regulation in bacteria. *Nat Rev Microbiol* **14**: 775–784.

Brantl, S. (2007) Regulatory mechanisms employed by cis-encoded antisense RNAs. *Curr Opin Microbiol* **10**: 102–109.

Braun, F., Durand, S., and Condon, C. (2017) Initiating ribosomes and a 5'/3-'UTR interaction control ribonuclease action to tightly couple *B. subtilis hbs* mRNA stability with translation. *Nucleic Acids Res* **45**: 11386–11400.

Brennan, R.G., and Link, T.M. (2007) Hfq structure, function and ligand binding. *Curr Opin Microbiol* **10**: 125–133.

Britton, R.A., Wen, T., Schaefer, L., Pellegrini, O., Uicker, W.C., Mathy, N., *et al.* (2007) Maturation of the 5' end of *Bacillus subtilis* 16S rRNA by the essential ribonuclease YkqC/RNase J1. *Mol Microbiol* **63**: 127–138.

Brockhurst, M.A., Harrison, E., Hall, J.P.J., Richards, T., McNally, A., and MacLean, C. (2019) The ecology and evolution of pangenomes. *Curr Biol* **29**: R1094–R1103.

Broeke-Smits, ten, N.J.P., Pronk, T.E., Jongerius, I., Bruning, O., Wittink, F.R., Breit, T.M., *et al.* (2010) Operon structure of *Staphylococcus aureus. Nucleic Acids Res* **38**: 3263–3274.

Bronesky, D., Desgranges, E., Corvaglia, A., François, P., Caballero, C.J., Prado, L., *et al.* (2019) A multifaceted small RNA modulates gene expression upon glucose limitation in *Staphylococcus aureus*. *EMBO J* **38**: e99363.

Bronesky, D., Wu, Z., Marzi, S., Walter, P., Geissmann, T., Moreau, K., *et al.* (2016) *Staphylococcus aureus* RNAIII and its regulon link quorum sensing, stress responses, metabolic adaptation, and regulation of virulence gene expression. *Annu Rev Microbiol* **70**: 299–316.

Bækkedal, C., and Haugen, P. (2015) The Spot 42 RNA: A regulatory small RNA with roles in the central metabolism. *RNA Biol* **12**: 1071–1077.

Caballero, C.J., Menendez-Gil, P., Catalan-Moreno, A., Vergara-Irigaray, M., García, B., Segura, V., *et al.* (2018) The regulon of the RNA chaperone CspA and its auto-regulation in *Staphylococcus aureus*. *Nucleic Acids Res* **46**: 1345–1361.

Cahová, H., Winz, M.-L., Höfer, K., Nübel, G., and Jäschke, A. (2015) NAD captureSeq indicates NAD as a bacterial cap for a subset of regulatory RNAs. *Nature* **519**: 374–377.

Callaghan, A.J., Aurikko, J.P., Ilag, L.L., Günter Grossmann, J., Chandran, V., Kühnel, K., *et al.* (2004) Studies of the RNA degradosome-organizing domain of the *Escherichia coli* ribonuclease RNase E. *J Mol Biol* **340**: 965–979.

Caron, M.-P., Lafontaine, D.A., and Massé, E. (2010) Small RNA-mediated regulation at the level of transcript stability. *RNA Biol* **7**: 140–144.

Carroll, R.K., Weiss, A., Broach, W.H., Wiemels, R.E., Mogen, A.B., Rice, K.C., and Shaw, L.N. (2016) Genome-wide annotation, identification, and global transcriptomic analysis of regulatory or small RNA gene expression in *Staphylococcus aureus. MBio* **7**: e01990–15.

Chabelskaya, S., Bordeau, V., and Felden, B. (2014) Dual RNA regulatory control of a *Staphylococcus aureus* virulence factor. *Nucleic Acids Res* **42**: 4847–4858.

Chao, Y., and Vogel, J. (2010) The role of Hfq in bacterial pathogens. *Curr Opin Microbiol* **13**: 24–33.

Chao, Y., and Vogel, J. (2016) A 3' UTR-derived small RNA provides the regulatory noncoding arm of the inner membrane stress response. *Mol Cell* **61**: 352–363.

Chao, Y., Li, L., Girodat, D., Förstner, K.U., Said, N., Corcoran, C., *et al.* (2017) *In vivo* cleavage map illuminates the central tole of RNase E in coding and non-coding RNA pathways. *Mol Cell* **65**: 39–51.

Chao, Y., Papenfort, K., Reinhardt, R., Sharma, C.M., and Vogel, J. (2012) An atlas of Hfq-bound transcripts reveals 3' UTRs as a genomic reservoir of regulatory small RNAs. *EMBO J* **31**: 4005–4019.

Charollais, J., Pflieger, D., Vinh, J., Dreyfus, M., and Iost, I. (2003) The DEAD-box RNA helicase SrmB is involved in the assembly of 50S ribosomal subunits in *Escherichia coli. Mol Microbiol* **48**: 1253–1265.

Charpentier, E., Anton, A.I., Barry, P., Alfonso, B., Fang, Y., and Novick, R.P. (2004) Novel cassette-based shuttle vector system for gram-positive bacteria. *Appl Environ Microbiol* **70**: 6076–6085.

Chen, Y.G., Kowtoniuk, W.E., Agarwal, I., Shen, Y., and Liu, D.R. (2009) LC/MS analysis of cellular RNA reveals NAD-linked RNA. *Nat Chem Biol* **5**: 879–881.

Cheng, C., Bhardwaj, N., and Gerstein, M. (2009) The relationship between the evolution of microRNA targets and the length of their UTRs. *BMC Genomics* **10**: 431.

Cheng, Z.-F., and Deutscher, M.P. (2005) An important role for RNase R in mRNA decay. *Mol Cell* **17**: 313–318.

Chesneau, O., Lailler, R., Derbise, A., and Solh, El, N. (1999) Transposition of IS1181 in the genomes of *Staphylococcus* and *Listeria*. *FEMS Microbiol Lett* **177**: 93–100.

Chevalier, C., Boisset, S., Romilly, C., Masquida, B., Fechter, P., Geissmann, T., *et al.* (2010) *Staphylococcus aureus* RNAIII binds to two distant regions of *coa* mRNA to arrest translation and promote mRNA degradation. *PLoS Pathog* **6**: e1000809.

Cirz, R.T., Jones, M.B., Gingles, N.A., Minogue, T.D., Jarrahi, B., Peterson, S.N., and Romesberg, F.E. (2007) Complete and SOS-mediated response of *Staphylococcus aureus* to the antibiotic ciprofloxacin. *J Bacteriol* **189**: 531–539.

Cohen, O., Doron, S., Wurtzel, O., Dar, D., Edelheit, S., Karunker, I., *et al.* (2016) Comparative transcriptomics across the prokaryotic tree of life. *Nucleic Acids Res* **44**: W46–53.

Collins, J.A., Irnov, I., Baker, S., and Winkler, W.C. (2007) Mechanism of mRNA destabilization by the *glmS* ribozyme. *Genes & Development* **21**: 3356–3368.

Commichau, F.M., Rothe, F.M., Herzberg, C., Wagner, E., Hellwig, D., Lehnik-Habrink, M., *et al.* (2009) Novel activities of glycolytic enzymes in *Bacillus subtilis*: interactions with essential proteins involved in mRNA processing. *Mol Cell Proteomics* **8**: 1350–1360.

Connolly, J.P.R., O'Boyle, N., Turner, N.C.A., Browning, D.F., and Roe, A.J. (2019) Distinct intraspecies virulence mechanisms regulated by a conserved transcription factor. *Proc Natl Acad Sci USA* **116**: 19695–19704.

Conway, T., Creecy, J.P., Maddox, S.M., Grissom, J.E., Conkle, T.L., Shadid, T.M., *et al.* (2014) Unprecedented high-resolution view of bacterial operon architecture revealed by RNA sequencing. *MBio* **5**: e01442–14.

Cramton, S.E., Gerke, C., Schnell, N.F., Nichols, W.W., and Götz, F. (1999) The intercellular adhesion (*ica*) locus is present in *Staphylococcus aureus* and is required for biofilm formation. *Infect Immun* **67**: 5427–5433.

Cramton, S.E., Schnell, N.F., Götz, F., and Brückner, R. (2000) Identification of a new repetitive element in *Staphylococcus aureus*. *Infect Immun* **68**: 2344–2348.

Crick, F. (1970) Central dogma of molecular biology. *Nature* **227**: 561–563.

Dar, D., and Sorek, R. (2018) High-resolution RNA 3'-ends mapping of bacterial Rho-dependent transcripts. *Nucleic Acids Res* **46**: 6797–6805.

Dar, D., Shamir, M., Mellin, J.R., Koutero, M., Stern-Ginossar, N., Cossart, P., and Sorek, R. (2016) Term-seq reveals abundant ribo-regulation of antibiotics resistance in bacteria. *Science* **352**: aad9822-1–aad9822-12.

Darling, A.C.E., Mau, B., Blattner, F.R., and Perna, N.T. (2004) Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* **14**: 1394–1403.

Davis, B.M., and Waldor, M.K. (2007) RNase E-dependent processing stabilizes MicX, a *Vibrio cholerae* sRNA. *Mol Microbiol* **65**: 373–385.

De Mets, F., Van Melderen, L., and Gottesman, S. (2019) Regulation of acetate metabolism and coordination with the TCA cycle via a processed small RNA. *Proc Natl Acad Sci USA* **116**: 1043–1052.

Delihas, N. (2011) Impact of small repeat sequences on bacterial genome evolution. *Genome Biol Evol* **3**: 959–973.

Deng, X., Chen, K., Luo, G.-Z., Weng, X., Ji, Q., Zhou, T., and He, C. (2015) Widespread occurrence of N6-methyladenosine in bacterial mRNA. *Nucleic Acids Res* **43**: 6557–6567.

Diep, B.A., and Otto, M. (2008) The role of virulence determinants in community-associated MRSA pathogenesis. *Trends Microbiol* **16**: 361–369.

Duval, B.D., Mathew, A., Satola, S.W., and Shafer, W.M. (2010) Altered growth, pigmentation, and antimicrobial susceptibility properties of *Staphylococcus aureus* due to loss of the major cold shock gene *cspB*. *Antimicrobial Agents and Chemotherapy* **54**: 2283–2290.

El-Mouali, Y., and Balsalobre, C. (2018) 3'untranslated regions: regulation at the end of the road. *Curr Genet* **13**: 497–5.

El-Mouali, Y., Gaviria-Cantin, T., Sánchez-Romero, M.A., Gibert, M., Westermann, A.J., Vogel, J., and Balsalobre, C. (2018) CRP-cAMP mediates silencing of *Salmonella* virulence at the post-transcriptional level. *PLoS Genet* **14**: e1007401.

Even, S., Pellegrini, O., Zig, L., Labas, V., Vinh, J., Bréchemmier-Baey, D., and Putzer, H. (2005) Ribonucleases J1 and J2: two novel endoribonucleases in *B. subtilis* with functional homology to *E.coli* RNase E. *Nucleic Acids Res* **33**: 2141–2152.

Fozo, E.M., Hemm, M.R., and Storz, G. (2008) Small toxic proteins and the antisense RNAs that repress them. *Microbiol Mol Biol Rev* **72**: 579–589.

Freitag, N.E., Port, G.C., and Miner, M.D. (2009) *Listeria monocytogenes* - from saprophyte to intracellular pathogen. *Nat Rev Microbiol* **7**: 623–628.

Fröhlich, K.S., Papenfort, K., Fekete, A., and Vogel, J. (2013) A small RNA activates CFA synthase by isoform-specific mRNA stabilization. *EMBO J* **32**: 2963–2979.

Gaupp, R., Wirf, J., Wonnenberg, B., Biegel, T., Eisenbeis, J., Graham, J., *et al.* (2016) RpiRc Is a Pleiotropic Effector of Virulence Determinant Synthesis and Attenuates Pathogenicity in *Staphylococcus aureus*. *Infect Immun* **84**: 2031–2041.

Gerstmeir, R., Wendisch, V.F., Schnicke, S., Ruan, H., Farwick, M., Reinscheid, D., and Eikmanns, B.J. (2003) Acetate metabolism and its regulation in *Corynebacterium glutamicum. J Biotechnol* **104**: 99–122.

Gor, V., Takemura, A.J., Nishitani, M., Higashide, M., Medrano Romero, V., Ohniwa, R.L., and Morikawa, K. (2019) Finding of Agr phase variants in *Staphylococcus aureus*. *MBio* **10**: e00796-19.

Gripenland, J., Netterling, S., Loh, E., Tiensuu, T., Toledo-Arana, A., and Johansson, J. (2010) RNAs: regulators of bacterial virulence. *Nat Rev Microbiol* **8**: 857–866.

Guillier, M., Allemand, F., Dardel, F., Royer, C.A., Springer, M., and Chiaruttini, C. (2005) Double molecular mimicry in *Escherichia coli:* binding of ribosomal protein L20 to its two sites in mRNA is similar to its binding to 23S rRNA. *Mol Microbiol* **56**: 1441–1456.

Guo, M.S., Updegrove, T.B., Gogol, E.B., Shabalina, S.A., Gross, C.A., and Storz, G. (2014) MicL, a new σ^E-dependent sRNA, combats envelope stress by repressing synthesis of Lpp, the major outer membrane lipoprotein. *Genes & Development* **28**: 1620–1634.

Gupta, R.K., Luong, T.T., and Lee, C.Y. (2015) RNAIII of the *Staphylococcus aureus* agr system activates global regulator MgrA by stabilizing mRNA. *Proc Natl Acad Sci USA* **112**: 14036–14041.

Halees, A.S., El-Badrawi, R., and Khabar, K.S.A. (2008) ARED Organism: expansion of ARED reveals AU-rich element cluster variations between human and mouse. *Nucleic Acids Res* **36**: D137–40.

Hardy, K.J., Ussery, D.W., Oppenheim, B.A., and Hawkey, P.M. (2004) Distribution and characterization of staphylococcal interspersed repeat units (SIRUs) and potential use for strain differentiation. *Microbiology (Reading, Engl)* **150**: 4045–4052.

Hedges, S.B., Marin, J., Suleski, M., Paymer, M., and Kumar, S. (2015) Tree of life reveals clock-like speciation and diversification. *Mol Biol Evol* **32**: 835–845.

Holmqvist, E., and Vogel, J. (2018) RNA-binding proteins in bacteria. *Nat Rev Microbiol* **2**: a003483.

Holmqvist, E., Li, L., Bischler, T., Barquist, L., and Vogel, J. (2018) Global maps of ProQ binding *in vivo* reveal target recognition via RNA structure and stability control at mRNA 3' ends. *Mol Cell* **70**: 971–982.

Holt, D.C., Holden, M.T.G., Tong, S.Y.C., Castillo-Ramirez, S., Clarke, L., Quail, M.A., *et al.* (2011) A very early-branching *Staphylococcus aureus* lineage lacking the carotenoid pigment staphyloxanthin. *Genome Biol Evol* **3**: 881–895.

Hood, M.I., and Skaar, E.P. (2012) Nutritional immunity: transition metals at the pathogen-host interface. *Nat Rev Microbiol* **10**: 525–537.

Horn, J., Stelzner, K., Rudel, T., and Fraunholz, M. (2018) Inside job: *Staphylococcus aureus* host-pathogen interactions. *Int J Med Microbiol* **308**: 607–624.

Horsburgh, M.J., Clements, M.O., Crossley, H., Ingham, E., and Foster, S.J. (2001) PerR controls oxidative stress resistance and iron storage proteins and is required for virulence in *Staphylococcus aureus. Infect Immun* **69**: 3744–3754.

Houston, P., Rowe, S.E., Pozzi, C., Waters, E.M., and O'Gara, J.P. (2011) Essential role for the major autolysin in the fibronectin-binding protein-mediated *Staphylococcus aureus* biofilm phenotype. *Infect Immun* **79**: 1153–1165.

Huntzinger, E., Boisset, S., Saveanu, C., Benito, Y., Geissmann, T., Namane, A., *et al.* (2005) *Staphylococcus aureus* RNAIII and the endoribonuclease III coordinately regulate *spa* gene expression. *EMBO J* **24**: 824–835.

Intile, P.J., Balzer, G.J., Wolfgang, M.C., and Yahr, T.L. (2015) The RNA helicase DeaD stimulates ExsA translation to promote expression of the *Pseudomonas aeruginosa* type III secretion system. *J Bacteriol* **197**: 2664–2674.

Jester, B.C., Romby, P., and Lioliou, E. (2012) When ribonucleases come into play in pathogens: a survey of gram-positive bacteria. *Int J Microbiol* **2012**: 592196.

Johansson, J., Mandin, P., Renzoni, A., Chiaruttini, C., Springer, M., and Cossart, P. (2002) An RNA thermosensor controls expression of virulence genes in *Listeria monocytogenes. Cell* **110**: 551–561.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., and Madden, T.L. (2008) NCBI BLAST: a better web interface. *Nucleic Acids Res* **36**: W5–9.

Joseph, S.J., Li, B., Petit, R.A., III, Qin, Z.S., Darrow, L., and Read, T.D. (2016) The single-species metagenome: subtyping *Staphylococcus aureus* core genome sequences from shotgun metagenomic data. *PeerJ* **4**: e2571.

Kawano, M., Reynolds, A.A., Miranda-Rios, J., and Storz, G. (2005) Detection of 5'- and 3-'UTR-derived small RNAs and *cis*-encoded antisense RNAs in *Escherichia coli. Nucleic Acids Res* **33**: 1040–1050.

Khabar, K.S.A. (2010) Post-transcriptional control during chronic inflammation and cancer: a focus on AU-rich elements. *Cell Mol Life Sci* **67**: 2937–2955.

Khemici, V., and Linder, P. (2016) RNA helicases in bacteria. *Curr Opin Microbiol* **30**: 58–66.

Kim, H.M., Shin, J.-H., Cho, Y.-B., and Roe, J.-H. (2014) Inverse regulation of Fe- and Ni-containing SOD genes by a Fur family regulator Nur through small RNA processed from 3'UTR of the *sodF* mRNA. *Nucleic Acids Res* **42**: 2003–2014.

Kingsford, C.L., Ayanbule, K., and Salzberg, S.L. (2007) Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake. *Genome Biol* **8**: R22.

Koch, G., Yepes, A., Förstner, K.U., Wermser, C., Stengel, S.T., Modamio, J., *et al.* (2014) Evolution of resistance to a last-resort antibiotic in *Staphylococcus aureus* via bacterial competition. *Cell* **158**: 1060–1071.

Kortmann, J., and Narberhaus, F. (2012) Bacterial RNA thermometers: molecular zippers and switches. *Nat Rev Microbiol* **10**: 255–265.

Kowtoniuk, W.E., Shen, Y., Heemstra, J.M., Agarwal, I., and Liu, D.R. (2009) A chemical screen for biological small molecule-RNA conjugates reveals CoA-linked RNA. *Proc Natl Acad Sci USA* **106**: 7768–7773.

Kröger, C., Dillon, S.C., Cameron, A.D.S., Papenfort, K., Sivasankaran, S.K., Hokamp, K., *et al.* (2012) The transcriptional landscape and small RNAs of *Salmonella enterica* serovar Typhimurium. *Proc Natl Acad Sci USA* **109**: E1277–86.

Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I., *et al.* (2001) Whole genome sequencing of meticillin-resistant *Staphylococcus aureus. Lancet* **357**: 1225–1240.

Lasa, I., Toledo-Arana, A., and Gingeras, T.R. (2012) An effort to make sense of antisense transcription in bacteria. *RNA Biol* **9**: 1039–1044.

Lasa, I., Toledo-Arana, A., Dobin, A., Villanueva, M., de los Mozos, I.R., Vergara-Irigaray, M., *et al.* (2011) Genome-wide antisense transcription drives mRNA processing in bacteria. *Proc Natl Acad Sci USA* **108**: 20172–20177.

Lee, J.C. (1995) Electrotransformation of *Staphylococci*. *Methods Mol Biol* **47**: 209–216.

Lehnik-Habrink, M., Pförtner, H., Rempeters, L., Pietack, N., Herzberg, C., and Stülke, J. (2010) The RNA degradosome in *Bacillus subtilis*: identification of CshA as the major RNA helicase in the multiprotein complex. *Mol Microbiol* **77**: 958–971.

Lehnik-Habrink, M., Schaffer, M., Mäder, U., Diethmaier, C., Herzberg, C., and Stülke, J. (2011) RNA processing in *Bacillus subtilis:* identification of targets of the essential RNase Y. *Mol Microbiol* **81**: 1459–1473.

Lerch, M.F., Schoenfelder, S.M.K., Marincola, G., Wencker, F.D.R., Eckart, M., Förstner, K.U., *et al.* (2019) A non-coding RNA from the intercellular adhesion (*ica*) locus of *Staphylococcus epidermidis* controls polysaccharide intercellular adhesion (PIA)-mediated biofilm formation. *Mol Microbiol* **111**: 1571–1591.

Li, Z., and Deutscher, M.P. (2002) RNase E plays an essential role in the maturation of *Escherichia coli* tRNA precursors. *RNA* **8**: 97–109.

Lioliou, E., Sharma, C.M., Caldelari, I., Helfer, A.-C., Fechter, P., Vandenesch, F., *et al.* (2012) Global regulatory functions of the *Staphylococcus aureus* endoribonuclease III in gene expression. *PLoS Genet* **8**: e1002782–21.

Liu, M.Y., Gui, G., Wei, B., Preston, J.F., Oakford, L., Yüksel, U., *et al.* (1997) The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in *Escherichia coli. J Biol Chem* **272**: 17502–17510.

Loh, E., Dussurget, O., Gripenland, J., Vaitkevicius, K., Tiensuu, T., Mandin, P., *et al.* (2009) A trans-acting riboswitch controls expression of the virulence regulator PrfA in *Listeria monocytogenes*. *Cell* **139**: 770–779.

Loh, E., Righetti, F., Eichner, H., Twittenhoff, C., and Narberhaus, F. (2018) RNA thermometers in bacterial pathogens. *Microbiol Spectr* **6**: RWR–0012–2017.

López-Garrido, J., Puerta-Fernández, E., and Casadesús, J. (2014) A eukaryotic-like 3' untranslated region in *Salmonella enterica hilD* mRNA. *Nucleic Acids Res* **42**: 5894–5906.

Luo, H., Tang, J., Friedman, R., and Hughes, A.L. (2011) Ongoing purifying selection on intergenic spacers in group A streptococcus. *Infect Genet Evol* **11**: 343–348.

Luttinger, A., Hahn, J., and Dubnau, D. (1996) Polynucleotide phosphorylase is necessary for competence development in *Bacillus subtilis. Mol Microbiol* **19**: 343–356.

Lybecker, M., Zimmermann, B., Bilusic, I., Tukhtubaeva, N., and Schroeder, R. (2014) The double-stranded transcriptome of *Escherichia coli. Proc Natl Acad Sci USA* **111**: 3134–3139.

Lyon, B.R., Gillespie, M.T., and Skurray, R.A. (1987) Detection and characterization of IS256, an insertion sequence in *Staphylococcus aureus*. *J Gen Microbiol* **133**: 3031–3038.

Mackie, G.A. (1998) Ribonuclease E is a 5'-end-dependent endonuclease. *Nature* **395**: 720–723.

Maeda, T., and Wachi, M. (2012) 3' Untranslated region-dependent degradation of the *aceA* mRNA, encoding the glyoxylate cycle enzyme isocitrate lyase, by RNase E/G in *Corynebacterium glutamicum*. *Appl Environ Microbiol* **78**: 8753–8761.

Maira-Litrán, T., Kropec, A., Goldmann, D.A., and Pier, G.B. (2005) Comparative opsonic and protective activities of *Staphylococcus aureus* conjugate vaccines containing native or deacetylated Staphylococcal Poly-N-acetyl-beta-(1-6)-glucosamine. *Infect Immun* **73**: 6752–6762.

Manna, A., and Cheung, A.L. (2001) Characterization of *sarR*, a modulator of *sar* expression in *Staphylococcus aureus. Infect Immun* **69**: 885–896.

Marbaniang, C.N., and Vogel, J. (2016) Emerging roles of RNA modifications in bacteria. *Curr Opin Microbiol* **30**: 50–57.

Marincola, G., Wencker, F.D.R., and Ziebuhr, W. (2019) The many facets of the small non-coding RNA RsaE (RoxS) in metabolic niche adaptation of gram-positive bacteria. *J Mol Biol* **431**: 4684–4698.

Mata, J., Marguerat, S., and Bähler, J. (2005) Post-transcriptional control of gene expression: a genome-wide perspective. *Trends Biochem Sci* **30**: 506–514.

Mayr, C. (2017) Regulation by 3'-untranslated regions. *Annu Rev Genet* **51**: 171–194.

Mayr, C., and Bartel, D.P. (2009) Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* **138**: 673–684.

Mazumder, B., Seshadri, V., and Fox, P.L. (2003) Translational control by the 3′ -UTR: the ends specify the means. *Trends Biochem Sci* **28**: 91–98.

Mellin, J.R., and Cossart, P. (2015) Unexpected versatility in bacterial riboswitches. *Trends Genet* **31**: 150–156.

Menendez-Gil, P., Caballero, C.J, Catalan-Moreno A., Irurzun N., Barrio-Hernandez I., Caldelari I. and Toledo-Arana A. (2020a) Differential evolution in 3'UTRs leads to specific gene expression in *Staphylococcus*. *Nucleic Acids Res* **48**: 2544–2563,

Menendez-Gil, P., Caballero, C.J., Solano, C., and Toledo-Arana A. (2020b) Fluorescent molecular beacons mimicking RNA secondary structures to study RNA chaperone activity. *Methods Mol Biol* **2106**: 41–58.

Mercurio, S., Serra, L., and Nicolis, S.K. (2019) More than just stem cells: functional roles of the transcription factor Sox2 in differentiated glia and neurons. *Int J Mol Sci* **20**: 4540.

Michaux, C., Holmqvist, E., Vasicek, E., Sharan, M., Barquist, L., Westermann, A.J., *et al.* (2017) RNA target profiles direct the discovery of virulence functions for the cold-shock proteins CspC and CspE. *Proc Natl Acad Sci USA* **114**: 6824–6829.

Miczak, A., Kaberdin, V.R., Wei, C.L., and Lin-Chao, S. (1996) Proteins associated with RNase E in a multicomponent ribonucleolytic complex. *Proc Natl Acad Sci USA* **93**: 3865–3869.

Miyakoshi, M., Chao, Y., and Vogel, J. (2015a) Regulatory small RNAs from the 3' regions of bacterial mRNAs. *Curr Opin Microbiol* **24**: 132–139.

Miyakoshi, M., Chao, Y., and Vogel, J. (2015b) Cross talk between ABC transporter mRNAs via a target mRNA-derived sponge of the GcvB small RNA. *EMBO J* **34**: 1478–1492.

Miyakoshi, M., Matera, G., Maki, K., Sone, Y., and Vogel, J. (2019) Functional expansion of a TCA cycle operon mRNA by a 3' end-derived small RNA. *Nucleic Acids Res* **47**: 2075–2088.

Mohanty, B.K., and Kushner, S.R. (2016) Regulation of mRNA decay in bacteria. *Annu Rev Microbiol* **70**: 25–44.

Molina, N., and van Nimwegen, E. (2008) Universal patterns of purifying selection at noncoding positions in bacteria. *Genome Res* **18**: 148–160.

Morita, T., Maki, K., and Aiba, H. (2005) RNase E-based ribonucleoprotein complexes: mechanical basis of mRNA destabilization mediated by bacterial noncoding RNAs. *Genes & Development* **19**: 2176–2186.

Morita, T., Mochizuki, Y., and Aiba, H. (2006) Translational repression is sufficient for gene silencing by bacterial small noncoding RNAs in the absence of mRNA destruction. *Proc Natl Acad Sci USA* **103**: 4858–4863.

Morrissey, J.A., Cockayne, A., Brummell, K., and Williams, P. (2004) The staphylococcal ferritins are differentially regulated in response to iron and manganese and via PerR and Fur. *Infect Immun* **72**: 972–979.

Møller, T., Franch, T., Højrup, P., Keene, D.R., Bächinger, H.P., Brennan, R.G., and Valentin-Hansen, P. (2002) Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction. *Mol Cell* **9**: 23–30.

Nahvi, A., Sudarsan, N., Ebert, M.S., Zou, X., Brown, K.L., and Breaker, R.R. (2002) Genetic control by a metabolite binding mRNA. *Chem Biol* **9**: 1043–1049.

Nair, D., Memmi, G., Hernandez, D., Bard, J., Beaume, M., Gill, S., *et al.* (2011) Whole-genome sequencing of *Staphylococcus aureus* strain RN4220, a key laboratory strain used in virulence research, identifies mutations that affect not only virulence factors but also the fitness of the strain. *J Bacteriol* **193**: 2332–2335.

Nechooshtan, G., Elgrably-Weiss, M., Sheaffer, A., Westhof, E., and Altuvia, S. (2009) A pH-responsive riboregulator. *Genes & Development* **23**: 2650–2662.

Newman, J.A., Hewitt, L., Rodrigues, C., Solovyova, A., Harwood, C.R., and Lewis, R.J. (2011) Unusual, dual endo- and exonuclease activity in the degradosome explained by crystal structure analysis of RNase J1. *Structure* **19**: 1241–1251.

Nicolas, P., Mäder, U., Dervyn, E., Rochat, T., Leduc, A., Pigeonneau, N., *et al.* (2012) Condition-dependent transcriptome reveals high-level regulatory architecture in *Bacillus subtilis*. *Science* **335**: 1103–1106.

Nielsen, J.S., Lei, L.K., Ebersbach, T., Olsen, A.S., Klitgaard, J.K., Valentin-Hansen, P., and Kallipolitis, B.H. (2009) Defining a role for Hfq in Gram-positive bacteria: evidence for Hfq-dependent antisense regulation in *Listeria monocytogenes*. *Nucleic Acids Res* **38**: 907–919.

Nomura, M. (1970) Bacterial ribosome. *Bacteriol Rev* **34**: 228–277.

Nomura, M. (1999) Regulation of ribosome biosynthesis in *Escherichia coli* and *Saccharomyces cerevisiae*: diversity and common principles. *J Bacteriol* **181**: 6857–6864.

Novick, R.P. (2003) Autoinduction and signal transduction in the regulation of staphylococcal virulence. *Mol Microbiol* **48**: 1429–1449.

Novick, R.P., Ross, H.F., Projan, S.J., Kornblum, J., Kreiswirth, B., and Moghazeh, S. (1993) Synthesis of staphylococcal virulence factors is controlled by a regulatory RNA molecule. *EMBO J* **12**: 3967–3975.

Oberto, J. (2013) SyntTax: a web server linking synteny to prokaryotic taxonomy. *BMC Bioinformatics* **14**: 4–10.

Oliveira, D., Borges, A., and Simões, M. (2018) *Staphylococcus aureus* toxins and their molecular activity in infectious diseases. *Toxins* **10**: 252.

Opdyke, J.A., Kang, J.-G., and Storz, G. (2004) GadY, a small-RNA regulator of acid response genes in *Escherichia coli. J Bacteriol* **186**: 6698–6705.

Oren, Y., Smith, M.B., Johns, N.I., Kaplan Zeevi, M., Biran, D., Ron, E.Z., *et al.* (2014) Transfer of noncoding DNA drives regulatory rewiring in bacteria. *Proc Natl Acad Sci USA* **111**: 16112–16117.

Ow, M.C., Perwez, T., and Kushner, S.R. (2003) RNase G of *Escherichia coli* exhibits only limited functional overlap with its essential homologue, RNase E. *Mol Microbiol* **49**: 607–622.

Painter, K.L., Krishna, A., Wigneshweraraj, S., and Edwards, A.M. (2014) What role does the quorum-sensing accessory gene regulator system play during *Staphylococcus aureus* bacteremia? *Trends Microbiol* **22**: 676–685.

Peng, T., Berghoff, B.A., Oh, J.-I., Weber, L., Schirmer, J., Schwarz, J., *et al.* (2016) Regulation of a polyamine transporter by the conserved 3' UTR-derived sRNA SorX confers resistance to singlet oxygen and organic hydroperoxides in *Rhodobacter sphaeroides. RNA Biol* **13**: 988–999.

Peselis, A., and Serganov, A. (2014) Themes and variations in riboswitch structure and function. *Biochim Biophys Acta* **1839**: 908–918.

Peters, J.M., Vangeloff, A.D., and Landick, R. (2011) Bacterial transcription terminators: the RNA 3'-end chronicles. *J Mol Biol* **412**: 793–813.

Pfeiffer, V., Papenfort, K., Lucchini, S., Hinton, J.C.D., and Vogel, J. (2009) Coding sequence targeting by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation. *Nat Struct Mol Biol* **16**: 840–846.

Phadtare, S., and Severinov, K. (2010) RNA remodeling and gene regulation by cold shock proteins. *RNA Biol* **7**: 788–795.

Pragman, A.A., Ji, Y., and Schlievert, P.M. (2007) Repression of *Staphylococcus aureus* SrrAB using inducible antisense *srrA* alters growth and virulence factor transcript levels. *Biochemistry* **46**: 314–321.

Prud'homme-Généreux, A., Beran, R.K., Iost, I., Ramey, C.S., Mackie, G.A., and Simons, R.W. (2004) Physical and functional interactions among RNase E, polynucleotide phosphorylase and the cold-shock protein, CsdA: evidence for a 'cold shock degradosome'. *Mol Microbiol* **54**: 1409–1421.

Purves, J., Blades, M., Arafat, Y., Malik, S.A., Bayliss, C.D., and Morrissey, J.A. (2012) Variation in the genomic locations and sequence conservation of STAR elements among staphylococcal species provides insight into DNA repeat evolution. *BMC Genomics* **13**: 515–13.

Queck, S.Y., Jameson-Lee, M., Villaruz, A.E., Bach, T.-H.L., Khan, B.A., Sturdevant, D.E., *et al.* (2008) RNAIII-independent target gene control by the *agr* quorum-sensing system: insight into the evolution of virulence regulation in *Staphylococcus aureus*. *Mol Cell* **32**: 150–158.

Raghavan, R., Kacharia, F.R., Millar, J.A., Sislak, C.D., and Ochman, H. (2015) Genome rearrangements can make and break small RNA genes. *Genome Biol Evol* **7**: 557–566.

Rajkowitsch, L., Chen, D., Stampfl, S., Semrad, K., Waldsich, C., Mayer, O., *et al.* (2007) RNA chaperones, RNA annealers and RNA helicases. *RNA Biol* **4**: 118–130.

Redko, Y., Bechhofer, D.H., and Condon, C. (2008) Mini-III, an unusual member of the RNase III family of enzymes, catalyses 23S ribosomal RNA maturation in *B. subtilis*. *Mol Microbiol* **68**: 1096–1106.

Ren, G.-X., Guo, X.-P., and Sun, Y.-C. (2017) Regulatory 3' untranslated regions of bacterial mRNAs. *Front Microbiol* **8**: 1276.

Resch, A., Večerek, B., Palavra, K., and Bläsi, U. (2010) Requirement of the CsdA DEAD-box helicase for low temperature riboregulation of *rpoS* mRNA. *RNA Biol* **7**: 796–802.

Romeo, T., and Babitzke, P. (2018) Global regulation by CsrA and its RNA antagonists. *Microbiol Spectr* **6:** RWR-0009-2017.

Roßmanith, J., and Narberhaus, F. (2017) Modular arrangement of regulatory RNA elements. *RNA Biol* **14**: 287–292.

Roux, C.M., DeMuth, J.P., and Dunman, P.M. (2011) Characterization of components of the *Staphylococcus aureus* mRNA degradosome holoenzyme-like complex. *J Bacteriol* **193**: 5520–5526.

Ruiz de los Mozos, I., Vergara-Irigaray, M., Segura, V., Villanueva, M., Bitarte, N., Saramago, M., *et al.* (2013) Base pairing interaction between 5'- and 3-'UTRs controls *icaR* mRNA translation in *Staphylococcus aureus*. *PLoS Genet* **9**: e1004001.

Sáenz-Lahoya, S., Bitarte, N., García, B., Burgui, S., Vergara-Irigaray, M., Valle, J., *et al.* (2019) Noncontiguous operon is a genetic organization for coordinating bacterial gene expression. *Proc Natl Acad Sci USA* **116**: 1733-1738.

Schoenfelder, S.M.K., Lange, C., Prakash, S.A., Marincola, G., Lerch, M.F., Wencker, F.D.R., *et al.* (2019) The small non-coding RNA RsaE influences extracellular matrix composition in *Staphylococcus epidermidis* biofilm communities. *PLoS Pathog* **15**: e1007618.

Schubert, M., Lapouge, K., Duss, O., Oberstrass, F.C., Jelesarov, I., Haas, D., and Allain, F.H.-T. (2007) Molecular basis of messenger RNA recognition by the specific bacterial repressing clamp RsmA/CsrA. *Nat Struct Mol Biol* **14**: 807–813.

Sesto, N., Wurtzel, O., Archambaud, C., Sorek, R., and Cossart, P. (2013) The excludon: a new concept in bacterial antisense RNA-mediated gene regulation. *Nat Rev Microbiol* **11**: 75–82.

Shahbabian, K., Jamalli, A., Zig, L., and Putzer, H. (2009) RNase Y, a novel endoribonuclease, initiates riboswitch turnover in *Bacillus subtilis. EMBO J* **28**: 3523–3533.

Sharma, C.M., Hoffmann, S., Darfeuille, F., Reignier, J., Findeiß, S., Sittka, A., *et al.* (2010) The primary transcriptome of the major human pathogen *Helicobacter pylori. Nature* **464**: 250–255.

Sherwood, A.V., and Henkin, T.M. (2016) Riboswitch-mediated gene regulation: novel RNA architectures dictate gene expression responses. *Annu Rev Microbiol* **70**: 361–374.

Siguier, P., Gourbeyre, E., and Chandler, M. (2014) Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiol Rev* **38**: 865–891.

Siguier, P., Perochon, J., Lestrade, L., Mahillon, J., and Chandler, M. (2006) ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res* **34**: D32–6.

Singleton, M.R., Dillingham, M.S., and Wigley, D.B. (2007) Structure and mechanism of helicases and nucleic acid translocases. *Annu Rev Biochem* **76**: 23–50.

Skinner, M.E., Uzilov, A.V., Stein, L.D., Mungall, C.J., and Holmes, I.H. (2009) JBrowse: a next-generation genome browser. *Genome Res* **19**: 1630–1638.

Soper, T., Mandin, P., Majdalani, N., Gottesman, S., and Woodson, S.A. (2010) Positive regulation by small RNAs and the role of Hfq. *Proc Natl Acad Sci USA* **107**: 9602–9607.

Spickler, C., and Mackie, G.A. (2000) Action of RNase II and polynucleotide phosphorylase against RNAs containing stem-loops of defined structure. *J Bacteriol* **182**: 2422–2427.

Sridhar, J., Sabarinathan, R., Balan, S.S., Rafi, Z.A., Gunasekaran, P., and Sekar, K. (2011) Junker: an intergenic explorer for bacterial genomes. *Genomics Proteomics Bioinformatics* **9**: 179–182.

Stead, M.B., Marshburn, S., Mohanty, B.K., Mitra, J., Pena Castillo, L., Ray, D., *et al.* (2011) Analysis of *Escherichia coli* RNase E and RNase III activity *in vivo* using tiling microarrays. *Nucleic Acids Res* **39**: 3188–3203.

Stork, M., Di Lorenzo, M., Welch, T.J., and Crosa, J.H. (2007) Transcription termination within the iron transport-biosynthesis operon of *Vibrio anguillarum* requires an antisense RNA. *J Bacteriol* **189**: 3479–3488.

Suzuki, H., Lefébure, T., Bitar, P.P., and Stanhope, M.J. (2012) Comparative genomic analysis of the genus *Staphylococcus* including *Staphylococcus aureus* and its newly described sister species *Staphylococcus simiae*. *BMC Genomics* **13**: 38.

Thorpe, H.A., Bayliss, S.C., Hurst, L.D., and Feil, E.J. (2017) Comparative analyses of selection operating on nontranslated intergenic regions of diverse bacterial species. *Genetics* **206**: 363–376.

Timmermans, J., and Van Melderen, L. (2010) Post-transcriptional global regulation by CsrA in bacteria. *Cell Mol Life Sci* **67**: 2897–2908.

Tjaden, B. (2015) *De novo* assembly of bacterial transcriptomes from RNA-seq data. *Genome Biol* **16**: 1.

Toledo-Arana, A., Dussurget, O., Nikitas, G., Sesto, N., Guet-Revillet, H., Balestrino, D., *et al.* (2009) The *Listeria* transcriptional landscape from saprophytism to virulence. *Nature* **459**: 950–956.

Toledo-Arana, A., Merino, N., Vergara-Irigaray, M., Debarbouille, M., Penadés, J.R., and Lasa, I. (2005) *Staphylococcus aureus* develops an alternative, *ica*-independent biofilm in the absence of the *arlRS* two-component system. *J Bacteriol* **187**: 5318–5329.

Tramonti, A., De Canio, M., and De Biase, D. (2008) GadX/GadW-dependent regulation of the *Escherichia coli* acid fitness island: transcriptional control at the *gadY-gadW* divergent promoters and identification of four novel 42 bp GadX/GadW-specific binding sites. *Mol Microbiol* **70**: 965–982.

Trefflich, S., Dalmolin, R.J.S., Ortega, J.M., and Castro, M.A.A. (2019) Which came first, the transcriptional regulator or its target genes? An evolutionary perspective into the construction of eukaryotic regulons. *Biochim Biophys Acta Gene Regul Mech* 194472.

Updegrove, T.B., Shabalina, S.A., and Storz, G. (2015) How do base-pairing small RNAs evolve? *FEMS Microbiol Rev* **39**: 379–391.

Valentin-Hansen, P., Eriksen, M., and Udesen, C. (2004) The bacterial Sm-like protein Hfq: a key player in RNA transactions. *Mol Microbiol* **51**: 1525–1533.

Valle, J., Toledo-Arana, A., Berasain, C., Ghigo, J.-M., Amorena, B., Penadés, J.R., and Lasa, I. (2003) SarA and not SigmaB is essential for biofilm development by *Staphylococcus aureus. Mol Microbiol* **48**: 1075–1087.

Van Assche, E., Van Puyvelde, S., Vanderleyden, J., and Steenackers, H.P. (2015) RNA-binding proteins involved in post-transcriptional regulation in bacteria. *Front Microbiol* **6**: 141.

Vandecraen, J., Chandler, M., Aertsen, A., and Van Houdt, R. (2017) The impact of insertion sequences on bacterial genome plasticity and adaptability. *Crit Rev Microbiol* **43**: 709–730.

Vlasova, I.A., and Bohjanen, P.R. (2008) Posttranscriptional regulation of gene networks by GU-rich elements and CELF proteins. *RNA Biol* **5**: 201–207.

Wagner, E.G.H. (2009) Kill the messenger: bacterial antisense RNA promotes mRNA decay. *Nat Struct Mol Biol* **16**: 804–806.

Wagner, E.G.H., and Romby, P. (2015) Small RNAs in bacteria and archaea: who they are, what they do, and how they do it. *Adv Genet* **90**: 133–208.

Walter, N.G., and Engelke, D.R. (2002) Ribozymes: catalytic RNAs that cut things, make things, and do odd and useful jobs. *Biologist (London)* **49**: 199–203.

Wang, W., and Bechhofer, D.H. (1997) *Bacillus subtilis* RNase III gene: cloning, function of the gene in *Escherichia coli,* and construction of *Bacillus subtilis* strains with altered *rnc* loci. *J Bacteriol* **179**: 7379–7385.

Wang, Y., Li, Y., Toth, J.I., Petroski, M.D., Zhang, Z., and Zhao, J.C. (2014) N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. *Nat Cell Biol* **16**: 191–198.

Waters, L.S., and Storz, G. (2009) Regulatory RNAs in bacteria. *Cell* **136**: 615–628.

Weilbacher, T., Suzuki, K., Dubey, A.K., Wang, X., Gudapaty, S., Morozov, I., *et al.* (2003) A novel sRNA component of the carbon storage regulatory system of *Escherichia coli*. *Mol Microbiol* **48**: 657–670.

Wilkie, G.S., Dickson, K.S., and Gray, N.K. (2003) Regulation of mRNA translation by 5'- and 3-'UTR-binding factors. *Trends Biochem Sci* **28**: 182–188.

Xue, L., Chen, Y.Y., Yan, Z., Lu, W., Wan, D., and Zhu, H. (2019) Staphyloxanthin: a potential target for antivirulence therapy. *Infect Drug Resist* **12**: 2151–2160.

Yang, M.-K., Yang, Y.-H., Chen, Z., Zhang, J., Lin, Y., Wang, Y., *et al.* (2014) Proteogenomic analysis and global discovery of posttranslational modifications in prokaryotes. *Proc Natl Acad Sci USA* **111**: E5633–42.

Zhao, J.-P., Zhu, H., Guo, X.-P., and Sun, Y.-C. (2018) AU-rich long 3' untranslated region regulates gene expression in bacteria. *Front Microbiol* **9**: 3080.

Zhu, H., Mao, X.-J., Guo, X.-P., and Sun, Y.-C. (2016) The *hmsT* 3' untranslated region mediates c-di-GMP metabolism and biofilm formation in *Yersinia pestis*. *Mol Microbiol* **99**: 1167–1178.

Zhu, Y., Nandakumar, R., Sadykov, M.R., Madayiputhiya, N., Luong, T.T., Gaupp, R., *et al.* (2011) RpiR homologues may link *Staphylococcus aureus* RNAIII synthesis and pentose phosphate pathway regulation. *J Bacteriol* **193**: 6187–6196.

Ziebuhr, W., Krimmer, V., Rachid, S., Lössner, I., Götz, F., and Hacker, J. (1999) A novel mechanism of phase variation of virulence in *Staphylococcus epidermidis*: evidence for control of the polysaccharide intercellular adhesin synthesis by alternating insertion and excision of the insertion sequence element IS256. *Mol Microbiol* **32**: 345–356.

Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**: 3406–3415.

Zühlke, D., Dörries, K., Bernhardt, J., Maaß, S., Muntel, J., Liebscher, V., *et al.* (2016) Costs of life - Dynamics of the protein inventory of *Staphylococcus aureus* during anaerobiosis. *Sci Rep* **6**: 28172.

ANNEXES

Annex 1. Strains used in this study

Strains	Relevant characteristic(s)	BGR ID[a]	Source or reference
Staphylococcus argenteus			
MSHR1132	Strain isolated from blood culture of a 55-year-old indigenous Australian female (DMS Nº 28299).	314	Leibniz Institute DSMZ, (Holt *et al.*, 2011)
Staphylococcus simiae			
CCM 7213T	Strain isolated from faeces of a South American squirrel monkey.	315	CCM, (Suzuki *et al.*, 2012)
Staphylococcus capitis			
SK14	Strain isolated from normal skin of the right arm of a 58-year-old man. Catalog No. HM-117.	317	BEI Resources
Staphylococcus epidermidis			
RP62A	Strain isolated from a catheter sepsis (Ref. ATCC 35984).	313	NARSA-BEI Resources
Staphylococcus aureus			
RN4220	A cloning intermediate also used in virulence, resistance, and metabolic studies.	13	(Nair *et al.*, 2011)
N315	Hospital-acquired MRSA (methicillin resistant *Staphylococcus aureus*) strain.	15	(Kuroda *et al.*, 2001)
15981	MSSA (methicillin sensitive *Staphylococcus aureus*) clinical isolate from an otitis infection; biofilm positive; PIA-PNAG-dependent biofilm matrix.	8	(Valle *et al.*, 2003)
MW2	Typical community-acquired MRSA strain isolated in 1998 in North Dakota, USA.	10	(Baba *et al.*, 2002)
RN4220 p^{3xF}RpiRc	RN4220 carrying the p^{3xF}RpiRc plasmid.	551	This study
RN4220 p^{3xF}RpiRcΔ3'UTR	RN4220 carrying the p^{3xF}RpiRcΔ3'UTR plasmid.	552	This study
RN4220 p^{3xF}SarR	RN4220 carrying the p^{3xF}SarR plasmid.	543	This study
RN4220 p^{3xF}SarRΔ3'UTR	RN4220 carrying the p^{3xF}SarRΔ3'UTR plasmid.	544	This study
RN4220 p^{3xF}FtnA	RN4220 carrying the p^{3xF}FtnA plasmid.	541	This study
RN4220 p^{3xF}FtnAΔ3'UTR	RN4220 carrying the p^{3xF}FtnAΔ3'UTR plasmid.	542	This study
RN4220 p^{3xF}RecA	RN4220 carrying the p^{3xF}RecA plasmid.	538	This study
RN4220 p^{3xF}RecAΔ3'UTR	RN4220 carrying the p^{3xF}RecAΔ3'UTR plasmid.	539	This study
RN4220 p^{3xF}AtlR	RN4220 carrying the p^{3xF}AtlR plasmid.	737	This study
RN4220 p^{3xF}AtlRΔ3'UTR	RN4220 carrying the p^{3xF}AtlRΔ3'UTR plasmid.	738	This study
RN4220 p^{3xF}PerR	RN4220 carrying the p^{3xF}PerR plasmid.	599	This study
RN4220 p^{3xF}PerRΔ3'UTR	RN4220 carrying the p^{3xF}PerRΔ3'UTR plasmid.	600	This study
15981 p^{3xF}FtnA	15981 carrying the p^{3xF}FtnA plasmid.	793	This study
15981 p^{3xF}FtnAΔ3'UTR	15981 carrying the p^{3xF}FtnAΔ3'UTR plasmid.	794	This study
15981 p^{3xF}FtnAΔ3'UTR$^{57-93}$	15981 carrying the p^{3xF}FtnAΔ3'UTR$^{57-93}$ plasmid.	1657	This study
15981 pCN40	15981 carrying the pCN40 plasmid.	559	(Ruiz de los Mozos *et al.*, 2013)

Continued in the following page

Annex1. Continued

Strains	Relevant characteristic(s)	BGR ID[a]	Source or reference
15981 pIcaRm	15981 carrying the pIcaRm plasmid.	63	(Ruiz de los Mozos *et al.*, 2013)
15981 pIcaRmΔ3'UTR	15981 carrying the pIcaRmΔ3'UTR plasmid.	395	(Ruiz de los Mozos *et al.*, 2013)
15981 pIcaR+3'UTR^{Sarg}	15981 carrying the pIcaR+3'UTR^{Sarg} plasmid.	1395	This study
15981 pIcaR+3'UTR^{Ssim}	15981 carrying the pIcaR+3'UTR^{Ssim} plasmid.	1397	This study
15981 pIcaR+3'UTR^{Sepi}	15981 carrying the pIcaR+3'UTR^{Sepi} plasmid.	1396	This study
15981 pIcaR+3'UTR^{Scap}	15981 carrying the pIcaR+3'UTR^{Scap} plasmid.	1518	This study
15981 pGFP-3UTR-*ftnA*	15981 carrying the pGFP-3UTR-ftnA plasmid.	1644	This study
15981 pGFP-Δ3UTR-*ftnA*	15981 carrying the pGFP-Δ3UTR-*ftnA* plasmid.	1809	This study
15981 pGFP-3'UTR-*rpiRc*	15981 carrying the pGFP-3'UTR-*rpiRc* plasmid.	1643	This study
15981 pGFP-3'UTR-*rpiRc*+IS256	15981 carrying the pGFP-3'UTR-*rpiRc*+IS256 plasmid.	1915	This study
15981 pGFP-3'UTR-*rpiRc*+IS1181	15981 carrying the pGFP-3'UTR-*rpiRc*+IS1181 plasmid.	1916	This study
15981 pGFP-Δ3'UTR-*rpiRc*	15981 carrying the pGFP-Δ3'UTR-*rpiRc* plasmid.	1808	This study
15981 Δ*icaRADBC*	15981 mutant with the whole *icaR* and *icaADBC* locus deleted.	225	J. Valle/I. Lasa
Δ*icaRADBC* p^{3xF}IcaRm	15981 Δ*icaRADBC* carrying the p^{3xF}IcaRm plasmid.	1853	This study
Δ*icaRADBC* p^{3xF}IcaRmΔ3'UTR	15981 Δ*icaRADBC* carrying the p^{3xF}IcaRmΔ3'UTR plasmid.	1854	This study
Δ*icaRADBC* p^{3xF}IcaR+3'UTR^{Sarg}	15981 Δ*icaRADBC* carrying the p^{3xF}IcaR+3'UTR^{Sarg} plasmid.	1855	This study
Δ*icaRADBC* p^{3xF}IcaR+3'UTR^{Ssim}	15981 Δ*icaRADBC* carrying the p^{3xF}IcaR+3'UTR^{Ssim} plasmid.	1856	This study
Δ*icaRADBC* p^{3xF}IcaR+3'UTR^{Sepi}	15981 Δ*icaRADBC* carrying the p^{3xF}IcaR+3'UTR^{Sepi} plasmid.	1857	This study
Δ*icaRADBC* p^{3xF}IcaR+3'UTR^{Scap}	15981 Δ*icaRADBC* carrying the p^{3xF}IcaR+3'UTR^{Scap} plasmid.	1858	This study
15981 Δ*ftnA*	15981 carrying a chromosomal deletion of *ftnA* gene.	933	This study
Δ*ftnA* p^{3xF}FtnA	15981 Δ*ftnA* carrying the p^{3xF}FtnA plasmid.	1831	This study
Δ*ftnA* p^{3xF}FtnAΔ3'UTR	15981 Δ*ftnA* carrying the p^{3xF}FtnAΔ3'UTR plasmid.	1832	This study
Δ*ftnA* p^{3xF}FtnA+3'UTR^{Sarg}	15981 Δ*ftnA* carrying the p^{3xF}FtnA+3'UTR^{Sarg} plasmid.	1833	This study
Δ*ftnA* p^{3xF}FtnA+3'UTR^{Ssim}	15981 Δ*ftnA* carrying the p^{3xF}FtnA+3'UTR^{Ssim} plasmid.	1834	This study
Δ*ftnA* p^{3xF}FtnA+3'UTR^{Sepi}	15981 Δ*ftnA* carrying the p^{3xF}FtnA+3'UTR^{Sepi} plasmid.	1835	This study
Δ*ftnA* p^{3xF}FtnA+3'UTR^{Scap}	15981 Δ*ftnA* carrying the p^{3xF}FtnA+3'UTR^{Scap} plasmid.	1836	This study
15981 Δ*rpiRc*	15981 carrying a deletion *rpiRc gene.*	920	This study
Δ*rpiRc* p^{3xF}RpiRc	15981 Δ*rpiRc* carrying the p^{3xF}RpiRc plasmid.	1681	This study

Continued in the following page

Annex 1. Continued

Strains	Relevant characteristic(s)	BGR ID[a]	Source or reference
Δ*rpiRc* p^{3xF}RpiRcΔ3'UTR	15981 Δ*rpiRc* with the p^{3xF}RpiRcΔ3'UTR plasmid.	1684	This study
Δ*rpiRc* p^{3xF}RpiRc+3'UTR^{Sarg}	15981 Δ*rpiRc* with the p^{3xF}RpiRc+3'UTR^{Sarg} plasmid.	1827	This study
Δ*rpiRc* p^{3xF}RpiRc+3'UTR^{Ssim}	15981 Δ*rpiRc* with the p^{3xF}RpiRc+3'UTR^{Ssim} plasmid.	1828	This study
Δ*rpiRc* p^{3xF}RpiRc+3'UTR^{Sepi}	15981 Δ*rpiRc* with the p^{3xF}RpiRc+3'UTR^{Sepi} plasmid.	1829	This study
Δ*rpiRc* p^{3xF}RpiRc+3'UTR^{Scap}	15981 Δ*rpiRc* with the p^{3xF}RpiRc+3'UTR^{Scap} plasmid.	1830	This study
Δ*rpiRc* p^{3xF}RpiRc+3'UTR^{IS256}	15981 Δ*rpiRc* with the p^{3xF}RpiRc+3'UTR^{IS256} plasmid.	1682	This study
Δ*rpiRc* p^{3xF}RpiRc+3'UTR^{IS1181}	15981 Δ*rpiRc* with the p^{3xF}RpiRc+3'UTR^{IS1181} plasmid.	1683	This study
15981 Δ*rnc*	15981 with a deletion of the *rnc* gene.	1760	This study
Δ*rnc* p^{3xF}FtnA	15981 Δ*rnc* carrying the p^{3xF}FtnA plasmid.	1771	This study
Δ*rnc* p^{3xF}FtnAΔ3'UTR	15981 Δ*rnc* carrying the p^{3xF}FtnAΔ3'UTR plasmid.	1772	This study
Δ*rnc* pGFP-3UTR-*ftnA*	15981 Δ*rnc* carrying the pGFP-3UTR-*ftnA* plasmid.	1774	This study
15981 Δ*pnpA*	15981 with a deletion of the *pnpA* gene.	242	A. Toledo-Arana
Δ*pnpA* p^{3xF}FtnA	15981 Δ*pnpA* carrying the p^{3xF}FtnA plasmid.	1628	This study
Δ*pnpA* p^{3xF}FtnAΔ3'UTR	15981 Δ*pnpA* carrying the p^{3xF}FtnAΔ3'UTR plasmid.	1629	This study
Δ*pnpA* pGFP-3UTR-*ftnA*	15981 Δ*pnpA* carrying the pGFP-3UTR-*ftnA* plasmid.	1646	This study
15981 Δ*rnr*	15981 with a deletion of the *rnr* gene.	243	I. Lasa
Δ*rnr* p^{3xF}FtnA	15981 Δ*rnr* carrying the p^{3xF}FtnA plasmid.	1630	This study
Δ*rnr* p^{3xF}FtnAΔ3'UTR	15981 Δ*rnr* carrying the p^{3xF}FtnAΔ3'UTR plasmid.	1631	This study
15981 Δ*mrnc*	15981 with a deletion of the *mrnc* gene.	1762	This study
Δ*mrnc* p^{3xF}FtnA	15981 Δ*mrnc* carrying the p^{3xF}FtnA plasmid.	1777	This study
Δ*mrnc* p^{3xF}FtnAΔ3'UTR	15981 Δ*mrnc* carrying the p^{3xF}FtnAΔ3'UTR plasmid.	1778	This study
15981 Δ*rny*	15981 with a deletion of the *rny* gene.	1761	This study
Δ*rny* p^{3xF}FtnA	15981 Δ*rny* carrying the p^{3xF}FtnA plasmid.	1783	This study
Δ*rny* p^{3xF}FtnAΔ3'UTR	15981 Δ*rny* carrying the p^{3xF}FtnAΔ3'UTR plasmid.	1784	This study
15981 Δ*rnjA*	15981 with a deletion of the *rnjA* gene.	1768	This study
Δ*rnjA* p^{3xF}FtnA	15981 Δ*rnjA* carrying the p^{3xF}FtnA plasmid.	1797	This study
Δ*rnjA* p^{3xF}FtnAΔ3'UTR	15981 Δ*rnjA* carrying the p^{3xF}FtnAΔ3'UTR plasmid.	1798	This study
15981 *ftnA*Δ3'UTR	15981 carrying a deletion of the *ftnA* 3'UTR.	931	This study
MW2 *rpiRc*Δ3'UTR	MW2 carrying a chromosomal deletion of the *rpiRc* 3'UTR.	918	This study
MW2 *rpiRc*-3'UTR+IS1181	MW2 carrying IS1181 inserted in the 3'UTR of the *rpiRc* gene.	1922	This study
MW2 Δ*rpiRc*	MW2 carrying a deletion of the *rpiRc* gene.	934	This study

[a] Identification number of the strains stored at the Laboratory of Bacterial Gene Regulation, IdAB-CSIC.

Annex 2. Plasmids used in this study

Plasmids	Relevant characteristic(s)	Source and/or reference
pJET 1.2/blunt	*E. coli* cloning vector carrying a lethal restriction enzyme that becomes active when plasmid recircularizes. Amp^R.	Thermo Scientific
pGEM-T Easy	*E. coli* cloning vector carrying *lacZ* gene for colony screening. Amp^R.	Promega
pCN40	*E. coli-S. aureus* shuttle vector to express genes under the control of the P*blaZ* promoter. Low copy number. Amp^R-Em^R.	NARSA-BEI Resources (Charpentier *et al.*, 2004)
pCN47	*E. coli-S. aureus* shuttle vector for cloning genes under the control of its own promoter. Low copy number. Amp^R-Em^R.	NARSA-BEI Resources (Charpentier *et al.*, 2004)
pEW	A derivative pCN40 plasmid including the transcriptional terminator region of the pCN47 plasmid downstream of the multiple cloning site.	This study
pCN51	*E. coli-S. aureus* shuttle vector to express genes under the control of the P*cad* promoter. Low copy number. Amp^R-Em^R.	(Charpentier *et al.*, 2004)
pAD-cGFP	*Listeria monocytogenes* plasmid carrying the GFP gene with the 5'UTR from *hly* gene under the control of the P*hyper* promoter.	(Balestrino *et al.*, 2010)
pMAD	*E. coli-S. aureus* shuttle vector with a thermosensitive origin of replication for Gram-positive bacteria. The vector contains the *bgaB* gene encoding a β-galactosidase under the control of a constitutive promoter as reporter of plasmid presence. Amp^R, Erm^R.	(Arnaud *et al.*, 2004)
pMAD-ΔftnA	pMAD plasmid containing the allele for deletion of *ftnA* gene.	This study
pMAD-Δ3'UTR *ftnA*	pMAD plasmid containing the allele for deletion of the 3'UTR of *ftnA* gene.	This study
pMAD-Δ*rnc*	pMAD plasmid containing the allele for deletion of *rnc* gene.	This study
pMAD-Δ*mrnc*	pMAD plasmid containing the allele for deletion of *mrnc* gene.	This study
pMAD-Δ*rny*	pMAD plasmid containing the allele for deletion of *rny* gene.	This study
pMAD-Δ*rnjA*	pMAD plasmid containing the allele for deletion of *rnjA* gene.	This study
pMAD-Δ3'UTR *rpiRc*	pMAD plasmid containing the allele for deletion of the 3'UTR of *rpiRc* gene.	This study
pMAD-*rpiRc*-3'UTR+IS1181	pMAD plasmid containing the allele for the insertion of IS1181 in the 3'UTR of *rpiRc* gene.	This study
pMAD-Δ*rpiRc*	pMAD plasmid containing the allele for deletion of *rpiRc* gene.	This study
p^{3xF}RpiRc	pEW plasmid expressing the 3xFLAG-tagged *rpiRc* mRNA.	This study
p^{3xF}RpiRcΔ3'UTR	pEW plasmid expressing the 3xFLAG-tagged *rpiRc* mRNA lacking the 3'UTR while preserving the transcriptional terminator.	This study
p^{3xF}SarR	pEW plasmid expressing the 3xFLAG-tagged *sarR* mRNA.	This study
p^{3xF}SarRΔ3'UTR	pEW plasmid expressing the 3xFLAG-tagged *sarR* mRNA lacking the 3'UTR while preserving the transcriptional terminator.	This study

Continued in the following page

Annex 2. Continued

Plasmids	Relevant characteristic(s)	Source and/or reference
p^{3xF}FtnA	pEW plasmid expressing the 3xFLAG-tagged *ftnA* mRNA.	This study
p^{3xF}FtnAΔ3'UTR	pEW plasmid expressing the 3xFLAG-tagged *ftnA* mRNA lacking the 3'UTR while preserving the transcriptional terminator.	This study
p^{3xF}RecA	pEW plasmid expressing the 3xFLAG-tagged *recA* mRNA.	This study
p^{3xF}RecAΔ3'UTR	pEW plasmid expressing the 3xFLAG-tagged *recA* mRNA lacking the 3'UTR while preserving the transcriptional terminator.	This study
p^{3xF}AtlR	pCN51 plasmid expressing the 3xFLAG-tagged *atlR* mRNA.	This study
p^{3xF}AtlRΔ3'UTR	pCN51 plasmid expressing the 3xFLAG-tagged *atlR* mRNA lacking the 3'UTR while preserving the transcriptional terminator.	This study
p^{3xF}PerR	pEW plasmid expressing the 3xFLAG-tagged *perR* mRNA.	This study
p^{3xF}PerRΔ3'UTR	pEW plasmid expressing the 3xFLAG-tagged *perR* mRNA lacking the 3'UTR while preserving the transcriptional terminator.	This study
p^{3xF}IcaRm	pCN40 plasmid expressing the 3xFLAG-tagged *icaR* mRNA from *S. aureus* 132 strain.	(Ruiz de los Mozos *et al.*, 2013)
p^{3xF}IcaRmΔ3'UTR	pCN40 plasmid expressing the 3xFLAG-tagged *icaR* mRNA lacking the 3'UTR while preserving the transcriptional terminator.	(Ruiz de los Mozos *et al.*, 2013)
p^{3xF}IcaR+3'UTR^{Sarg}	pCN40 plasmid expressing the chimeric *icaR* mRNA that comprises the *S. aureus* 3xFLAG-tagged IcaR CDS and the *S. argenteus icaR* 3'UTR.	This study
p^{3xF}IcaR+3'UTR^{Ssim}	pCN40 plasmid expressing the chimeric *icaR* mRNA that comprises the *S. aureus* 3xFLAG-tagged IcaR CDS and the *S. simiae icaR* 3'UTR.	This study
p^{3xF}IcaR+3'UTR^{Sepi}	pCN40 plasmid expressing the chimeric *icaR* mRNA that comprises the *S. aureus* 3xFLAG-tagged IcaR CDS and the *S. epidermidis icaR* 3'UTR.	This study
p^{3xF}IcaR+3'UTR^{Scap}	pCN40 plasmid expressing the chimeric *icaR* mRNA that comprises the *S. aureus* 3xFLAG-tagged IcaR CDS and the *S. capitis icaR* 3'UTR.	This study
pIcaRm	pCN40 plasmid expressing the wild type *icaR* mRNA from *S. aureus* 132 strain.	(Ruiz de los Mozos *et al.*, 2013)
pIcaRmΔ3'UTR	pCN40 plasmid expressing the *icaR* mRNA lacking the 3'UTR while preserving the transcriptional terminator.	(Ruiz de los Mozos *et al.*, 2013)
pIcaR+3'UTR^{Sarg}	pCN40 plasmid expressing the chimeric *icaR* mRNA that comprises the *S. aureus* IcaR CDS and the *S. argenteus icaR* 3'UTR.	This study
pIcaR+3'UTR^{Ssim}	pCN40 plasmid expressing the chimeric *icaR* mRNA that comprises the *S. aureus* IcaR CDS and the *S. simiae icaR* 3'UTR.	This study
pIcaR+3'UTR^{Sepi}	pCN40 plasmid expressing the chimeric *icaR* mRNA that comprises the *S. aureus* IcaR CDS and the *S. epidermidis icaR* 3'UTR.	This study
pIcaR+3'UTR^{Scap}	pCN40 plasmid expressing the chimeric *icaR* mRNA that comprises the *S. aureus* IcaR CDS and the *S. capitis icaR* 3'UTR.	This study

Continued in the following page

Annex 2. Continued

Plasmids	Relevant characteristic(s)	Source and/or reference
p^{3xF}FtnA+3'UTR^{Sarg}	pEW plasmid expressing the chimeric *ftnA* mRNA that comprises the *S. aureus* FtnA CDS carrying the 3xFLAG tag at its N-terminus and the *S. argenteus ftnA* 3'UTR.	This study
p^{3xF}FtnA+3'UTR^{Ssim}	pEW plasmid expressing the chimeric *ftnA* mRNA that comprises the *S. aureus* FtnA CDS carrying the 3xFLAG tag at its N-terminus and the *S. simiae ftnA* 3'UTR.	This study
p^{3xF}FtnA+3'UTR^{Sepi}	pEW plasmid expressing the chimeric *ftnA* mRNA that comprises the *S. aureus* FtnA CDS carrying the 3xFLAG tag at its N-terminus and the *S. epidermidis ftnA* 3'UTR.	This study
p^{3xF}FtnA+3'UTR^{Scap}	pEW plasmid expressing the chimeric *ftnA* mRNA that comprises the *S. aureus* FtnA CDS carrying the 3xFLAG tag at its N-terminus and the *S. capitis ftnA* 3'UTR.	This study
p^{3xF}FtnAΔ3'UTR$^{57-93}$	pEW plasmid expressing the 3xFLAG-tagged *ftnA* mRNA lacking a region of the 3'UTR including nt 57-93 after the stop codon while preserving the transcriptional terminator.	This study
p^{3xF}RpiRc+3'UTR^{Sarg}	pEW plasmid expressing the chimeric *rpiRc* mRNA that comprises the *S. aureus* RpiRc CDS carrying the 3xFLAG tag at its N-terminus and the *S. argenteus rpiRc* 3'UTR.	This study
p^{3xF}RpiRc+3'UTR^{Ssim}	pEW plasmid expressing the chimeric *rpiRc* mRNA that comprises the *S. aureus* RpiRc CDS carrying the 3xFLAG tag at its N-terminus and the *S. simiae rpiRc* 3'UTR.	This study
p^{3xF}RpiRc+3'UTR^{Sepi}	pEW plasmid expressing the chimeric *rpiRc* mRNA that comprises the *S. aureus* RpiRc CDS carrying the 3xFLAG tag at its N-terminus and the *S. epidermidis rpiRc* 3'UTR.	This study
p^{3xF}RpiRc+3'UTR^{Scap}	pEW plasmid expressing the chimeric *rpiRc* mRNA that comprises the *S. aureus* RpiRc CDS carrying the 3xFLAG tag at its N-terminus and the *S. capitis rpiRc* 3'UTR.	This study
p^{3xF}RpiRc+3'UTR^{IS256}	pEW plasmid expressing the chimeric *rpiRc* mRNA that comprises the *S. aureus* 3xFLAG-tagged RpiRc CDS at its N-terminus and the IS256 inserted in the 3'UTR.	This study
p^{3xF}RpiRc+3'UTR^{IS1181}	pEW plasmid expressing the chimeric *rpiRc* mRNA that comprises the *S. aureus* 3xFLAG-tagged RpiRc CDS at its N-terminus and the IS1181 inserted in the 3'UTR.	This study
pGFP-3'UTR-*ftnA*	pEW plasmid expressing a chimeric mRNA including the *gfp* gene fused to the 3'UTR of *ftnA*.	This study
pGFP-Δ3'UTR-*ftnA*	pEW plasmid expressing a chimeric mRNA including the *gfp* gene fused to the transcriptional terminator of the *ftnA* mRNA.	This study
pGFP-3'UTR-*rpiRc*	pEW plasmid expressing a chimeric mRNA including the *gfp* gene fused to the 3'UTR of *rpiRc*.	This study
pGFP-3'UTR-*rpiRc*+IS256	pEW plasmid expressing a chimeric mRNA including the *gfp* gene fused to the 3'UTR of *rpiRc* carrying the IS256 insertion.	This study
pGFP-3'UTR-*rpiRc*+IS1181	pEW plasmid expressing a chimeric mRNA including the *gfp* gene fused to the 3'UTR of *rpiRc* carrying the IS1181 insertion.	This study
pGFP-Δ3'UTR-*rpiRc*	pEW plasmid expressing a chimeric mRNA including the *gfp* gene fused to the transcriptional terminator of the *rpiRc* mRNA.	This study

Annex 3. Primers used in this study

Oligonucleotide name	Sequence [a]
Simultaneous mapping of 5' and 3' ends (mRACE)	
RACE-icaR-1	TATCATCAAGTGTTGTACCGTCAT
RACE-icaR-2	TCAAAGATGAAGTGTATTCGCTAC
RACE-icaR-ssim-A	CACACTTTTAGCTATATCATCA
RACE-icaR-ssim-B	ATAGCCGTGTAATATATTTGTAAT
RACE-icaR-sepi-A	GACTAGCCTTTTTTATATTTACAC
RACE-icaR-sepi-B	AATAATTTTTGTAACTAGTATGTAAC
RACE-icaR-scap-A	CACTTTTAGCGATATCATCAAG
RACE-icaR-scap-B	AAATAATTTTGGTAACTAGTATGTAA
RACE-ftnA-sau-A	GTACGATTCTTTATCACAGTATGC
RACE-ftnA-sau-B	CACATCAATTATTTAACTCGTATC
RACE-ftnA-ssim-A	GAAACCTTCATAAGATTCTTTATCAC
RACE-ftnA-ssim-B	GTGACGATAGTAATGCACTTTAC
RACE-ftnA-sepi-A	CATAGATTTTTTTACCGTGGAAAC
RACE-ftnA-sepi-B	CATATAGATTACCTTACTCGTATTG
RACE-ftnA-scap-A	CATAAGATTCTTTATCGCAATAAG
RACE-ftnA-scap-B	GAAAGAACTTGCTAATCGCTC
RACE-RpiRc-sau-A	TTCTGTTAGTACGTTTGACAT
RACE-RpiRc-sau-B	GATAACTATCGTAAACATTTATCAAA
RACE-RpiRc-ssim-A	GTTTGACATATTTATCACTCCAGT
RACE-RpiRc-ssim-B	CAAACATTAAATGACGAGTG
RACE-3XFLAG-A	ACCGTCATGGTCTTTGTAGT
Construction of chromosomic mutant strains	
Dftn-A (BamHI)	*GGATCC*GTGTTTCCTCCTCAAATTTC
Dftn-B (NheI)	*GCTAGC*TTTTGATACACCTCTTATTTGTA
D3UTR-ftn-A (BamHI)	*GGATCC*GCAAACTTCTTCATTCAACAAG
D3UTR-ftn-B (NheI)	*GCTAGC*CTGTCTATTGTAGTGATGTTTAAT
Dftn-C (NheI)	*GCTAGC*ACGGAGATCACTAGATTCATTT
Dftn-D (EcoRI)	*GAATTC*GTAGTCAATCCTTTCAATTAATTAAATG
Dftn-E	CAATATCATCAACTTGCTCTG
Dftn-F	CAACATCTTCTGGTTGTATG
DrpiRc-A (BamHI)	*GGATCC*CAACTTGTTCAAGAAACACAT
DrpiRc-B (NheI)	*GCTAGC*CCAATTTATACATTATGTACTCATC
D3UTR-rpiRc-A (BamHI)	*GGATCC*CAACTTGTTCAAGAAACACAT
D3UTR-rpiRc-B (NheI)	*GCTAGC*AAGAATTTATATAAGACTGTTAATG
DrpiRc-C (NheI)	*GCTAGC*AACAAACGCAGTTGATTAAC
DrpiRc-D (EcoRI)	*GAATTC*GTATGGTATTTAATGTTGAAACTAG
DrpiRc-E	TCGCTACAAATATCAGATTACTATC
DrpiRc-F	CGTATACGTTTATAGTGCGTC
Drnc-A (BamHI)	*GGATCC*GGTGAATCGACGTGGAAAAT
Drnc-B (KpnI)	*GGTACC*TTCTAAAACGATTAACTATCTCAC
Drnc-C (KpnI)	*GGTACC*GATTTTAAAACACAATTCCAAGA
Drnc-D (EcoRI)	*GAATTC*AGAACACATGTATACGATATTTTAG
Drnc-E-n	CAGAATTTCTCCCTAAGAAAC
Drnc-F-n	CACCTTTATCGAATTGAACATTG
Dmrnc-A (BamHI)	*GGATCC*CACATTAAATTATTGAATCCATTG
Dmrnc-B (KpnI)	*GGTACC*GCTTCAAAATATCCATTTCTTC
Dmrnc-C (KpnI)	*GGTACC*GAACGATTAGAGGCATTATTAA
Dmrnc-D (EcoRI)	*GAATTC*CTAATTTAGATTTTGGTACAGTTTG
Dmrnc-E	GCAAGGAAAAACAAAGATTTTG
Dmrnc-F	GTACTGTCAATAAACCTTCTT
Drny-A (BamHI)	*GGATCC*CAATAGTTTTATAATCGAGCTTC
Drny_B (KpnI)	*GGTACC*CTCCAACAACTCCTAGAATGATC
Drny-C (KpnI)	*GGTACC*CGATTGGCTAGAGATATTAAAAATC
Drny-D (EcoRI)	*GAATTC*GAAAACCAATCATCTTTATAGGTTTA

Continued in the following page

Annex 3. Continued

Oligonucleotide name	Sequence [a]
Drny-E	CAAATATCCTTATAGGATTGATTG
Drny-F	CTGCAGAAGTTATAAAAGAATTAAAG
DrnjA-A (BamHI)	*GGATCC*GAGTGGGACAGAAATGA
DrnjA-B (KpnI)	*GGTACC*TCAAAAAGCTACTAACTTTGAAGT
DrnjA-C (KpnI)	*GGTACC*TTATTTAGCAATCTCCACATTA
DrnjA-D (EcoRI)	*GAATTC*GATTTAACTGAAATTTTAGTGTTATT
DrnjA-E	CAATTAAACGAGGCAAAGAG
DrnjA-F	CTCATTTAAATTTTACCGTTTCA
pMAD-1	GGAAGCGAGAAGAATCATAATG
pMAD-2	CTAGCTAATGTTACGTTAC
Construction of plasmids for screening of regulatory 3'UTRs	
SarR-A	AATTTTGTGAGCAAGCCAT*C*
+1-SarR (BamHI)	*GGATCC*TAACATAGTTGGATAGAGTTTCG
3XFLAG-SarR-izqda	CCGTCATGGTCTTTGTAGTCCATTATTAAACCACTCCTCTG
3XFLAG-SarR-drcha	GACTACAAAGACCATGACGGTGATTATAAAGATCATGATATCGACTA CAAAGATGACGACGATAAAAGTAAAATTAATGACATTAATG
D3UTR-SarR-term (EcoRI)	*GAATTC*AAAATAAAAAGACTAGTGTACCTTGTTTCAAGCTTAATAAGC TGAAATAAGGTATGGCTAGTCTTTAACTTACGCAATTAACCTTGATTT AAT
Term-SarR (EcoRI)	*GAATTC*AAAATAAAAAGACTAGTGTACCT
RecA-A	AACAGACTAATCCTACGATTG
+1-recA (BamHI)	*GGATCC*AAGATAATTAAAAGATAGCAATTTC
3XFLAG-recA-izqda	CCGTCATGGTCTTTGTAGTCCAAAGCGAGACCTCCTAATT
3XFLAG-recA-drcha	GACTACAAAGACCATGACGGTGATTATAAAGATCATGATATCGACTA CAAAGATGACGACGATAAAGATAACGATCGTCAAAAAGCT
D3UTR-recA-term (EcoRI)	*GAATTC*AAAAAACCTTATAGTCTCGATTTGTAGTGTATCCCATAAAGT TAGATATTTAGATATAAATTTGTGTACTATT
Term-recA (EcoRI)	*GAATTC*AAAAAACCTTATAGTCTCGATT
AtlR-A	ATTTATCGCCACTAGTGTAG
+1-atlR (BamHI)	*GGATCC*AAATAATATGTAGTGGAGTAGAAC
3XFLAG-atlR-izqda	CCGTCATGGTCTTTGTAGTCCATGTTGTTCTACTCCACTAC
3XFLAG-atlR-drcha	GACTACAAAGACCATGACGGTGATTATAAAGATCATGATATCGACTA CAAAGATGACGACGATAAATATAAACAACTTGAAAAACTTATTACAC
D3UTR-atlR-term (EcoRI)	*GAATTC*ATAGCGGTGATTGTTTACCATCAATTTTAACAATCACCGCT AGTGTTTGCTATTGTCGTGTTAC
Term-atlR (EcoRI)	*GAATTC*ATAGCGGTGATTGTTTACCA
PerR-A	GTAACTATAACAGCGCATATAACT
+1-PerR (BamHI)	*GGATCC*TAAGTAATAATAATTATTATATAAGAAAGATGGTG
3XFLAG-PerR-izqda	CCGTCATGGTCTTTGTAGTCCATCTATATCACCATCTTTCTTAT
3XFLAG-PerR-drcha	GACTACAAAGACCATGACGGTGATTATAAAGATCATGATATCGACTA CAAAGATGACGACGATAAAAGTGTTGAAATAGAATCAATTGAAC
D3UTR-PerR-term (EcoRI)	*GAATTC*AAAAAAAGTTAAACAACTTTCATGTCGTTTAACTTCATACTA CCAAAGTTAAATTATT
Term-PerR (EcoRI)	*GAATTC*AAAAAAAGTTAAACAACTTTCATG
Construction of plasmids expressing chimeric *icaR* mRNAs	
IcaR+1 (BamHI)	*GGATCC*GAAATATTTGTAATTGCA
IcaR-Term (EcoRI)	*GAATTC*CCTTTTAAAAAGATGTGGGTA
IcaR-MSHR1132-Term-EcoRI	*GAATTC*CATACTCTGTCCTTGATA
IcaR-Simiae-Term-EcoRI	*GAATTCC*CAAATTGCCTACCCTAGTTAC
IcaR-Sepi+1-BamHI	*GGATCC*TAATATTTGTAATTTTAACTTAATTTTTCTG
IcaR-Sepi-Term-MfeI	*CAATTG*TTTCGCGTTTATAGTAAGTTTTCA
IcaR-capitis+1-BamHI	*GGATCCGTAATATTTGTAATTTTAACTTAATTTTCC*
IcaR-capitis-Term-EcoRI	*GAATTC*TTACATCATTGACTTGTAATTAAATTGC
Construction of plasmids expressing chimeric *ftnA* mRNAs	
Ftn-A	GTGTTTCCTCCTCAAATTTC
+1-ftn (BamHI)	*GGATCC*AATTCATAGTAATTTTAATTTACAA
3XFLAG-ftn-izqda	CCGTCATGGTCTTTGTAGTCCATTTTGATACACCTCTTATTTG

Continued in the following page

Annex 3. Continued

Oligonucleotide name	Sequence [a]
3XFLAG-ftn-drcha	GACTACAAAGACCATGACGGTGATTATAAAGATCATGATATCGACTA CAAAGATGACGACGATAAATTAAGTAAAAATTTATTAGAAGCA
CDS-stop-ftn	TTATTCTTCGTCGAATGTACGA
D3UTR-ftn-term (KpnI)	*GGTACC*AAAAAACGCAGATCAATGATTCAGAAAATGAATCTAGTGAT CTCCGTCTATTGTAGTGATGTTTAATTAT
Term-ftn (KpnI)	*GGTACC*AAAAAACGCAGATCAATGAT
3UTR-ftn-term-1/2 (KpnI)	*GGTACC*AAAAAACGCAGATCAATGATTCAGAAAATGAATCTAGTGAT CTCCGTGACCCAAATGCCTATCAT
3UTR-ftn-Sarg-fw	TCGTACATTCGACGAAGAATAATTAACAAACTATATTAACAGACAGA
3UTR-ftn-Sarg-rvs (KpnI)	*GGTACC*GAAGGCCCTTAATAATAAATAGTAC
3UTR-ftn-Ssim-fw	TCGTACATTCGACGAAGAATAAATTATAATTTCATAGAGATATAATTT
3UTR-ftn-Ssim-rvs (KpnI)	*GGTACC*GACCAATGTGTTTGAGAACAC
3UTR-ftn-Sepi-fw	TCGTACATTCGACGAAGAATAATACAACAACTTATTCCTAAATATTGT
Term-ftn-Sepi (KpnI)	*GGTACC*AAAAAAACCCTTAAATCATTTATCT
3UTR-ftn-Scap-fw	TCGTACATTCGACGAAGAATAATTCGTAACTCTTGTCTTAAATTC
3UTR-ftn-Scap-rvs (KpnI)	*GGTACC*GCTCTACTTTTATAAACTAACTCTAAT
Construction of plasmids expressing chimeric *rpiRc* mRNAs	
+1-RpiRc (BamHI)	*GGATCC*GATTAAGATGGTAGATTTAGTCTT
3XFLAG-RpiRc-izqda	CCGTCATGGTCTTTGTAGTCCATATTAATCACTCCACTTTAACG
3XFLAG-RpiRc-drcha	GACTACAAAGACCATGACGGTGATTATAAAGATCATGATATCGACTA CAAAGATGACGACGATAAATCAAACGTACTAACAGAAATAGAT
CDS-stop-RpiRc	TTAATGTTTAAAGTTAATATTTGAT
D3UTR-RpiRc-term (EcoRI)	*GAATTC*AAAATAACAAACGCAATTGATATATGGAATGAATCCGTTAAT CAACTGCGTTTGTTAAGAATTTATATAAGACTGTTAATG
Term-RpiRc (EcoRI)	*GAATTC*AAAATAACAAACGCAATTGAT
3UTR-RpiRc-Sarg-fw	AATATTAACTTTAAACATTAACAGTCTTATATAACTTCTTTAAATG
3UTR-RpiRc-Sarg-rvs (EcoRI)	*GAATTC*GGATCACAATCATTGATAATCTG
3UTR-RpiRc-Ssim-fw	AATATTAACTTTAAACATTAAATGACGAGTGAAACAAATGA
3UTR-RpiRc-Ssim-rvs-2 (EcoRI)	*GAATTC*ATCATTCAGTAAATAAATAATCATCT
CDS-stop-RpiRc3UTRepi (EcoRI)	*GAATTC*AAAGCGATTCTTTCATACATTATCTTTTCTAGGATAAGTATA AAGAATCGCTTTTTATTATTTTAATGTTTAAAGTTAATATTTGAT
3UTR-RpiRc-Scap-fw	AATATTAACTTTAAACATTAAATAAATAGAAAAAGCGATTCGTC
3UTR-RpiRc-Scap-rvs (EcoRI)	*GAATTC*TAGCTATATATAGTTAAACAAAAAAGC
RpiRc-CDS-SpeI-fw	CAATTACTAGTACAAGGGATAATC
3UTR-RpiRc-42-IS256-rvs	CAGGAGTCTGGACTTGACTTAAATTATTATATGACATGATTAAAG
IS256-fw	AGTCAAGTCCAGACTCCTGT
IS256-3UTR-RpiRc-rvs	AAGACATGTAAGCTTTGAAAATGATGCATAAATTATGATAAAGTCCG TATAATTGTGT
3UTR RpiRc-IS1181-fw	TTCAAAGCTTACATGTCTTTTATCTTGTATAACATACGCATATGACCA ATGATTTGAGGTTCTCCACCAAATGTGG
IS1181-rvs	CTTCCCACCATAAAAGATGAAGAACCTGATTTGATAAGGCCATTCAA
pCN_univ_rv_AT	GTTTTGGTTCATCTTCTGTTAACTTACTAA
Construction of plasmids expressing GFP	
SalI-GFP-fw	*GTCGAC*ATAAAGCAAGCATATAATATTGC
BcuI-TT-BamHI-GFP-rvs	*ACTAGT*AAATGCCTATCCAAGAGGATAGGCATTTTGGATCCTTATTT GTATAGTTCATCCAT
BamHI-EcoRI-3UTR-ftn-fw	*GGATCCGAATTC*TTAAACATCACTACAATAGACAGAT
SmaI-3UTR-ftn-rvs	*CCCGGG*AAAAAACGCAGATCAATGAT
KpnI-D3UTR-term-ftn	*GGTACC*AAAAAACGCAGATCAATGATTCAGAAAATGAATCTAGTGAT CTCCCTGTCTATTGTAGTGATGTTTAAG
BamHI-EcoRI-3UTR-RpiRc-fw	*GGATCCGAATTC*CAGTCTTATATAAATTCTTTAATCATGT
SmaI-3UTR-RpiRc-rvs	*CCCGGG*AAAATAACAAACGCAATTGAT
KpnI-term-RpiRc	*GGTACC*AAAATAACAAACGCAATTGAT
KpnI-D3UTR-term-RpiRc	*GGTACC*AAAATAACAAACGCAATTGATATATGGAATGAATCCGTTAA TCAACTGCGTTTGTTAAGAATTTATATAAGACTGGAATT

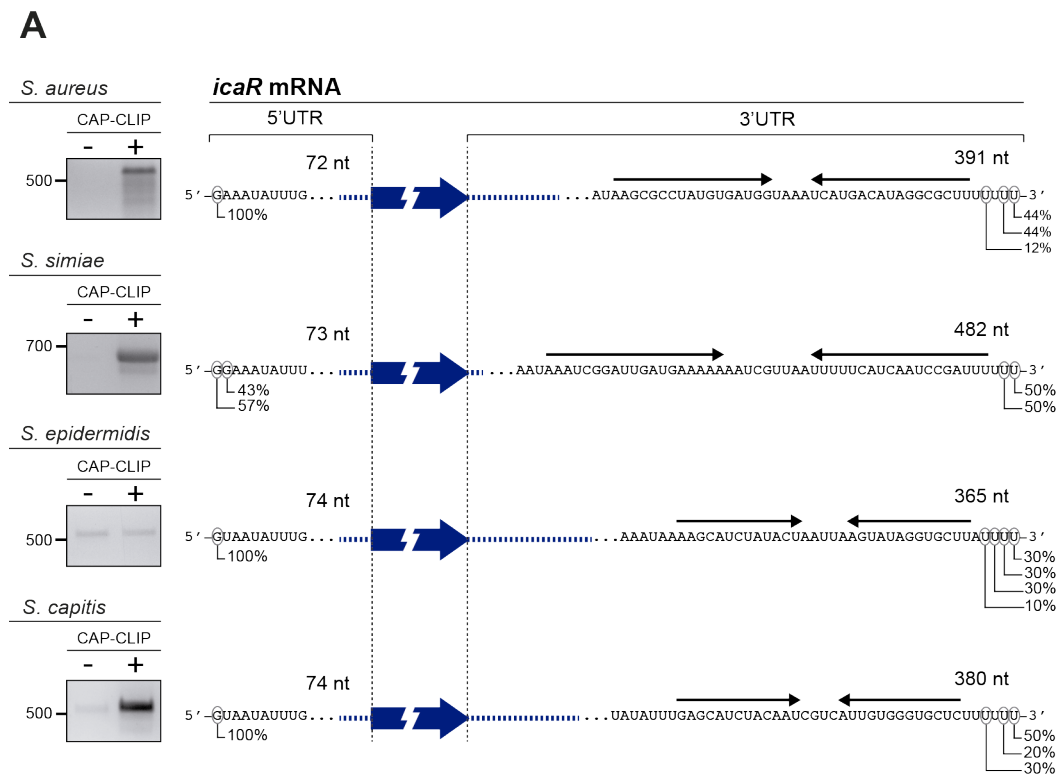
Continued in the following page

Annex 3. Continued

Oligonucleotide name	Sequence [a]
Synthesis of riboprobes	
NB-probe-icaR-fw	GAAAAAAAGTATTTACGAACAAAG
T7prom-NB-probe-rvs	**TAATACGACTCACTATAGG**GCTTTCAAATACCAACTTTCAAG
NB-probe-ftn-fw	GAAGAACGTTTCCATGGACAAAA
T7-NB-probe-ftn-rvs	**TAATACGACTCACTATAGGG**ACGAGCGCCAAGTTCTTTTTC
Riboprobe-RpiRc-fw	CTAAACTCCATACACGTACTACAC
Riboprobe-T7prom-RpiRc-rvs	**TAATACGACTCACTATAGGG**GTCATCAATAACTTTAACCATTG
***In vitro* transcription of the 5' and 3'UTRs from IcaR**	
T7-5UTR-fw	**TAATACGACTCACTATAGGG**AAATATTTGTAATTGCCAACTTAATTT
IcaR-5rev	ATCATCAAGTGTTGTACCGTCATACCCCTT
T7-3UTR IcaR-fw	**TAATACGACTCACTATAGGG**ATTTTTGAAGAAATAATTTTTGTTA
Term-IcaR-rvs	AAAAAGCGCCTATGTCATGATTTACCATCA
3UTR-IcaR-Sarg-rvs	AAAAATAAGCGTCAATGACATG
3UTR-IcaR-Ssim-rvs	TCCCAAACAAAACAAAAAATCGGATTG
3UTR-IcaR-Sepi-rvs	AAAATAAGCACCTATACTTAATTAG
3UTR-IcaR-Scap-rvs	AAAAAGAGCACCCACAATGAC

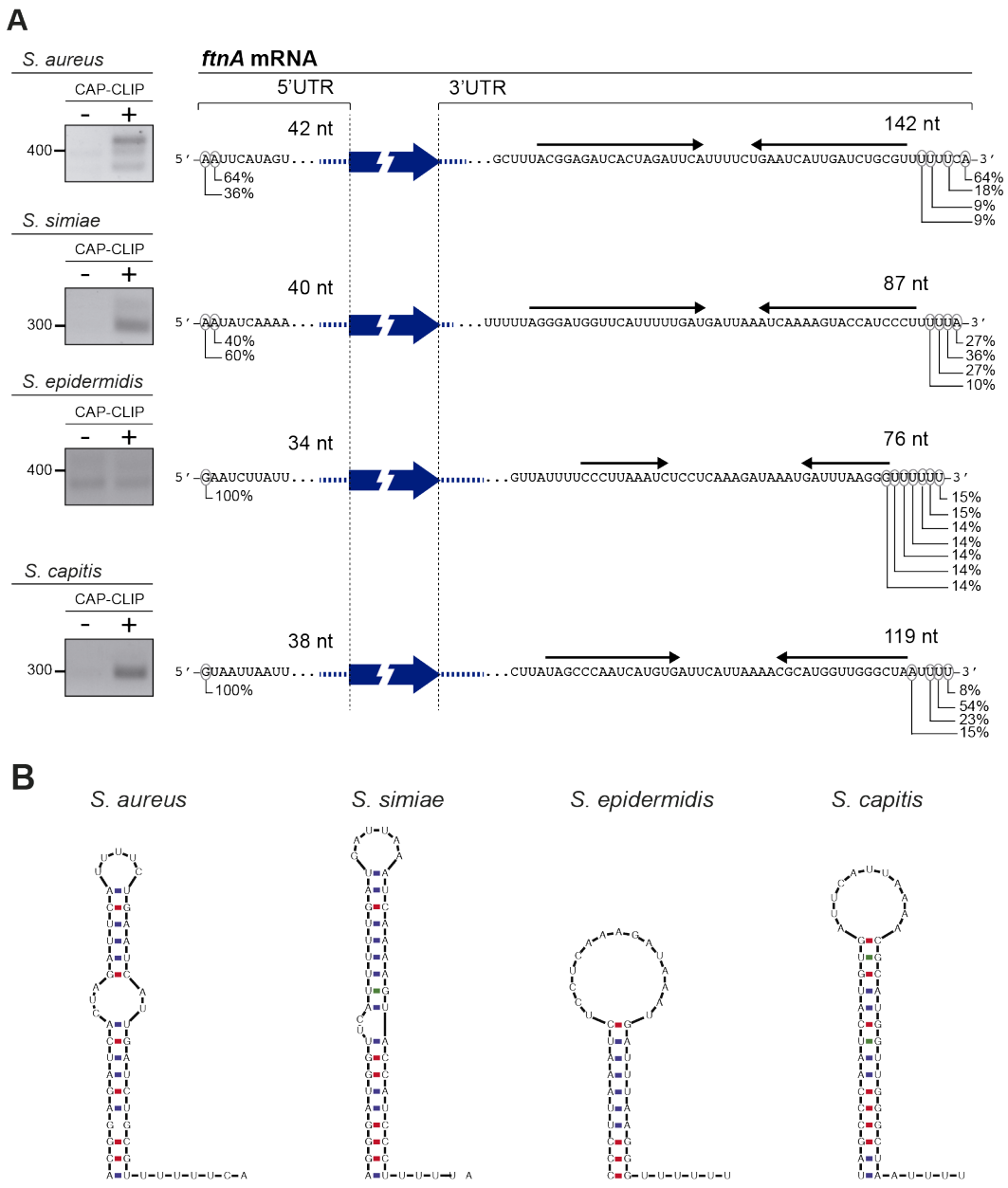
[a] Restriction enzymes sites, T7 promoter and 3xFLAG sequences included into the oligonucleotides are indicated in italic, bold and underlined format, respectively.

Annex 4. Determination of *icaR* mRNA boundaries by mRACE analysis



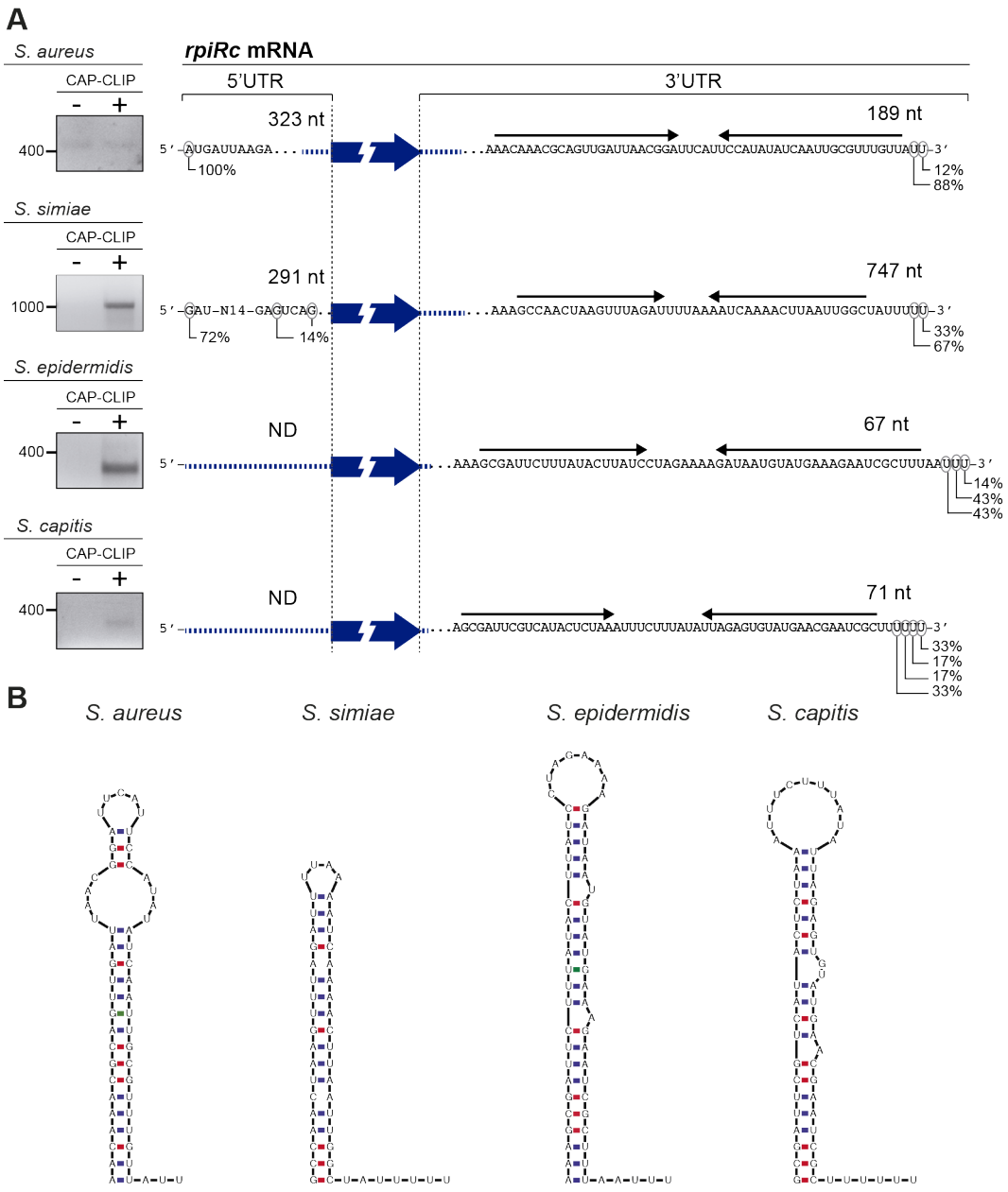
(A) Simultaneous mapping of 5' and 3' *icaR* mRNA ends by a modified rapid amplification of cDNAs ends technique (mRACE) using total RNAs extracted from *S. aureus*, *S. simiae*, *S. epidermidis* and *S. capitis* strains. mRACE RT-PCR amplification products were run in agarose gels. Representative gel images are shown. Bands corresponding to CAP-CLIP treated samples were excised from the gel and cloned into pGEM-T plasmid. Plasmids were purified from 10 bacterial clones and plasmid sequences were identified by Sanger sequencing. The 5' and 3' ends and the percentage of clones showing the same nucleotide position are indicated in the corresponding sequence. **(B)** Transcriptional terminator structures of *icaR* mRNA predicted by Mfold in the indicated staphylococcal species (Zuker, 2003).

Annex 5. Determination of *ftnA* mRNA boundaries by mRACE analysis



(A) Simultaneous mapping of 5' and 3' *ftnA* mRNA ends by a modified rapid amplification of cDNAs ends technique (mRACE) using total RNAs extracted from *S. aureus, S. simiae, S. epidermidis* and *S. capitis* strains. mRACE RT-PCR amplification products were run in agarose gels. Representative gel images are shown. Bands corresponding to CAP-CLIP treated samples were excised from the gel and cloned into pGEM-T plasmid. Plasmids were purified from 10 bacterial clones and plasmid sequences were identified by Sanger sequencing. The 5' and 3' ends and the percentage of clones showing the same nucleotide position are indicated in the corresponding sequence. **(B)** Transcriptional terminator structures of *ftnA* mRNA predicted by Mfold in the indicated staphylococcal species (Zuker, 2003).

Annex 6. Determination of *rpiRc* mRNA boundaries by mRACE analysis



(A) Simultaneous mapping of 5' and 3' *rpiRc* mRNA ends by a modified rapid amplification of cDNAs ends technique (mRACE) using total RNAs extracted from *S. aureus* and *S. simiae* WT strains and *S. aureus* Δ*rpiRc* strains carrying the p^{3xF}RpiRc+3'UTR^{Sepi} and p^{3xF}RpiRc+3'UTR^{Scap} plasmids. mRACE RT-PCR amplification products were run in agarose gels. Representative gel images are shown. Bands corresponding to CAP-CLIP treated samples were excised from the gel and cloned into pGEM-T plasmid. Plasmids were purified from 10 bacterial clones and plasmid sequences were identified by Sanger sequencing. The 5' and 3' ends and the percentage of clones showing the same nucleotide position are indicated in the corresponding sequence. **(B)** Transcriptional terminator structures of *rpiRc* mRNA predicted by Mfold in the indicated staphylococcal species (Zuker, 2003).

Annex 7. Alignment disruptions on *rpiRc* 3'UTR identify the presence of ISs

S. aureus strain	Aligned nucleotides	3'UTR position[a]	Contig IDs matched by blastn	Reason of alignment disruption	Insertion sequence[b]
C1555	1-1167 1197-1385	-27 3	LAMV01000070.1 LAMV01000029.1	Query matched two different contigs. Missing 30 nt in comparison to reference.	ND
N13/1/648	105-1197 1260-1385	3 66	NAIZ01000307.1 NAIZ01000166.1	Query matched two different contigs. Missing 63 nt in comparison to reference.	ND
Q13/1/145	1-1197 1189-1385	3 -5	NAJG01000081.1 NAJG01000005.1	Query matched two different contigs. 8 nt duplication found at the contig ends. Putative IS insertion.	IS?
ST59-1	1-1197 1189-1385	3 -5	QQPK01000081.1 QQPK01000005.1	Query matched two different contigs. 8 nt duplication found at the contig ends. Putative IS insertion.	IS?
SA-085	1-1205 1252-1385	11 58	JXIF01000064.1 JXIF01000188.1	Query matched two different contigs. Missing 47 nt in comparison to reference.	ND
UCL369	1-1208 1226-1385	14 32	LPWR01000137.1 LPWR01000151.1	Query matched two different contigs. Missing 18 nt in comparison to reference.	ND
UP72	1-1211 1200-1385	17 6	LGWY01000005.1 LGWY01000017.1	Query matched two different contigs. Inserted sequence.	←IS256←
BY2	1-1211 1200-1385	17 6	PTCK01000007.1 PTCK01000010.1	Query matched two different contigs. Inserted sequence.	←IS256←
AY20	1-1211 1200-1385	17 6	PTCM01000012.1 PTCM01000001.1	Query matched two different contigs. Inserted sequence.	←IS256←
AB4	1-1211 1200-1385	17 6	PTDD01000011.1 PTDD01000019.1	Query matched two different contigs. Inserted sequence.	←IS256←
CM95	1-1219 1202-1385	25 8	PZVS01000071.1 PZVS01000027.1	Query matched two different contigs. 17 nt duplication.	ND

Continued in the following page

Annex 7. Continued

S. aureus strain	Aligned nucleotides	3'UTR position[a]	Contig IDs matched by blastn	Reason of alignment disruption	Insertion sequence[b]
MM66RVI-4	232-1223 1329-1385	29 135	JMBU01000159.1 JMBU01000018.1	Query matched two different contigs. Missing 106 nt in comparison to reference.	ND
HU09-1102013	1-1227 1196-1385	33 2	MSRE01000085.1 MSRE01000086.1	Query matched two different contigs. Inserted sequence.	Repeated *rpiRc* 3'UTR fragment
2010-60-6511-5	1-1236 1225-1385	42 31	KK220570.1 KK220550.1	Query matched two different contigs. Inserted sequence.	←IS256←
42	1-1239 1251-1385	45 57	SULE01000232.1 SULE01000221.1	Query matched two different contigs. Missing 12 nt in comparison to reference.	ND
H8/16	1-1269 1197-1385	75 3	NKCP01000017.1 NKCP01000092.1	Query matched two different contigs. 72 nt duplication.	ND
PN235B0	1-1280 1262-1385	86 68	PDUY01000154.1 PDUY01000106.1	Query matched two different contigs. 18 nt duplication. Putative IS insertion.	IS?
DAR1183	1-1302 1295-1385	108 101	KK099086.1	Inserted sequence.	←IS1181←
07-03168	1-1314 1349-1385	120	CTTN01000002.1	NNNNs at contig sequence. Missing 35 nt.	ND
NCTC5657	1-1322 1323-1385	138 129	UHAO01000001.1	Inserted sequence (1870 nt).	ND
TUM9458	1-1326 1323-1385	132 129	AP019305.1	Inserted sequence (1023 nt).	←IS← (Similar to IS1062, IS30 family)
TUM9463	1-1326 1323-1385	132 129	AP019306.1	Inserted sequence (1063 nt).	←IS← (Similar to IS1062, IS30 family transposase)
CM154	1-1331 1324-1385	137 130	PZRV01000009.1 PZRV01000004.1	Query matched two different contigs. Inserted sequence.	←IS1181←
M1339	1-1332 1325-1385	138 131	JDEK01000007.1 JDEK01000026.1	Query matched two different contigs. Inserted sequence.	→IS1181→
M1360	1-1332 1325-1385	138 131	KI996572.1 KI996573.1	Query matched two different contigs. Inserted sequence.	→IS1181→

Continued in the following page

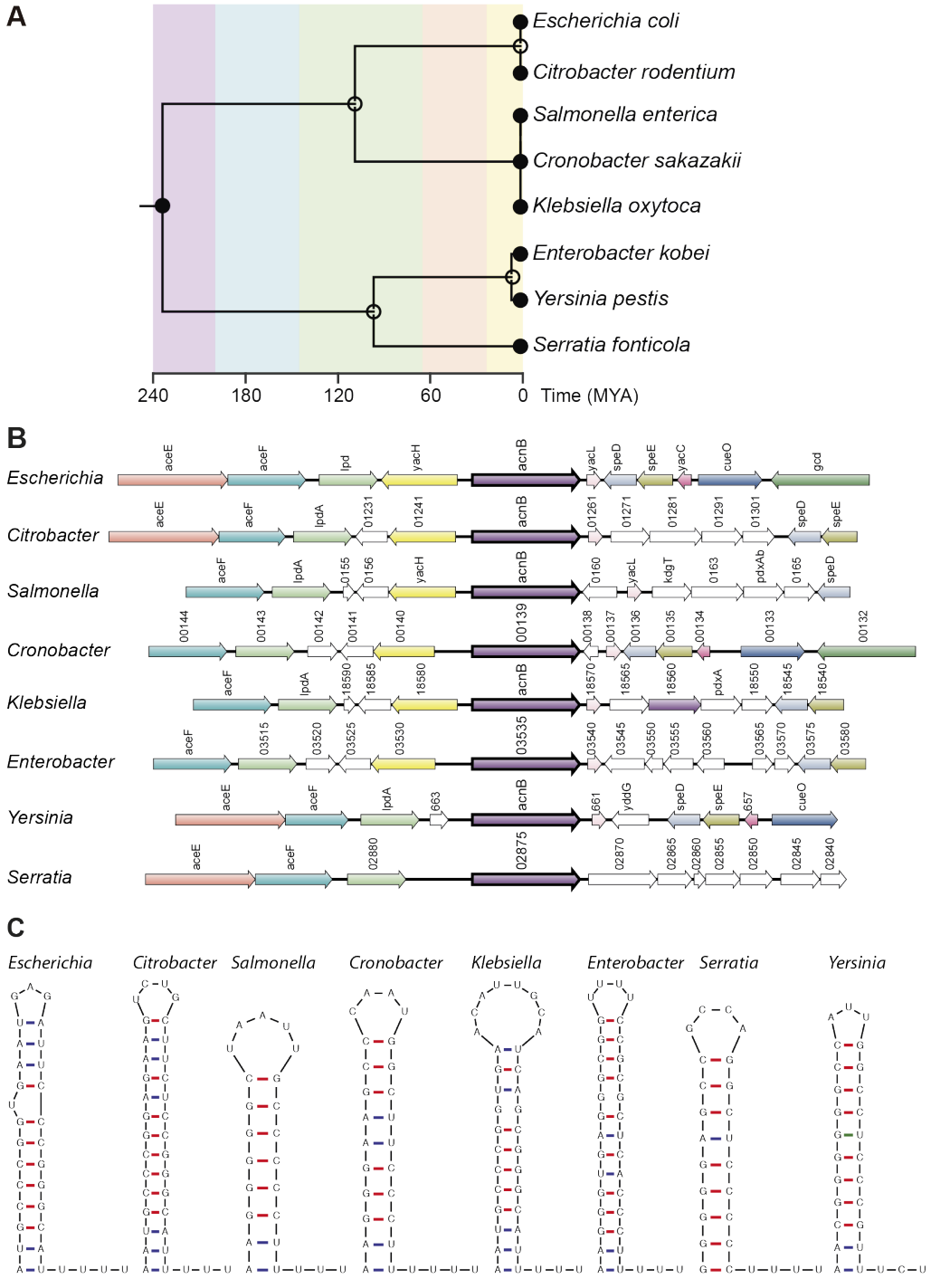
Annex 7. Continued

S. aureus strain	Aligned nucleotides	3'UTR position[a]	Contig IDs matched by blastn	Reason of alignment disruption	Insertion sequence[b]
M1361	1-1332 1325-1385	138 131	KI996660.1 KI996703.1	Query matched two different contigs. Inserted sequence.	→IS1181→
M1362	1-1332 1325-1385	138 131	JDEO01000009.1 JDEO01000024.1	Query matched two different contigs. Inserted sequence.	→IS1181→
M1365	1-1332 1325-1385	138 131	KI996771.1 KI996719.1	Query matched two different contigs. Inserted sequence.	→IS1181→
21272	1-1333 1326-1385	139 132	AHJY01000026.1 AHJY01000029.1	Query matched two different contigs. 7 nt duplication found at the contig ends. Putative IS insertion.	ND
M6K049	1-1333 1326-1385	139 132	BEAN01000005.1 BEAN01000006.1	Query matched two different contigs. Inserted sequence.	→IS1181→
M1532	1-1333 1326-1385	139 132	KI998229.1 KI998207.1	Query matched two different contigs. Inserted sequence.	←IS1181←
103	1-1339 1361-1385	145 167	NAID01000066.1 NAID01000010.1	Query matched two different contigs. Missing 22 nt in comparison to reference.	ND
SAV1149	1-1374 1364-1385	180 170	QYAT01000195.1 QYAT01000051.1	Query matched two different contigs. 10 nt duplication found at the contig ends. Putative IS insertion.	IS?

[a] Pairwise alignment disruption position at the *rpiRc* 3'UTR. Distance from RpiRc stop codon is indicated.

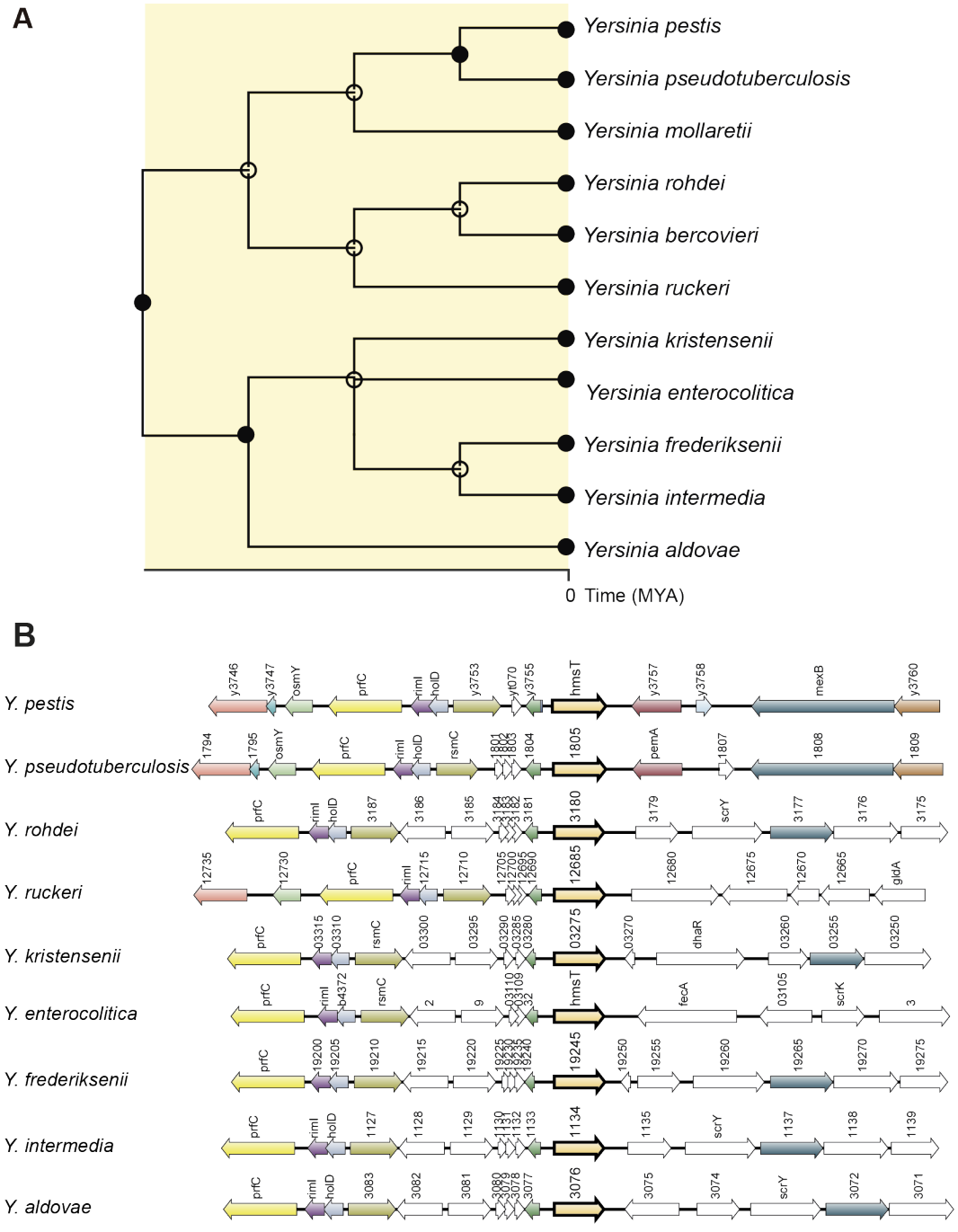
[b] ND, not determined due to missing sequence information. Arrows indicated the transposase gene orientation relative to RpiRc CDS.

Annex 8. Comparison of *acnB* mRNA genomic configurations and predicted TTs among some relevant species of the *Enterobacteriaceae* family



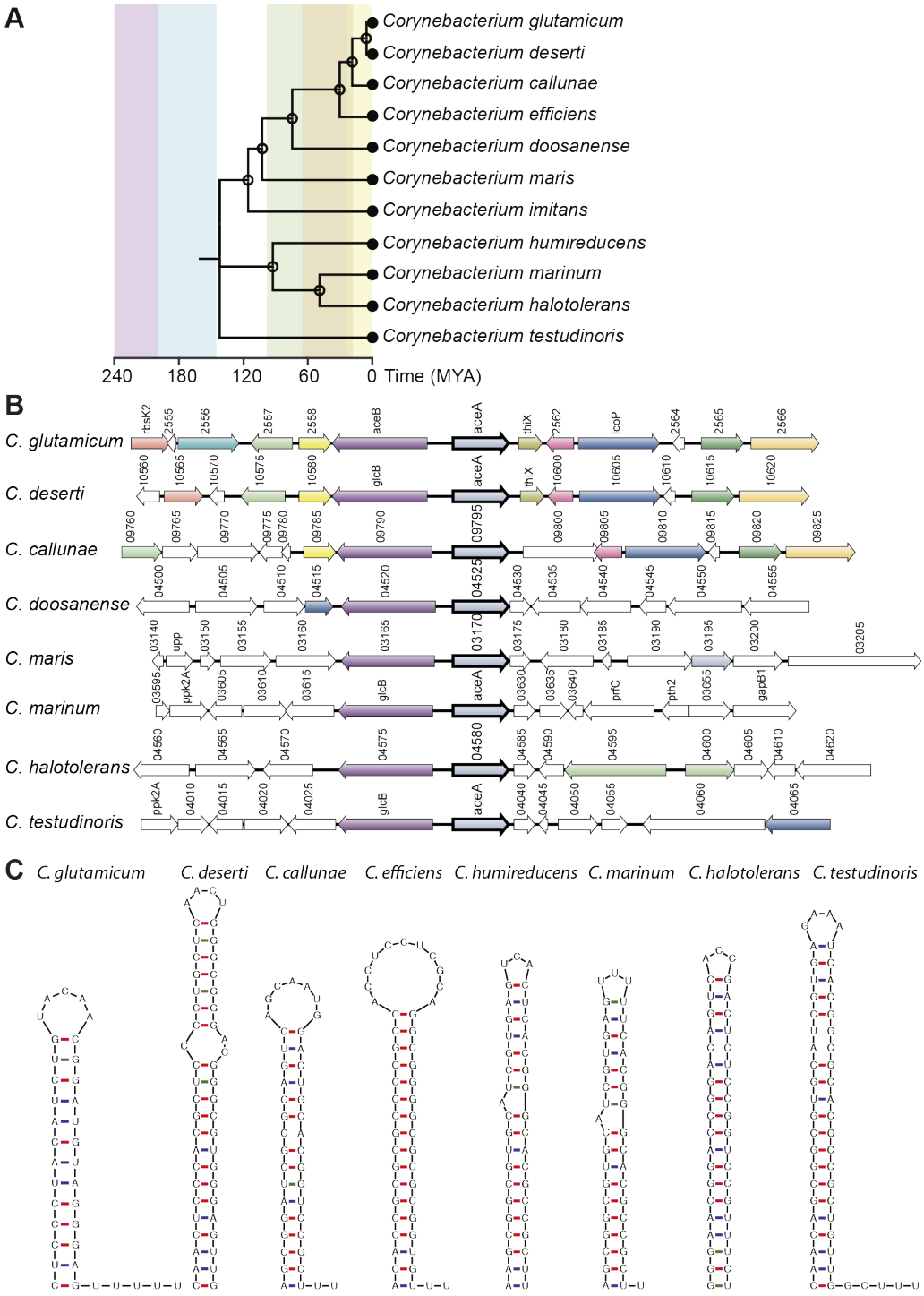
(A) Phylogenetic tree representation showing evolutionary time scales according to TimeTree (http://www.timetree.org) (Hedges *et al.*, 2015). MYA, million years ago. **(B)** Synteny analysis of the chromosomic *acnB* region performed using SyntTax, the Prokaryotic Synteny & Taxonomy Explorer (http://archaea.u-psud.fr/synttax/) (Oberto, 2013). **(C)** Putative transcriptional terminator structures predicted by Mfold in different enterobacterial species (Zuker, 2003).

Annex 9. Comparison of *hmsT* mRNA genomic configurations among some relevant species of *Yersina* genus



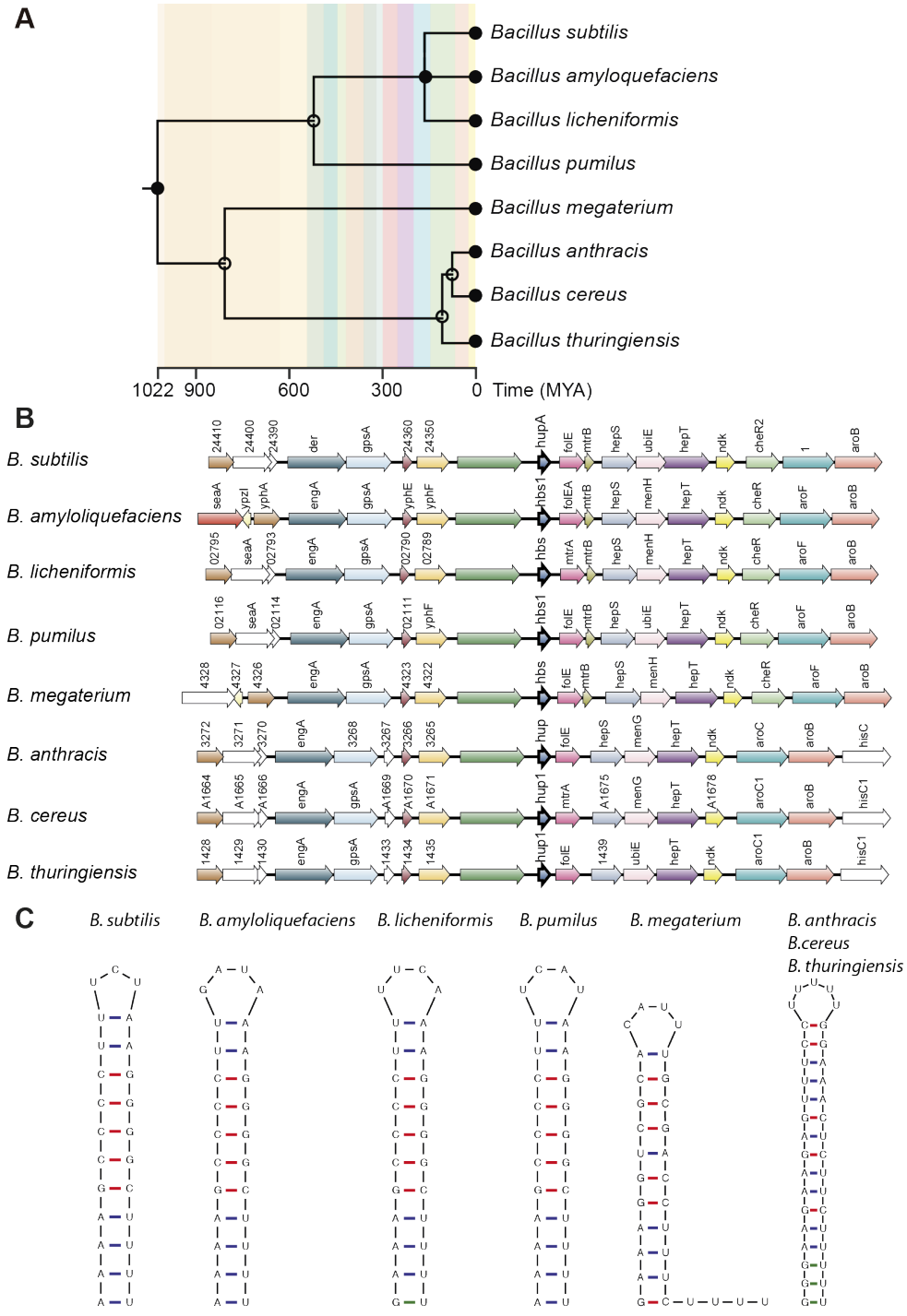
(A) Phylogenetic tree representation showing evolutionary time scales according to TimeTree (http://www.timetree.org) (Hedges *et al.*, 2015). MYA, million years ago. **(B)** Synteny analysis of the chromosomic *hmsT* region performed using SyntTax, the Prokaryotic Synteny & Taxonomy Explorer (http://archaea.u-psud.fr/synttax/) (Oberto, 2013).

Annex 10. Comparison of *aceA* mRNA genomic configurations and predicted TTs among some relevant species of *Corynebacterium* genus

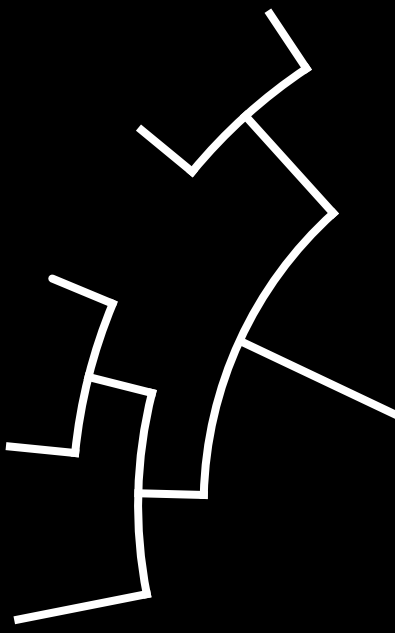


(A) Phylogenetic tree representation showing evolutionary time scales according to TimeTree (http://www.timetree.org) (Hedges *et al.*, 2015). MYA, million years ago. **(B)** Synteny analysis of the chromosomic *aceA* region performed using SyntTax, the Prokaryotic Synteny & Taxonomy Explorer (http://archaea.u-psud.fr/synttax/) (Oberto, 2013). **(C)** Putative transcriptional terminator structures predicted by Mfold in different *Corynebacterium* species (Zuker, 2003).

Annex 11. Comparison of *hbs* mRNA genomic configurations and predicted TTs among some relevant species of *Bacillus* genus



(A) Phylogenetic tree representation showing evolutionary time scales according TimeTree (http://www.timetree.org) (Hedges *et al.*, 2015). MYA, million years ago. **(B)** Synteny analysis of the chromosomic *hbs* region, performed using SyntTax, the Prokaryotic Synteny & Taxonomy Explorer (http://archaea.u-psud.fr/synttax/) (Oberto, 2013). **(C)** Putative transcriptional terminator structures predicted by Mfold in different *Bacillus* species (Zuker, 2003).



IdAB
Instituto de Agrobiotecnología
Agrobioteknologiako Institutua

GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA E INNOVACIÓN

CSIC
Consejo Superior de Investigaciones Científicas

Nafarroako Gobernua
Gobierno de Navarra

erc
European Research Council
Established by the European Commission

European Commission

Horizon 2020
European Union funding
for Research & Innovation

upna
Universidad Pública de Navarra
Nafarroako Unibertsitate Publikoa

upna
Universidad Pública de Navarra
Nafarroako Unibertsitate Publikoa