CLASSIFICATION. Major category: Biological Sciences. Minor category: Microbiology.

The non-contiguous operon: a novel genetic organization to coordinate bacterial gene expression

S. Saenz-Lahoya[a], N. Bitarte[a], B. García[a], S. Burgui[a], M. Vergara-Irigaray[a], J. Valle[a], C. Solano[a], A. Toledo-Arana[b] and I. Lasa[a*]

[a] Laboratory of Microbial Pathogenesis. Navarrabiomed, Complejo Hospitalario de Navarra (CHN)-Universidad Pública de Navarra (UPNA), IDISNA, Pamplona-31008, Spain

[b] Instituto de Agrobiotecnología. IDAB, CSIC-UPNA-Gobierno de Navarra. 31192-Mutilva, Navarra, Spain.

[*] Corresponding author: Iñigo Lasa.

To whom correspondence should be addressed: Iñigo Lasa. Laboratory of Microbial Pathogenesis. Navarrabiomed-Complejo Hospitalario de Navarra (CHN)-Universidad Pública de Navarra (UPNA), IDISNA, Pamplona-31008, Spain

E-mail: ilasa@unavarra.es

Keywords: operon, antisense transcription, overlapping transcription, *Staphylococcus aureus*, RNase III, menaquinone.

ABSTRACT

Bacterial genes are typically grouped into operons defined as clusters of adjacent genes encoding for proteins that fill related roles and are transcribed into a single polycistronic mRNA molecule. This simple organization provides an efficient mechanism to coordinate the expression of neighboring genes and is at the basis of gene regulation in bacteria. Here, we report the existence of a higher level of organization in operon structure that we named non-contiguous operon and consists in an operon containing a gene (s) that is transcribed in the opposite direction to the rest of the operon. This novel transcriptional architecture is exemplified by the genes *menE-menC-MW1733-ytkD-MW1731* involved in menaquinone synthesis in the major human pathogen *Staphylococcus aureus*. We show that *menE-menC-ytkD-MW1731* genes are transcribed as a single transcription unit whereas the *MW1733* gene, located between *menC* and *ytkD*, is transcribed in the opposite direction. This genomic organization generates overlapping transcripts whose expression is mutually regulated by transcriptional interference and RNase III processing at the overlapping region. In the light of our results, the canonical view of operon structure should be revisited by including this novel operon arrangement in which co-transcription and overlapping transcription are combined to coordinate functionally related gene expression.

Significance

In bacteria, functionally related genes are often co-transcribed in a single mRNA molecule under the same upstream promoter, forming a polycistronic operon unit. With this strategy, bacteria guarantee that production of all proteins related to a specific cellular process is simultaneously switched on or off. Here, we report the identification of a novel transcriptional organization consisting in operons that contain a gene(s) that is transcribed in the opposite direction to the rest of the genes of the operon. As a consequence, the resulting mRNA is fully complementary to the operon transcript. This genetic arrangement leads to mutual regulation of the overlapping transcripts expression and thus provides an additional strategy for coordinating the expression of functionally related genes within an operon.

\body

INTRODUCTION

The term operon was first proposed by F. Jacob and J. Monod as a functional genomic DNA unit containing a group of genes that are transcribed together under the control of a single promoter (1). This concept served to explain a revolutionary model of bacterial gene regulation, in which expression of a cluster of genes was negatively controlled by a repressor acting at a single operator that coordinated the process. Over time, it has been shown that operon gene regulation is much more complex than originally expected, with operons that can be positively and/or negatively regulated at different levels other than transcription initiation (2-4). The overall simplicity of operon organization for coordinating gene expression explains why a substantial fraction of functionally related bacterial genes are organized into operons (5, 6).

The list of operons in a specific genome can be predicted with reasonable accuracy based on features such as the distance between each adjacent gene, the likelihood of a pair of genes to be neighbors in a group of reference genomes and the phylogenetic distance (7-9). However, these features are predisposed by the classical operon concept and consequently, the identification of variations in genetic organization inside operon structure through bioinformatic predictions has been very limited. The development of RNA deep sequencing technologies is helping to elucidate ‘operons’ in their full complexities by precisely defining the beginning and the end of mRNA molecules and revealing the changes in the structure of statically predicted operons under different experimental conditions (4, 10-17). An unexpected finding unraveled by the precise determination of transcript boundaries is that very often convergent operons overlap at their 3’ end (tail to tail) and divergent operons overlap at their 5’ end (head to head). In these situations, the mRNA contributes to the expression of operon genes, and at the same time, a region of the mRNA acts as an antisense transcript, affecting the expression of the contiguous operon. The simultaneous sense and antisense functions for transcript boundaries were reported under the term excludon in

Listeria monocytogenes (18). This genomic organization allows the establishment of a regulatory relationship that results in the “exclusive” expression of only one of both coding regions. The mechanisms for excludon mediated regulation is multifaceted and it can include transcription interference, transcription attenuation, degradation of the double-stranded overlapping RNAs or stabilization of the RNAs after cleavage (19, 20). In a previous work, and through a genome wide transcriptome profiling of the pathogen *Staphylococcus aureus* (21), we identified several examples of groups of genes that were apparently transcribed together despite they were separated by gene(s) transcribed in the opposite direction. This transcriptional organization is an extreme example of an excludon, since the mRNA encoded on the opposite strand of DNA to the operon would serve as a canonical mRNA that encodes for a protein while acting as an antisense RNA, base-pairing all along its length with an internal untranslated region of the polycistronic mRNA. Here, we report the existence of this new transcriptional organization in an operon involved in the synthesis of menaquinone in *S. aureus*. Our results demonstrate that the expression of both overlapping transcripts is mutually regulated by transcriptional interference and endoribonuclease mediated digestion. The existence of this genetic arrangement that we named non-contiguous operon, confirms overlapping transcription as a specific mechanism for regulating gene expression within an operon. In addition, it underlines the relevance of reviewing operon structures in bacterial genomes to identify all protein partners whose expression is coordinated in a particular cellular process.

RESULTS

Identification of "non-contiguous operons" in the *S. aureus* genome

We screened genome-wide the transcriptome data obtained from the clinical isolates *S. aureus* 15981 (21) and *S. aureus* MW2 to identify genes cotranscribed together despite being separated by a gene transcribed in the opposite direction. We found six examples that fit the predicted model (Fig. 1, S1 and S2). RNA sequencing data of published results from different laboratories (http://rnamaps.unavarra.es/) confirmed the existence of identical transcriptional organizations in five other genetically unrelated *S. aureus* strains (Fig.1, S1 and S2). The function of most of the proteins encoded by such operons is unknown. CoaD, which is part of the coenzyme A biosynthesis pathway, MenADB and MenEC, required for menaquinone synthesis, and MoaABCED, required for molybdate transport, are among the proteins with known functions.

To explore the significance of this transcriptional organization, we chose the region comprising *menE*, *menC*, *MW1733, ytkD* and *MW1731* genes based on the size of the transcripts and the relevance of menaquinone synthesis during *S. aureus* infections (22) (Fig. 1). *menE-menC* and *ytkD-MW1731* are listed as two independent operons in the prokaryotic operons database (http://csbl.bmb.uga.edu/DOOR/) (23). However, transcriptome data indicated that both operons are transcribed as a single transcriptional unit (Fig. 1). These results agreed with published results obtained by mapping of transcriptional start sites (TSS) by differential RNA-seq that revealed a unique TSS upstream the *menE* gene (24) (Fig. 1). To experimentally confirm the transcriptome results and because the environmental conditions controlling the expression of *menE* remain unknown, we first generated two derivatives of the wild type strain in which the promoter region upstream the *menE* gene was deleted (ΔP*men* strain) or replaced by the constitutive *blaZ* promoter (P*blaZ-men* strain) (Fig. S3). For each of these strains, we generated derivatives in which the chromosomal copy of either *menC* or *MW1731* genes was tagged with the 3xFLAG sequence (Fig. S3) and then examined MenC and MW1731 protein levels by western blotting. Consistent with

the transcriptome results, the ΔP*men* mutation correlated with inhibition of both MenC and MW1731 proteins expression. Note that the ΔP*men* deletion did not completely abolish protein production. On the contrary, the strains containing the constitutive BlaZ promoter produced considerably higher levels of MenC and MW1731 compared with the wild type strain (Fig. 2A). Thus, these findings suggested that expression of *ytkD-MW1731* depends on the promoter region upstream the *menE* gene.

To further validate the cotranscription of *menE-menC-ytkD-MW1731* genes, we performed northern-blot analysis using strand-specific riboprobes corresponding to *menE-menC* (probe A) and *ytkD-MW1731* (probe B) coding regions with total RNA from exponentially growing cells of the wild type strain, and its two isogenic derivatives, ΔP*men* and P*blaZ-men*. Results showed that the mRNA expression levels are very low in the wild type strain, because neither probe was able to detect the mRNA (Fig. 2B). In contrast, both probe A and B hybridizations with P*blaZ-men* RNA revealed an increased accumulation of a fuzzy band of approximately 4 kb that was compatible with cotranscription of *menE-menC* with the downstream genes *ytkD-MW1731* (Fig. 2B). Note that probe B also clearly detects an additional processing band (≈ 1.2kb). Together, these results strongly suggest that *menE-menC* and *ytkD-MW1731* expression depends on the promoter located upstream *menE*, which is consistent with transcriptome data indicating that *menE-menC-ytkD-MW1731* genes comprise a single transcriptional unit.

Transcriptome data also indicated that the *MW1733* gene (258 bp long) is transcribed at high levels with a short 5' UTR of 26 nucleotides and a 3' UTR of 60 nucleotides that overlaps the 3' end of the *menC* coding sequence. To confirm transcriptome data, we generated two additional strains in which 27 nucleotides of the promoter region upstream the *MW1733* gene were deleted (ΔP*MW1733* strain) or replaced by the constitutive *blaZ* promoter (P*blaZ-MW1733* strain) (Fig. S3). For each of these strains, we generated a derivative in which the chromosomal copy of the *MW1733* gene was

tagged with the 3xFLAG sequence (Fig. S3). As expected, analysis of MW1733 protein levels revealed that ΔP*MW1733* mutation correlated with inhibition of MW1733 protein production whereas transcription from the constitutive BlaZ promoter led to higher levels of MW1733 compared with those in the wild type strain (Fig. 2A). Finally, we performed a northern-blot analysis using a strand-specific riboprobe corresponding to the *MW1733* coding region (probe C) with total RNA from exponentially growing cells of the wild type, ΔP*MW1733* and P*blaZ-MW1733* strains. Results showed the presence of a discrete band of approximately 350 nucleotides in the wild type strain (Fig. 2B). Hybridization with ΔP*MW1733* and P*blaZ-MW1733* RNA revealed a decreased and increased accumulation of the RNA product, respectively.

Overall, these results describe an operon organization which is novel in the sense that all the genes of the operon are not contiguous and therefore, we refer to it as “non-contiguous operon”. The first two genes are followed by a gene transcribed from the opposite strand, this gene is then followed by two other genes cotranscribed with the first two. In this spatial transcriptional organization, the non-coding region of the tetracistronic transcript completely overlaps the monocistronic unit.

The expression of the menE-menC-ytkD-MW1731 **operon and the** ***MW1733 gene*** **is** ***reciprocally*** **regulated**

To determine whether transcriptional levels of *menE-menC-ytkD-MW1731* have an effect on the amount of *MW1733* mRNA**,** we compared by northern blot the transcript levels of *MW1733* in the wild type, ΔP*men* and P*blaZ-men* strains using probe C to detect *MW1733* mRNA**.** Results showed that *MW1733* transcript levels slightly increased when transcription of the operon was inhibited and on the other hand, markedly decreased in P*blaZ-men* strain, that is under the presence of an excess of the overlapping tetracistronic transcript (Fig. 3A). To confirm the regulation of *MW1733* expression at a protein level, we constructed derivatives of ΔP*men* and P*blaZ-men*

containing the chromosomal copy of the *MW1733* gene tagged with a 3xFLAG epitope at the C-terminus (Fig. S4). Consistent with northern-blot results, MW1733 protein levels significantly decreased in P*blaZ-men* compared to those in the wild type strain (Fig. 3A).

Next, we investigated the possibility of a reciprocal effect of *MW1733* mRNA levels on the expression of the tetracistronic operon. To do so, we firstly analysed by northern blot, and with the use of probe A, *menE-menC-ytkD-MW1731* mRNA levels in the wild type, ΔP*MW1733* and P*blaZ-MW1733* strains. In agreement with the low level of expression of the tetracistronic mRNA in the wild type strain (Fig. 2B), we could not find a significant difference in *menE-menC-ytkD-MW1731* mRNA levels between strains when probe A was used (Fig. 3B). Thus, we repeated the northern blot assay with the use of probe B, specific for *ytkD-MW1731*. Again, the *ytkD-MW1731* transcript was undetectable in the wild type and ΔP*MW1733* strains. However, when *MW1733* was overexpressed, a specific processing transcript was detected. The size of the discrete band (≈ 1.5kb) is consistent with a transcript including *ytkD-MW1731* that might be obtained upon processing of the *menE-menC-ytkD-MW1731* mRNA (Fig. 3C). Next, we constructed derivatives of ΔP*MW1733* and P*blaZ-MW1733* harbouring a chromosomal copy of either *menC* or *MW1731* tagged with the 3xFLAG epitope in the carboxy-terminal domain (Fig. S4). Notably, constitutive expression of *MW1733* caused a clear reduction in the levels of the MenC protein (Fig. 3B) and a significant accumulation of MW1731 protein levels in P*blaZ-MW1733* compared to the wild type strain (Fig. 3C).

Collectively, these results support the notion that in the non-contiguous operon, transcriptional units generated from opposite strands regulate each other's expression. Thus, in the non-contiguous operon under study, an increase in tetracistronic operon transcription negatively regulates the expression of the interspersed *MW1733* gene. Reciprocally, an increase in *MW1733* mRNA discoordinates expression within the

overlapped operon, by strongly elevating *ytkD-MW1731* mRNA levels while reducing *menE-menC* expression.

Analysis of the mechanisms underlying the regulation of non-contiguous operons expression

RNase III endoribonuclease is responsible for processing overlapping sense/antisense transcripts genome wide in bacteria (21, 25, 26). Thus, we examined the importance of RNase III activity in the reduction of *MW1733* transcript levels when an excess of *menE-menC-ytkD-MW1731* is transcribed. Deletion of RNase III both in the wild type strain and in the strain overproducing the tetracistronic operon (P*blaZ-men*) (Fig. S5), caused a slight increase in the amount of *MW1733* mRNA (Fig. 4A). Consequently, MW1733 protein levels only moderately increased in *rnc* mutants compared to those in the respective RNase III producing strains (Fig. 4A). On the other hand, we studied the involvement of RNase III in *menE-menC-ytkD-MW1731* mRNA processing when *MW1733* is overexpressed. A northern blot, using probe A, with RNA from cells of the wild type, P*blaZ-MW1733*, and their corresponding *rnc* mutants showed no significant differences between strains, given the low detectability of *menE-menC-ytkD-MW1731* mRNA (Fig. 4B). Secondly, we carried out a similar northern blot, but with the use of probe B, specific to detect *ytkD-MW1731* mRNA. Results revealed that processing of the tetracistronic mRNA when an excess of *MW1733* is transcribed still occurred in the absence of RNase III. However, in this case, the processing pattern of the operon changed, leading to a significant decrease in the amount of the discrete 1.5 kb transcript and to the appearance of two additional larger mRNA fragments (Fig. 4C). Accordingly, MW1731 protein levels decreased in *rnc* mutants of the wild type and P*blaZ-MW1733* strains when compared to those in their respective RNase III producing strains (Fig. 4C and Fig. S5). Overall, these results indicated that RNase III explains, only to a certain extent, the *MW1733* mediated cleavage of *menE-menC-ytkD-MW1731*

mRNA, suggesting that additional ribonuclease(s) might also be responsible for this processing.

Besides processing by RNase III, another possible explanation for the reciprocal regulation of overlapping transcripts described above might be transcriptional interference (27), defined as the suppressive influence that the convergent RNA synthesis machinery from one DNA strand causes in *cis* on the transcription of the neighboring gene. Thus, we next sought to determine whether the observed antisense regulation of *menE-menC-ytkD-MW1731* over *MW1733* occurred when the *MW1733* gene was expressed in another location of the chromosome. To do so, we inserted a 3xFLAG tagged *MW1733* gene under its own promoter next to the innocuous *attB* site of the lipase gene in both ΔP*MW1733* and ΔP*MW1733* P*blaZ-men* genetic backgrounds (Fig. S6). Importantly, and contrary to what happens when *MW1733* is located in its natural location, northern blot analysis of *MW1733* transcript levels showed that these were only slightly reduced in the presence of an excess of *menE-menC-ytkD-MW1731* mRNA when the *MW1733* gene was placed and expressed *in trans* (Fig. 4D). Note that there is a marked difference in the size and abundance of *MW1733* mRNA when it is ectopically expressed from the *attB* chromosomal location. Consistent with northern blot results, western blot analysis showed that MW1733 protein levels were unaffected in the ΔP*MW1733* P*blaZ-men MW1733 trans* strain compared with those in the wild type strain (Fig. 4D). These results indicated that *menE-menC-ytkD-MW1731* mediated suppressive influence on *MW1733* expression requires *cis* localization of both transcripts. Lastly, to reinforce these results, we overexpressed a 3xFLAG tagged *MW1733* gene ectopically from a plasmid in the wild type strain harboring a chromosomal copy of either *menC* or *MW1731* tagged with the 3xFLAG epitope (Fig. S6) and analyzed MenC and MW1731 levels by western blot. Overexpression of *MW1733 in trans* did not have any impact on MenC or MW1731 production, showing that *MW1733* effect in discoordinating *menE-menC-YtkD-MW1731* operon expression also requires *cis* localization of both transcripts (Fig. 4E). Overall,

the above results indicate the existence of a transcriptional interference mechanism of gene regulation between the machinery that synthesizes the non-contiguous operon mRNA and the one synthesizing the mRNA of the interspersed gene.

High transcriptional levels of the *MW1733* gene can lead to the appearance of SCVs

The experiments shown above demonstrated that overexpression of *MW1733* mRNA leads to reduced MenC protein levels (Fig. 3B). In *S. aureus*, the inhibition of the synthesis of menaquinone has been associated with a slowed growth phenotype, known as small colony variants (SCVs) (28). SCVs are frequently isolated from clinical samples obtained from patients experiencing chronic infections by *S. aureus*. We observed that the ΔP*men* strain constructed in this work, which still shows some residual production of the MenC protein (Fig. 2A), produces colonies whose size are smaller than the ones corresponding to the wild type strain though they are not as small as the SCVs generated by deletion of *menE-menC* genes (Δ*menEC*) (Fig. S7A). Therefore, we wondered whether constitutive expression of *MW1733* might be followed by the appearance of SCVs phenotypic hallmarks. To test this hypothesis, the promoter of *MW1733* was replaced by the constitutive BlaZ promoter in ΔP*men* strain (Fig. S8). The resulting strain produced colonies significantly smaller than the ΔP*men* strain and exhibited several characteristics associated to *S. aureus* SCVs such as decreased pigmentation and increased resistance to aminoglycosides (tobramycin, streptomycin, gentamycin, and amikacin) than the wild-type strain (29) (Fig. S7B and C). These results suggest that overexpression of *MW1733* suppresses the expression of its convergent *menE-menC* genes, which in turn leads to suppressed menaquinone synthesis and the appearance of a SCV phenotype.

To confirm that appearance of SCVs by *MW1733* overexpression in ΔP*men* strain exclusively happened when *MW1733* and *menE-menC-YtkD-MW1731* mRNAs were

expressed *in cis*, we overexpressed the *MW1733* gene ectopically from a plasmid in ΔP*men* strain and analyzed colony size on TSA plates. The resulting strain, ΔP*men* pCN40::*MW1733* (Fig. S7A and Fig. S8), showed the same phenotype as the ΔP*men* strain. Thus, we conclude that this non-contiguous operon transcriptional organization constitutes an effective mechanism for regulating gene expression and ultimately for controlling cell growth.

DISCUSSION

The novelty introduced by the “non-contiguous operon” concept is that genes within an operon can be interspersed with genes divergently transcribed and that, consequently, they do not necessarily need to be contiguous in the genome. This transcriptional arrangement does not fit within the classical operon paradigm, explaining why it has passed previously unnoticed. It is important to note that in all the examples of non-contiguous operons in the *S. aureus* genome, coding sequences of the operon never overlap the coding region of the interspersed gene. Thus, it appears that the non-contiguous operon transcriptional architecture may be a result of evolutionary pressure to minimize genome size and provide an additional strategy for coupling the expression of functionally related polypeptides. Our results provide evidence of two mechanisms by which the non-contiguous operon arrangement can coordinate gene expression. The first mechanism is related with the generation of double stranded templates between complementary overlapping RNAs that can modify mRNA stability or translation (30, 31). We showed that RNase III digestion of the mRNA duplexes is partially responsible for both the repression of *MW1733* expression and also for the cleavage of the tetracistronic mRNA into two independent transcripts. The resulting two halves might be translated into proteins at a similar or different rate than before the cleavage. Our results indicate that transcriptional induction of the *MW1733* gene leads on one hand, to a reduction in MenE protein levels and on the other, to the stabilization of the *ytkD-MW1731* half and thus, to the accumulation of higher levels of MW1731

protein compared to the wild type strain. Specific RNase III cleavage at intercistronic regions with alternative outcomes for the resulting mRNAs has been previously reported in *E. coli* (32). Opdyke *et al.* showed that binding of the *cis* non-coding RNA *gadY* to the intercistronic region of *gadXW* mRNA resulted in RNase III cleavage and monocistronic transcripts accumulation, probably due to increased stability of single transcripts. Similarly, binding of a *cis*-encoded non-coding RNA to the *cII-O* mRNA of λ phage has been shown to be responsible for an RNase III processing event that is followed by degradation of the upstream *cII* fragment while the downstream *O* mRNA remains stable. Because the sRNA partially overlaps the *cII* coding sequence at its 3′ end, it was concluded that degradation of the *cII* transcript is due to RNase III processing occurring at that region (33). Regarding the mechanisms underlying the stabilization process, it is possible that cleavage might alter the secondary structure of the transcripts so that they are less susceptible to degradation. RNase III is not the only endonuclease involved in *MW1733* dependent processing of the *menE-menC-ytkD-MW1731* operon because discrete RNA fragments from the tetracistronic operon are still detected in the absence of RNase III when *MW1733* is overexpressed. An important direction for future studies will be to identify such additional endoribonuclease(s).

The second mechanism that contributes to coordinating mRNA expression within the non-contiguous operon is transcriptional interference. Because the distance between promoters of the tetracistronic operon and the *MW1733* gene is longer than 200 nucleotides, the most obvious explanation for transcriptional interference is the collision between the RNA synthesis machinery from one DNA strand with the transcription machinery from the other strand (34, 35). A major finding consistent with the existence of transcriptional interference is that tetracistronic operon overexpression did not cause any effect on *MW1733* mRNA levels when this was expressed in *trans* from a separate genomic location. Similarly, the expression of *menC* and *MW1731* was unaffected when *MW1733* was overexpressed in *trans* from a plasmid. Pairing between

complementary transcripts can occur regardless of whether they are expressed in *cis* or *trans*, and therefore, digestion of overlapping transcripts by RNase III and additional endoribonucleases should take place when *MW1733* is produced in *trans.* Thus, we currently do not understand why *MW1733* overexpression in *trans* does not affect *menC* and *MW1731* expression. One possibility is that pairing and processing of the overlapping transcripts is less efficient when both complementary transcripts are produced from separate genomic locations.

What are the benefits of the non-contiguous operon organization compared to regular operons? The exact functions of overlapping transcription are still a matter of debate and several authors defend that overlapping transcription are mainly the product of transcriptional noise, arising at spurious promoters throughout the genome (36). The existence and maintenance of non-contiguous operon transcriptional architecture is a strong evidence that overlapping transcription represents a specific strategy for gene regulation. We can imagine a number of ways the non-contiguous operon may create higher-level organizational features that are adaptive compared to a regular operon. First, it enables a discoordinated expression within the genes of the operon upstream and downstream the overlapping gene, diminishing gene expression noise and ensuring a more precise stoichiometry. Second, it allows endoribonuclease-dependent removal of transcripts that escape the regular transcription repression process. Third, it allows downregulation (exclusion) of the overlapping gene expression by transcript-independent transcriptional interference. Finally, it saves space and decreases the genetic load associated with selecting for a regulatory given motif. All these theoretical benefits require future studies to fully explore the fitness advantages that this transcriptional organization provides to bacteria.

The *menE-menC-MW1733-YtkD-MW1731* genetic arrangement is conserved across the *Staphylococcus* genus, a fact that suggests high functional relevance (Fig. S9). We have found an example of the regulatory possibilities of this transcriptional arrangement in the emergence of a Small-Colony Variant (SCVs) phenotype

associated to menaquinone synthesis deficiency in *S. aureus* (37). Many efforts have been made to identify auxotrophic mutations that result in the appearance of SCVs (28, 38). However, when examining *S. aureus* clinical and tissue-cultured induced SCVs, only around 20% can be assigned to a defined auxotrophy implying that other pathways underlying SCVs formation probably exist (37). Here, we have seen that an increase in the transcription of *MW1733* can account for the induction of SCVs under low polycistronic operon transcription levels, without the need to generate a mutation. The generation of SCVs through this mechanism has the advantage of producing variants able to rapidly switch and revert to the fast-growing wild type phenotype at the earliest opportunity to generate and infection, without the fitness costs associated with the generation of mutations and revertant mutations. In this way, the formation and stability of SCVs would be modulated by environmental conditions affecting transcriptional levels of both the *menE-menC-ytkD-MW1731* operon and the *MW1733* gene and also by factors affecting the binding between overlapping transcripts and the RNase III processing rate. Further work is needed to identify environmental stimuli able to trigger SCVs through this mechanism.

Overall, our results add a further degree of complexity to the initial model of operon gene regulation described by F. Jacob and J. Monod and highlight the functional relevance of overlapping transcription as a mechanism to coordinate the expression levels of bacterial neighbouring genes.

MATERIALS AND METHODS

S. aureus strain 15981 was used as the genetic background for all genetic manipulations. A summary of strains used is provided in Table S1. Mutant strains, 3xFLAG tagged strains, strains harbouring the P*BlaZ* promoter instead of native promoters and strains containing a 3xFLAG tagged *MW1733* gene under its own promoter next to the *attB* site of the lipase gene were generated via allelic replacement using the pMAD vector (39) as described previously (40). For inactivation of *rnc* (RNaseIII encoding gene), the previously described pMAD Δ*rnc::cat86* plasmid (21) was used. Detailed materials and methods are described in SI Materials and Methods.

ACKNOWLEDGMENTS

We thank Jose R. Penadés for critical reading of the paper. This work was supported by the Spanish Ministry of Economy and Competitiveness grants BIO2014-53530-R and BIO2017-83035-R (AEI/FEDER, EU). A.T-A is supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 646869).

REFERENCES

1. Jacob F, Monod J (1961) Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology* 3:318–356.

2. Mattheakis LC, Nomura M (1988) Feedback regulation of the spc operon in *Escherichia coli*: translational coupling and mRNA processing. *J Bacteriol* 170(10):4484–4492.

3. Lim HN, Lee Y, Hussein R (2011) Fundamental relationship between operon organization and gene expression. *Proc Natl Acad Sci U S A* 108(26):10626–10631.

4. Conway T, et al. (2014) Unprecedented high-resolution view of bacterial operon architecture revealed by RNA sequencing. *mBio* 5(4):e01442–14.

5. Okuda S, et al. (2007) Characterization of relationships between transcriptional units and operon structures in *Bacillus subtilis* and *Escherichia coli*. *BMC Genomics* 8(1):48.

6. Rocha EPC (2008) The organization of the bacterial genome. *Annu Rev Genet* 42:211–233.

7. Zheng Y, Szustakowski JD, Fortnow L, Roberts RJ, Kasif S (2002) Computational identification of operons in microbial genomes. *Genome Res* 12(8):1221–1230.

8. Mao X, et al. (2014) DOOR 2.0: presenting operons and their functions through dynamic and integrated views. *Nucleic Acids Res* 42(Database issue):D654–9.

9. Pertea M, Ayanbule K, Smedinghoff M, Salzberg SL (2009) OperonDB: a comprehensive database of predicted operons in microbial genomes. *Nucleic Acids Res* 37(Database issue):D479–82.

10. Toledo-Arana A, et al. (2009) The *Listeria* transcriptional landscape from saprophytism to virulence. *Nature* 459(7249):950–956.

11. Cho B-K, et al. (2009) The transcription unit architecture of the *Escherichia coli* genome. *Nat Biotechnol* 27(11):1043–1049.

12. Güell M, et al. (2009) Transcriptome complexity in a genome-reduced bacterium. *Science* 326(5957):1268–1271.

13. Wurtzel O, et al. (2010) A single-base resolution map of an archaeal transcriptome. *Genome Res* 20(1):133–141.

14. Beaume M, et al. (2010) Cartography of Methicillin-Resistant S. aureus Transcripts: Detection, Orientation and Temporal Expression during Growth Phase and Stress Conditions. *PLoS ONE* 5(5):e10725.

15. Georg J, et al. (2009) Evidence for a major role of antisense RNAs in cyanobacterial gene regulation. *Molecular Systems Biology* 5:305.

16. Dornenburg JE, DeVita AM, Palumbo MJ, Wade JT (2010) Widespread Antisense Transcription in *Escherichia coli*. *mBio* 1(1):e00024–10–e00024–10.

17. Ruiz de Los Mozos I, et al. (2013) Base pairing interaction between 5“- and 3-”UTRs controls icaR mRNA translation in *Staphylococcus aureus. PLoS Genetics* 9(12):e1004001–e1004001.

18. Sesto N, Wurtzel O, Archambaud C, Sorek R, Cossart P (2013) The excludon: a new concept in bacterial antisense RNA-mediated gene regulation. *Nat Rev Micro* 11(2):75–82.

19. Thomason MK, Storz G (2010) Bacterial Antisense RNAs: How Many Are There, and What Are They Doing? *Annu Rev Genet* 44(1):167–188.

20. Georg J, Hess WR (2011) *cis*-antisense RNA, another level of gene regulation in bacteria. *Microbiol Mol Biol Rev* 75(2):286–300.

21. Lasa I, et al. (2011) Genome-wide antisense transcription drives mRNA processing in bacteria. *Proc Natl Acad Sci U S A* 108(50):20172–20177.

22. Dean MA, Olsen RJ, Long SW, Rosato AE, Musser JM (2014) Identification of point mutations in clinical *Staphylococcus aureus* strains that produce small-colony variants auxotrophic for menadione. *Infect Immun* 82(4):1600–1605.

23. Mao F, Dam P, Chou J, Olman V, Xu Y (2009) DOOR: a database for prokaryotic operons. *Nucleic Acids Res* 37(Database issue):D459–63.

24. Koch G, et al. (2014) Evolution of Resistance to a Last-Resort Antibiotic in *Staphylococcus aureus* via Bacterial Competition. *Cell* 158(5):1060–1071.

25. Lioliou E, et al. (2012) Global Regulatory Functions of the *Staphylococcus aureus* Endoribonuclease III in Gene Expression. *PLoS Genetics* 8(6):e1002782.

26. Lybecker M, Zimmermann B, Bilusic I, Tukhtubaeva N, Schroeder R (2014) The double-stranded transcriptome of *Escherichia coli*. *Proc Natl Acad Sci U S A* 111(8):3134–3139.

27. Shearwin KE, Callen BP, Egan JB (2005) Transcriptional interference - a crash course. *Trends Genet* 21(6):339–345.

28. Proctor RA, et al. (2006) Small colony variants: a pathogenic form of bacteria that facilitates persistent and recurrent infections. *Nat Rev Micro* 4(4):295–305.

29. Sendi P, Proctor RA (2009) *Staphylococcus aureus* as an intracellular pathogen: the role of small colony variants. *Trends in Microbiology* 17(2):54–58.

30. Brantl S (2007) Regulatory mechanisms employed by cis-encoded antisense RNAs. *Curr Opin in Microbiol* 10(2):102–109.

31. Lasa I, Toledo-Arana A, Gingeras TR (2012) An effort to make sense of antisense transcription in bacteria. *RNA biology* 9(8):1039–1044.

32. Opdyke JA, Fozo EM, Hemm MR, Storz G (2011) RNase III Participates in GadY-Dependent Cleavage of the gadX-gadW mRNA. *Journal of Molecular Biology* 406(1):29–43.

33. Krinke L, Wulff DL (1990) RNase III-dependent hydrolysis of lambda cII-O gene mRNA mediated by lambda OOP antisense RNA. *Genes & Development* 4(12A):2223–2233.

34. Prescott EM, Proudfoot NJ (2002) Transcriptional collision between convergent genes in budding yeast. *Proc Natl Acad Sci USA* 99(13):8796–8801.

35. Crampton N, Bonass WA, Kirkham J, Rivetti C, Thomson NH (2006) Collision events between RNA polymerases in convergent transcription studied by atomic force microscopy. *Nucleic Acids Res* 34(19):5416–5425.

36. Lloréns-Rico V, et al. (2016) Bacterial antisense RNAs are mainly the product of transcriptional noise. *Sci Adv* 2(3):e1501363.

37. Proctor RA, et al. (2014) *Staphylococcus aureus* Small Colony Variants (SCVs): a road map for the metabolic pathways involved in persistent infections. *Front Cell Inf Microbio* 4(70):99.

38. Johns BE, Purdy KJ, Tucker NP, Maddocks SE (2015) Phenotypic and genotypic characteristics of small colony variants and their role in chronic Infection. *Microbiol Insights* 8:15–23.

39. Arnaud M, Chastanet A, Débarbouillé M (2004) New vector for efficient allelic replacement in naturally nontransformable, low-GC-content, Gram-positive bacteria. *Appl Environ Microbiol* 70(11):6887–6891.

40. Valle J, et al. (2003) SarA and not sigmaB is essential for biofilm development by *Staphylococcus aureus*. *Mol Microbiol* 48(4):1075–1087.

41. King AN, et al. (2018) Guanine limitation results in CodY-dependent and -independent alteration of *Staphylococcus aureus* physiology and gene expression. *J Bacteriol*. doi:10.1128/JB.00136-18.

42. Poupel O, Proux C, Jagla B, Msadek T, Dubrac S (2018) SpdC, a novel virulence factor, controls histidine kinase activity in *Staphylococcus aureus*. *PLoS Pathog* 14(3):e1006917–32.

43. Bui LMG, Hoffmann P, Turnidge JD, Zilm PS, Kidd SP (2015) Prolonged growth of a clinical *Staphylococcus aureus* strain selects for a stable small-colony-variant cell type. *Infect Immun* 83(2):470–481.

44. Yeo W-S, et al. (2018) The FDA-approved anti-cancer drugs, streptozotocin and floxuridine, reduce the virulence of *Staphylococcus aureus*. *Sci Rep* 8(1):2521.

FIGURE LEGENDS

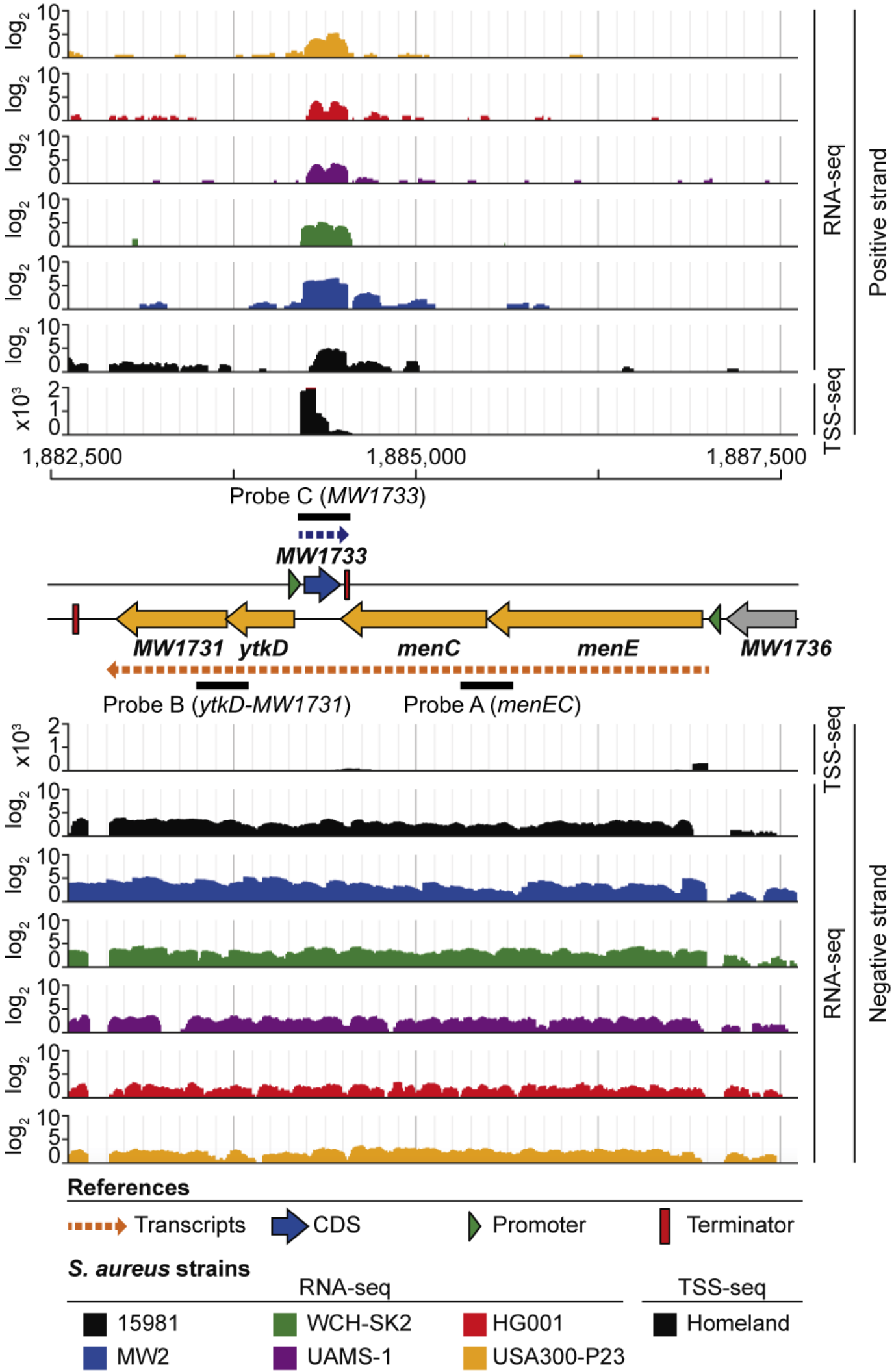
Figure 1. Analysis of the "non-contiguous operon" architecture. JBrowser software images showing RNA-seq or TSS-seq mapped reads distribution in the region comprising *menE-menC-MW1733-ytkD-MW1731* genes of seven unrelated *S. aureus* strains. The scale (log_2 or $x10^3$) indicates the number of mapped reads per nucleotide position. A schematic representation of the structure under study is shown in the middle of the panel. ORFs are represented as orange arrows for the genes that constitute the *menE-menC-ytkD-MW1731* operon and as a blue arrow for the *MW1733* gene. Promoters are shown as green triangles and transcriptional terminators as red rectangles. The transcript generated from the *menE-menC-ytkD-MW1731* operon is represented as a dashed orange arrow whilst the transcript generated from *MW1733* is presented as a dashed blue arrow. The top line denotes the position in base pairs of the *S. aureus* MW2 genome. All genetic information about the start and the end of transcription was obtained from a previous study (21). RNA seq data were obtained from 15981 (21), MW2 (this study), UAMS-1 (41), HG001 (42), WCH-SK2 (43), Homeland (24), USA300-P23 (44).

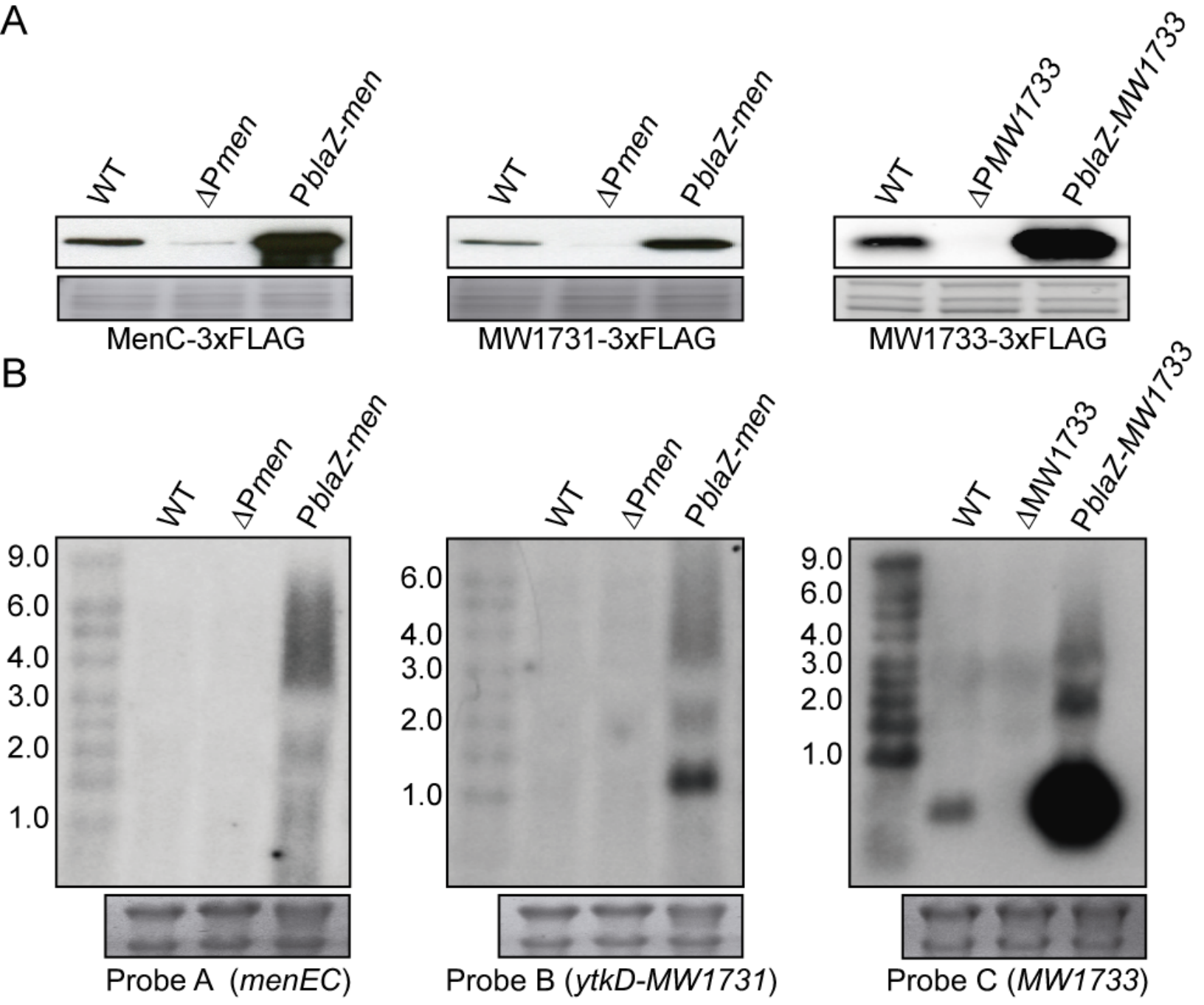
Figure 2. Experimental evidence showing that the region comprising *menE-menC-MW1733-ytkD-MW1731* genes exhibits an architecture characteristic of a non-contiguous operon. (A) Western blots showing MenC, MW1731 or MW1733 protein levels in the wild-type (WT) and the following derivative strains: ΔP*men*, P*blaZ-men*, ΔP*MW1733* and P*blaZ-MW1733*. The 3xFLAG tagged proteins were detected with commercial anti-3xFLAG antibodies. Coomassie stained or stain-free gel portions are shown as a loading control. (B) Northern-blot analysis of RNA harvested from the strains described in A. Blots were probed with specific riboprobes for *menEC*, *ytkD-MW1731* and *MW1733* regions. The positions of RNA standards are indicated. Lower panel shows 16S and 23S ribosome bands stained with ethidium bromide as loading control. The strains used in this figure are depicted in Fig. S3.

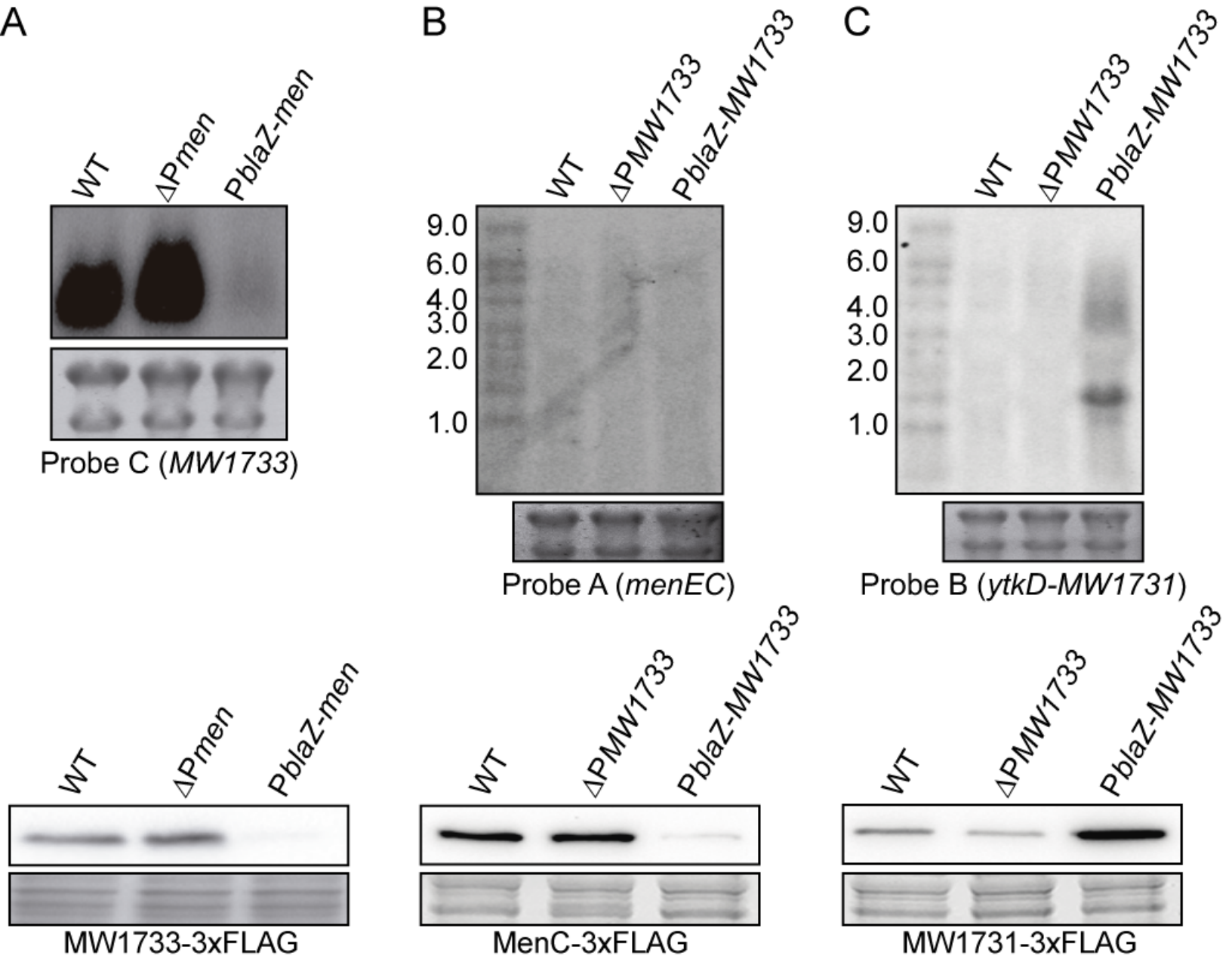
Figure 3. Mutual regulation of overlapping transcripts expression within the non-contiguous operon. (A) Top: Northern blot showing *MW1733* mRNA levels in the WT, ΔP*men* and P*blaZ-men* strains. A specific riboprobe (probe C) for *MW1733* was used; Bottom: Western blot showing MW1733 protein levels in the same strains producing a 3xFLAG tagged MW1733 protein. (B) Top: Northern blot showing *menE-menC* mRNA levels in the WT, ΔP*MW1733* and P*blaZ-MW1733* strains. A specific riboprobe (probe A) for *menEC* was used; Bottom: Western blot showing MenC protein levels in the same strains producing a 3xFLAG tagged MenC protein. (C) Top: Northern blot showing *ytkD-MW1731* mRNA levels in the strains described in B. A specific riboprobe (probe B) for *ytkD-MW1731* was used; Bottom: Western blot showing MW1731 protein levels in the same strains producing a 3xFLAG tagged MW1731 protein. The positions of RNA standards are indicated and16S and 23S ribosome bands stained with ethidium bromide are shown as loading controls. The 3xFLAG tagged proteins were detected with commercial anti-3xFLAG antibodies. Coomassie stained or stain-free gel portions are shown as a loading control. The strains used in this figure are depicted in Fig. S4.

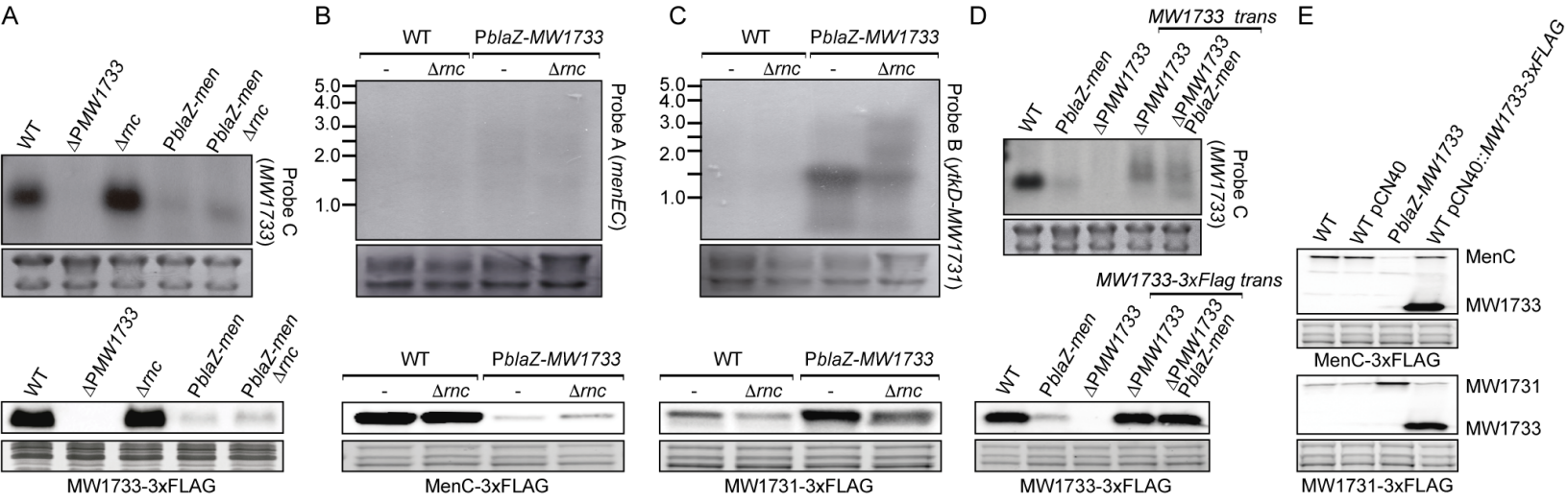
Figure 4. RNase III processing at the overlapping region and transcriptional interference are involved in reciprocal regulation of the overlapping transcripts generated from the non-contiguous operon. (A) Top: Northern blot showing *MW1733* mRNA levels in the WT, ΔP*MW1733*, Δ*rnc*, P*blaZ-men* and P*blaZ-men* Δ*rnc*. A specific riboprobe (probe C) for *MW1733* was used; Bottom: Western blot showing MW1733 protein levels in the same strains producing a 3xFLAG tagged MW1733 protein. (B) Top: Northern blot showing *menE-menC* mRNA levels in the WT, Δ*rnc,* P*blaZ-MW1733* and P*blaZ-MW1733* Δ*rnc*. A specific riboprobe (probe A) for *menE-menC* was used; Bottom: Western blot showing MenC protein levels in the same strains producing a 3xFLAG tagged MenC protein. (C) Top: Northern blot showing

591 *ytkD-MW1731* mRNA levels in the strains described in B. A specific riboprobe (probe
592 B) for *ytkD-MW1731* was used; Bottom: Western blot showing MW1731 protein levels
593 in the same strains producing a 3xFLAG tagged MW1731 protein. (D) Top: Northern
594 blot showing *MW1733* mRNA levels in the WT, P*blaZ-men*, ΔP*MW1733*, ΔP*MW1733*
595 *MW1733 trans* and ΔP*MW1733* P*blaZ-men MW1733 trans*. A specific riboprobe (probe
596 C) for *MW1733* was used; Bottom: Western blot showing MW1733 protein levels in the
597 same strains producing a 3xFLAG tagged MW1733 protein. (E) Western blots showing
598 MenC (top) and MW1733 (bottom) protein levels in the WT, WT pCN40, P*blaZ-*
599 *MW1733* and WT pCN40::*MW1733*-3xFLAG. Strains contained a chromosomal copy of
600 either *menC* or *MW1731* tagged with the 3xFLAG epitope. The positions of RNA
601 standards are indicated and16S and 23S ribosome bands stained with ethidium
602 bromide are shown as loading controls. The 3xFLAG tagged proteins were detected
603 with commercial anti-3xFLAG antibodies. Coomassie stained or stain-free gel portions
604 are shown as a loading control. The strains used in this figure are depicted in Fig. S5,
605 in the case of A, B, and C sections, and in Fig. S6 in the case of D and E sections.
606


log2
x10^3
RNA-seq
TSS-seq
Positive strand
1,882,500
1,885,000
1,887,500
Probe C (MW1733)
MW1733
MW1731
ytkD
menC
menE
MW1736
Probe B (ytkD-MW1731)
Probe A (menEC)
Negative strand
References
Transcripts
CDS
Promoter
Terminator
S. aureus strains
RNA-seq
TSS-seq
15981
WCH-SK2
HG001
Homeland
MW2
UAMS-1
USA300-P23


A
WT
ΔPmen
PblaZ-men
MenC-3xFLAG
WT
ΔPmen
PblaZ-men
MW1731-3xFLAG
WT
ΔPMW1733
PblaZ-MW1733
MW1733-3xFLAG
B
WT
ΔPmen
PblaZ-men
9.0
6.0
4.0
3.0
2.0
1.0
Probe A (menEC)
WT
ΔPmen
PblaZ-men
6.0
4.0
3.0
2.0
1.0
Probe B (ytkD-MW1731)
WT
ΔMW1733
PblaZ-MW1733
9.0
6.0
4.0
3.0
2.0
1.0
Probe C (MW1733)


A
WT
ΔPmen
PblaZ-men
Probe C (MW1733)
B
WT
ΔPMW1733
PblaZ-MW1733
9.0
6.0
4.0
3.0
2.0
1.0
Probe A (menEC)
C
WT
ΔPMW1733
PblaZ-MW1733
9.0
6.0
4.0
3.0
2.0
1.0
Probe B (ytkD-MW1731)
WT
ΔPmen
PblaZ-men
MW1733-3xFLAG
WT
ΔPMW1733
PblaZ-MW1733
MenC-3xFLAG
WT
ΔPMW1733
PblaZ-MW1733
MW1731-3xFLAG


A
WT
ΔPMW1733
Δrnc
PblaZ-men
PblaZ-men Δrnc
Probe C (MW1733)
MW1733-3xFLAG
B
WT
PblaZ-MW1733
- Δrnc - Δrnc
5.0
4.0
3.0
2.0
1.0
Probe A (menEC)
MenC-3xFLAG
C
Probe B (ytkD-MW1731)
MW1731-3xFLAG
D
MW1733 trans
MW1733-3xFlag trans
MW1733-3xFLAG
E
WT pCN40
PblaZ-MW1733
WT pCN40::MW1733-3xFLAG
MenC
MW1733
MW1731
MenC-3xFLAG
MW1731-3xFLAG