microLife, 2024, **5**, uqae007
DOI: 10.1093/femsml/uqae007
Advance access publication date: 30 March 2024
Research Article

Noncontiguous operon atlas for the *Staphylococcus aureus* genome

Pablo Iturbe[1], Alvaro San Martín[1], Hiroshi Hamamoto[2], Marina Marcet-Houben[3,4], Toni Galbaldón[3,4,5,6], Cristina Solano[1], Iñigo Lasa[1,*]

[1]Laboratory of Microbial Pathogenesis, Navarrabiomed-Universidad Pública de Navarra (UPNA)-Hospital Universitario de Navarra (HUN), IdiSNA, Irunlarrea 3, Pamplona, 31008 Navarra, Spain
[2]Faculty of Medicine, Department of Infectious diseases, Yamagata University, 2-2-2 Lida-Nishi, 990-9585 Yamagata, Japan
[3]Barcelona Supercomputing Centre (BSC-CNS). Plaça Eusebi Güell 1-3, 08034 Barcelona, Spain
[4]Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology, Baldiri Reixac, 10, 08028 Barcelona, Spain
[5]Catalan Institution for Research and Advanced Studies (ICREA), 08010 Barcelona, Spain
[6]CIBER de Enfermedades Infecciosas, Instituto de Salud Carlos III, 28029 Madrid, Spain
*Corresponding author. Laboratory of Microbial Pathogenesis, Navarrabiomed-Universidad Pública de Navarra (UPNA)-Hospital Universitario de Navarra (HUN), IdiSNA, Irunlarrea 3, Pamplona, 31008 Navarra, Spain. E-mail: ilasa@unavarra.es
Editor: [Carmen Buchrieser]

Abstract

Bacteria synchronize the expression of genes with related functions by organizing genes into operons so that they are cotranscribed together in a single polycistronic messenger RNA. However, some cellular processes may benefit if the simultaneous production of the operon proteins coincides with the inhibition of the expression of an antagonist gene. To coordinate such situations, bacteria have evolved noncontiguous operons (NcOs), a subtype of operons that contain one or more genes that are transcribed in the opposite direction to the other operon genes. This structure results in overlapping transcripts whose expression is mutually repressed. The presence of NcOs cannot be predicted computationally and their identification requires a detailed knowledge of the bacterial transcriptome. In this study, we used direct RNA sequencing methodology to determine the NcOs map in the *Staphylococcus aureus* genome. We detected the presence of 18 NcOs in the genome of *S. aureus* and four in the genome of the lysogenic prophage 80α. The identified NcOs comprise genes involved in energy metabolism, metal acquisition and transport, toxin–antitoxin systems, and control of the phage life cycle. Using the menaquinone operon as a proof of concept, we show that disarrangement of the NcO architecture results in a reduction of bacterial fitness due to an increase in menaquinone levels and a decrease in the rate of oxygen consumption. Our study demonstrates the significance of NcO structures in bacterial physiology and emphasizes the importance of combining operon maps with transcriptomic data to uncover previously unnoticed functional relationships between neighbouring genes.

Keywords: *Staphylococcus aureus*; noncontiguous operon; overlapping transcription; antisense transcription; excludon; menaquinone synthesis

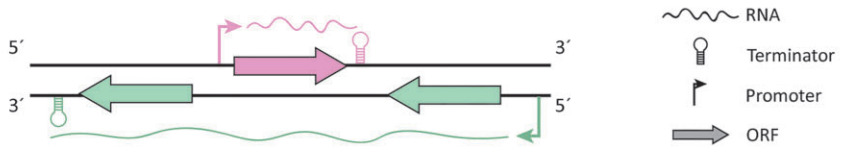
Introduction

A characteristic feature of bacterial genomes is that the genes involved in a biosynthetic pathway are grouped into functional units, called operons (Jacob and Monod 1961). Genes within an operon are cotranscribed into a single RNA molecule under the control of the same promoter. This clustering of genes not only facilitates their simultaneous transcription, but also their transfer between species, reduces the time needed for proteins to reach their intended destinations after synthesis, and facilitates the synchronization of different biological processes (Ballouz et al. 2010, Gilchrist et al. 2021). Adaptation to complex environmental conditions, such as temperature, pH, nutrient availability, or defence against the immune system, often requires the integration of individual regulatory circuits into more complex networks (Gottesman 1984). Normally, the regulatory link between different biological processes is provided by transcription factors that activate or repress the expression of different sets of genes (Feklístov et al. 2014). Recently, a new way of coordinating the expression of neighbouring genes by overlapping the 5' or 3' untranslated region (UTR) of mRNAs has been proposed (Lasa et al. 2012, Wurtzel et al. 2012, Sesto et al. 2013, Toledo-Arana and Lasa 2020). In this case, the collision between the transcriptional machinery of the two genes (transcriptional interference), together with the activity of a double-stranded RNase (RNaseIII) that digests the overlapping transcripts, allows the coordinated expression of adjacent genes. The sets of neighbouring genes that overlap their 5' and 3' UTRs have been termed excludons, because the expression of the more actively expressed gene can suppress the expression of the less actively transcribed gene in the pair. The existence of excludons explains, at least in part, the observation that genes with apparently unrelated functions in bacterial genomes maintain a much higher degree of proximity than would be expected if they were randomly arranged (Rogozin et al. 2002, Mihelčić et al. 2019, Marcet-Houben et al. 2023).

An extreme example of excludon organization is the noncontiguous operon (NcO) (Fig. 1A). NcOs consist of operons that contain one or more genes that are transcribed in the opposite direction to the rest of the operon genes, so that their mRNA over-

Received 11 November 2023; **revised** 20 March 2024; **accepted** 8 April 2024

(A)



(B)

Noncontiguous Operons	Nº of OR	Predicted Function (Based on TIGRFAMS, TheSEED and PFAM)
00208 00209 00210	1	Glucose transport
00470 glmU 00472 00473 00474	891	Magnesium binding
00692 00693 mgrA 00695 00696 00697	33	Gene regulation
00825 00826 00827	6	Toxin-antitoxin system
00979 menA 00981 menF menD menH menB	3	Menaquinone biosynthesis
01071 01072 01073	5	Uncharacterized
01072 01073 01074 cooD	19	Biosynthesis of cofactors, prosthetic groups, and carriers
recG fapR pslX fabD fabB 01200 acpP	1479	Fatty acid and phospholipid metabolism
rpsN 01330 01331 01332	46	Ribosomal structure
01351 01352 01353 01354	34	DNA replication, recombination, and repair
01912 ytkD 01914 menC menE	35	Menaquinone biosynthesis
02106 02107 02108 02109	1	Peptidoglycan biosynthesis
02376 02377 02378 02379	221	Purines, pyrimidines, nucleosides, and nucleotides metabolism
fmhB 02528 02529	10	Peptidoglycan production
02532 02533 02534 02535	6	Uncharacterized
moaA mobA moaD moaE mobB moeA moaC moaB moeB	47	Molybdenum metabolism
strA 02835 02836	3	Membrane protein anchoring
02991 02992 02993 02994 02995 msrA 02997	2	FMN binding
int xis ORF3 ORF4 ORF5 cI	2223	Phage lysogenic module
cro ORF8 ORF9 ORF10 ORF11 ORF12 ORF13	572	Phage lytic module
ORF17 ORF18 ORF19 ORF20 ORF21	322	Phage lytic module
ORF69 ORF70 ORF71 ORF72 ORF73	4	Phage lytic module

Figure 1. (A) NcO structure. Illustration depicting the genetic arrangement of an NcO. (B) List of NcOs of *S. aureus* NCTC8325 and phage 80α obtained by Oxford Nanopore direct RNA sequencing. The selected NcOs fulfil two conditions: there is at least one overlapping read (OR) between the two cotargeted genes and the average read coverage signal at three consecutive nucleotides in the intergenic region does not fall below 0.1 reads per nucleotide. Predictions regarding the function of NcOs were made using the TIGRFAMS, TheSEED and PFAM databases. Gene annotations are based on the *S. aureus* NCTC8325 strain. NcOs were named after the first gene that is cotranscribed with other genes of the operon.

laps with the mRNA of the operon along its entire length (Sáenz-Lahoya et al. 2019). Therefore, this operon structure combines co-transcription and overlapping transcription to coordinate gene expression. Importantly, NcOs identification may allow to uncover functional relationships between neighbouring genes that would otherwise go unnoticed. After the first description of the NcO architecture in *Staphylococcus aureus* (Sáenz-Lahoya et al. 2019), gene arrangements consistent with an NcO structure have been detected in *Bacillus subtilis* (Johnson et al. 2020), the phytopathogenic bacterium *Dickeya dadantii* (Forquet et al. 2022), a temperate phage of the Siphoviridae family from *S. aureus* (Rohmer et al. 2022) and large DNA viruses such as varicella, Epstein–barr, and Baculovirus (Prazsák et al. 2018, Fülöp et al. 2022, Torma et al. 2022). Altogether, these examples support the idea that NcOs are not exclusive to bacteria and are also present in the genomes of phages and viruses. However, our understanding remains limited as regards the number of NcOs within a bacterial genome, the specific biological processes that benefit from this transcriptional arrangement,

and the potential repercussions of disrupting this transcriptional structure on bacterial fitness.

The set of operons present in a bacterial genome can be computationally predicted with reasonable accuracy based on criteria such as the proximity of neighbouring genes with short intergenic distances across diverse bacterial species and functional relationship between the protein products (Mao et al. 2009, Taboada et al. 2012, 2014, 2018, Krishnakumar and Ruffing 2022). As these criteria are fully adapted to the classical operon concept, any variation in gene organization that does not meet these criteria is not considered in current operon identification tools. In the specific case of NcOs, the cotranscribed genes are separated by antisense gene(s) and therefore, do not fulfil the neighbourhood criterion. Identification of NcOs necessarily requires information from the bacterial transcriptome in which the existence of an mRNA encoded by the antisense gene(s) and fully complementary to an internal UTR of the operon mRNA is detected. The use of Nanopore direct RNA sequencing technology enables the sequencing of intact RNA molecules, providing a comprehensive picture of individual RNA strands as they exist in cells (Byrne et al. 2017, Garalde et al. 2018, Jain et al. 2022). Using this approach, here, we have determined the long RNA transcriptome of *S. aureus* and exploited this information to define the NcO map in this bacterium. Our findings revealed the presence of at least 18 NcOs within the *S. aureus* NCTC8325 genome, and four more in the model lysogenic phage 80α. Taking the *menEC-01914-ytkD-01912* NcO as a test case, we demostrate that disruption of the NcO architecture has a significant impact on the bacterium's fitness. Overall, this comprehensive analysis of NcOs opens new avenues for uncovering novel functional and regulatory connections between neighbouring genes.

Materials and methods

Oligonucleotides, plasmids, bacterial strains, and culture conditions

The oligonucleotides, bacterial strains, and plasmids used in this study are listed in Tables S1, S2, and S3 (Supporting Information), respectively. Oligonucleotides were synthesized by Stabvida (Caparica, Portugal). *Staphylococcus aureus* strains were grown in trypticase soy broth (TSB) (Pronadisa). *Escherichia coli* IM01B was grown in LB broth (Pronadisa). When required, growth media were supplemented with antibiotics at the following concentrations: erythromycin 1.5 μg ml^{-1} and ampicillin 100 μg ml^{-1}. Plasmids were purified using the NucleoSpin Plasmid Miniprep Kit (Macherey-Nagel) according to the manufacturer's protocol. FastDigest restriction enzymes and the Rapid DNA Ligation Kit (Thermo Scientific) were used according to the manufacturer's instructions. Plasmids were transformed into *E. coli* IM01B strain and *S. aureus* by electroporation using previously described protocols (Valle et al. 2003). Staphylococcal electrocompetent cells were generated as previously described (Schenk and Laddaga 1992). RNA for transcriptome analysis was purified and sequenced from *S. aureus* RN10359 (a derivative of NCTC8325, cured of three prophages and lysogenized by 80α prophage) (Ubeda et al. 2009) and *S. aureus* MW2 (Baba et al. 2002).

Plasmid construction

Strains carrying the *01912* gene tagged with the 3xFlag epitope were constructed using plasmid pMAD::01912_3xFlag (Sáenz-Lahoya et al. 2019). Strains 15981 *men* operon_*gfp* 01914_3xFlag and *S. aureus* 15981 *men* operon_*gfp* 01914 invert_3xFlag containing the *gfp* gene at the end of the *men* operon and also the *01914* gene tagged with a 3xFlag were constructed starting from a strain, which contains a 3xFlag tagged *01914* gene (*S. aureus* 15981 *01914*_3xFlag) (Sáenz-Lahoya et al. 2019) and the *S. aureus* 15981 *01914* invert_3xFlag strain constructed in the present study. The pMAD::*men* operon_*gfp* plasmid used to generate these strains was constructed as follows: the *gfp* gene was amplified using primers GFPmut2_Fw_Overlap_01912 and GFPmut2_Rv_Overlap_01912 from pCN52 plasmid (Table S1, Supporting Information). The PCR product was purified using the Silica Bead DNA Gel Extraction Kit (Thermo Scientific) and fused by overlap extension PCR with the corresponding AB and CD fragments, which were amplified with primers pMAD_01912_6xHIS_CD_Rv pMAD_01912_6xHIS_CD_EcoRI_Fw and, 01912_EcoRI_Fw and 01912_OverlapGFP_Rv, respectively (Table S1, Supporting Information). The PCR product was purified using the Silica Bead DNA Gel Extraction Kit and cloned into pJET1.2/blunt (Thermo Scientific). Then, pJET containing the overlap PCR product was digested with EcoRI and SalI and the fragment was cloned in the EcoRI/SalI pMAD sites.

To construct plasmids pCN51::*men* operon_*gfp* 01914_3xFlag and pCN51::*men* operon_*gfp* *01914* invert_3XFlag, the whole *men* operon from the strains, which contain the *gfp* and the *01914* gene tagged with the 3xFlag (15981 *menE* operon_*gfp* 01914_3xFlag and 15981 *menE* operon_*gfp* *01914* invert_3xFlag) was amplified with primers Men_Operon_Rv_AscI and Men_Operon_Fw_SmaI. The PCR product was purified using the Silica Bead DNA Gel Extraction Kit and cloned into pJET1.2/blunt (Thermo Scientific). Then, the resulting plasmid was digested with SmaI and AscI and the *men* operon was cloned in pCN51 (Charpentier et al. 2004) digested with SmaI/AscI.

The pMAD::*01914 invert* plasmid used to generate the *S. aureus* 15981 *01914* invert strain was constructed as follows: the *01914* gene was amplified using primers 01914_Fw_RBS_Overlap and 01914_Rv_Overlap from *S. aureus* 15981 strain (Table S1, Supporting Information). The PCR product was purified using the Silica Bead DNA Gel Extraction Kit (Thermo Scientific) and fused by overlap extension PCR with the corresponding AB and CD fragments, which were amplified with primers pMAD_01914 invert_AB_Fw_Overlap and pMAD_01914 invert_AB_Rv_SalI, pMAD_01914 invert_CD_Fw_EcoRI and pMAD_01914 invert_CD_Rv_Overlap, respectively (see Table S1, Supporting Information). The PCR product was purified using the Silica Bead DNA Gel Extraction Kit and cloned into pJET1.2/blunt (Thermo Scientific). Then, pJET containing the overlap PCR product was digested with EcoRI and SalI and the fragment was cloned in the EcoRI/SalI pMAD (Arnaud et al. 2004) sites.

The pMAD::*01914* invert_3xFlag plasmid used to generate the *S. aureus* 15981 *01914* invert_3xFlag strain was constructed as stated for the the pMAD::*01914* invert plasmid, with the following modification. The *01914* gene was amplified using primers 01914_Fw_RBS_Overlap and 01914_3xFlag_Rv_Overlap from *S. aureus* 15981 *01914*::3xFlag (Sáenz-Lahoya et al. 2019) (Tables S1 and S2, Supporting Information).

The pCN51 plasmids containing *01914*, *01912*, and *ytkD* were constructed by amplifying each gene with primers 01914_Rv_AscI/01914_Fw_SmaI, ytkD_Fw_SmaI/ytkD_Rv_AscI, and 01912_Fw_SmaI/01912_Rv_AscI, respectively. The PCR product was purified using the Silica Bead DNA Gel Extraction Kit and cloned into pJET1.2/blunt (Thermo Scientific). Then, pJET plasmids were digested with SmaI and AscI and the fragments were cloned in pCN51 digested with SmaI/AscI.

RNA extraction and RT-PCR

Bacteria were grown overnight in 5 ml of TSB at 37ºC under shaking conditions (200 rpm). Cultures were diluted 1:100 in 20 ml of media and incubated at 37°C under shaking conditions (200 rpm) in 100 ml flasks until they reached an OD_{600} of 0.8. At this point, one of the samples of *S. aureus* strain RN10359, was incubated in the presence of 2 µg ml^{-1} Mitomycin C at 30°C for 1 h. Bacterial cultures were centrifuged, and the pellets were frozen in liquid nitrogen and stored at −80°C until needed. Total RNA from bacterial pellets was extracted by using TRIzol reagent method as described (Toledo-Arana et al. 2009). Briefly, bacterial pellets were resuspended into 400 µl of solution A (glucose 10%, Tris 12.5 mM, pH 7.6, EDTA 10 mM) and mixed to 60 µl of 0.5 M EDTA. Resuspended cells were transferred into Lysing Matrix B tubes (MP Biomedicals) containing 500 µl of acid phenol (Ambion). Bacteria were mechanically lysed by using a Fastprep apparatus (MP Biomedicals) at speed 6.0 during 45 s at 4°C twice. After lysis, tubes were centrifuged for 10 min at 14 000 rpm at 4°C. The aqueous phase was transferred to 2-ml tubes containing 1 ml of TRIzol, mixed, and incubated for 5 min at room temperature. A volume of 100 µl of chloroform were added, mixed, and incubated for 3 min at room temperature. Tubes were centrifuged for 10 min at 14 000 rpm at 4°C. The aqueous phase was transferred into a tube containing 200 µl of chloroform, mixed, and incubated for 5 min at room temperature. Tubes were centrifuged for 5 min at 14 000 rpm at 4°C. A volume of 500 µl of isopropanol was added to the aqueous phase and the mixture was incubated at room temperature for 15 min to precipitate the RNA. Tubes were centrifuged for 15 min at 14 000 rpm at 4°C. RNA pellets were resuspended in DEPC-treated water. RNA was then treated with 5 µl TURBO DNase I in the presence of 1 µl of SUPERase-In RNase Inhibitor (Invitrogen) at 37ºC for 30 min. The DNA-free RNA solution was transferred into a tube with 300 µl of acid phenol, mixed, and incubated for 3 min at room temperature. Tubes were centrifuged for 5 min at 14 000 rpm at 4°C. The aqueous phase was transferred into a tube containing 300 µl of chloroform, mixed, and centrifuged for 5 min at 14 000 rpm at 4°C. RNA contained in the aqueous phase was precipitated by addition of 200 µl of ammonium acetate 5 M and 1 ml of ethanol mixed and incubated for 2 h at −20°C. Tubes were centrifuged for 30 min at 14 000 rpm at 4°C. RNA pellets were resuspended in DEPC-treated water. RNA concentration and RNA quality were determined by using Agilent RNA Nano LabChips (Agilent Technologies) (Lasa et al. 2011). RNAs were stored at −80°C until needed.

The full-length transcription of *men* operons was validated by RT-PCR. RNA (500 µg) was reverse transcribed using M-MLV (Promega) and 10 pmol of primer 3' (RT_PCR_01912_1 and RT_menF_1) (Table S1, Supporting Information). Samples without reverse transcriptase were used as negative controls. Subsequently, 1 µl of cDNA was subjected to a 30-cycle PCR using DreamTaq DNA Polymerase (ThermoFisher) and primers RT_PCR_menE_Fw and RT_PCR_01912 _Rv (Table S1, Supporting Information) for the *menE* NcO and primers RT_SAOUHSC_00979_Fw and RT_menF_Rv (Table S1, Supporting Information) for the *00979* NcO. The primers used to confirm the existence of the rest of NcOs are listed in Table S1 (Supporting Information).

Oxford nanopore direct RNA sequencing

Total RNA was purified from bacterial cells as described above (Lasa et al. 2011). Ribosomal RNAs were removed from 10 µg of total RNA with the RiboZero Bacteria kit (Illumina, CA). Enriched mRNA was treated with *E. coli* Poly(A) Polymerase (NEB) in the presence of 10 mM ATP for 30 min at 37ºC to add a polyA tail. PolyAdenylated mRNAs were cleaned using the RNA Clean & Concentrator kit (Zymo, CA). Purified mRNAs were then converted into a nanopore-compatible RNA library using the Direct RNA Sequencing Kit from Oxford Nanopore (SQC-RNA002) (Pust et al. 2021). The library was sequenced on an R9.4 flowcell in a GridION instrument (Oxford Nanopore) for 24 h and base called in real-time in the instrument with the software Guppy. Data obtained by Oxford Nanopore are available at NCBI under Bioproject PRJNA922758 (*S. aureus* RN10359) and Bioproject PRJNA860339 (*S. aureus* MW2). The resulting reads were mapped using the Geneious Prime tool (Biomatters) using the *S. aureus* NCTC8325 genome as a reference (accession number: NC_007795.1) (Baba et al. 2008). The transcriptome mapping information is available on our website (https://staph.unavarra.es:10443).

Prediction of NcOs

Prediction of NcOs was based on the alignment of sequencing data from *S. aureus* strain RN10359 and MW2 using Oxford Nanopore technology as follows. Oxford Nanopore direct RNA sequencing data were aligned to reference genomes (*S. aureus* NCTC8325, accession number NC_007795 and phage 80α, accession number NC_009526) using minimap2 (Li 2018) with a minimum alignment score of 0.99. This alignment method used 14-base-long substrings from the input sequences to find matching regions in the reference genome. Samtools view (Danecek et al. 2021) was employed to extract the coverage information for each position in the genome, both on the positive and negative strands, regardless of the depth of coverage. Counting of reads aligned to coding sequences or transfer RNA was performed using the GFF or GTF file containing genomic annotations. Only long reads that uniquely mapped to a single feature were counted. Reads with less than 80% identity were excluded using the Rsamtools package (https://bioconductor.org/packages/Rsamtools.). The median start and end sites of all overlapping reads, as well as the median read coverage, were calculated using the GenomicRanges package (Lawrence et al. 2013).

To annotate NcOs, overlapping reads were first combined into potential operons. The gene IDs of these potential operons were annotated. Then, the following procedure was used for each gene within a potential operon: (i) annotation of the 20 nucleotides downstream of the gene transcription termination site (TTS); (ii) estimation of the median read coverage within these 20 nucleotides, (iii) calculation of a 'split ratio' by dividing this median coverage by a predefined threshold set at 10; (iv) calculation of the read coverage for each nucleotide in the intergenic region before the TTS, and (v) comparison of this coverage with the split ratio. If the coverage of three consecutive nucleotides fell below this ratio, the operon was split at that position.

It is worth noting that the downstream length of 20 nucleotides, the split ratio of 10, and the 3-nucleotide window for determining operon splitting were arbitrarily chosen. Operons were labelled as NcOs if at least one gene on the opposite strand was flanked on both sides by genes from the same operon. Additionally, the number of reads overlapping with genes flanking the operon gene or with reverse genes was calculated for further information.

Taxonomic study of NcOs

In order to assess the taxonomic presence of NcOs we proceeded as follows. We downloaded the OMA database (https://omabrowser.org/oma/home/) (Altenhoff et al. 2020), which contained a broad set of species with gene order information. In par-

ticular this dataset contains 1719 bacterial genomes of which 32 are *Staphylococcus* species, including 23 *S. aureus* strains. Based on the reference genome, we also generated 100 random groups of consecutive genes of different sizes (three to seven) that were not related to known operons. We also downloaded the genomic sequences for the 32 *Staphylococcus* species included in the OMA database from ENA, then we used DFAST database (Tanizawa et al. 2018) to annotate the genomes and finally, we used exonerate (Slater and Birney 2005) to ensure all proteins present in the reference genome were also annotated properly in the genomes when present.

Given an operon, a BlastP search was performed for its sequences against each of the bacterial species included in OMA, limiting the blast to the best five hits. The analysis was also run against the 32 genomes of *Staphylococcus* reannotated by us. Results were then filtered using an e-value cut-off of $1e^{-05}$ and an overlap threshold of 0.5. Best hits were then kept and used to see whether they were found in contiguous positions in the genome. Up to three genes that were not part of the operon were allowed to be found between the homologous genes. Only results reporting at least two contiguous homologous genes were considered. For each new operon detected in an OMA species, we calculated a conservation score as previously used in Evolclust (Marcet-Houben and Gabaldón 2019). This score considers homologous genes, nonhomologous genes and the size of the two clusters. Given that sometimes the reference operon is much bigger than the found operon, scores can be below 0 as there are more nonhomologous genes than homologous genes. In such cases the scores were set to 0.

In addition to the search against OMA, we also searched for the operons in the String database to assess the conservation of their gene neighborhood across species and the connections between proteins encoded by the operon as predicted by String's different methodologies (Szklarczyk et al. 2022).

Western blots

01912, 01914, and GFP protein levels detection by western blot was monitored as follows. Bacteria were grown overnight in 5 ml of TSB at 37ºC under 200 rpm shaking conditions. In the case of GFP, $CdCl_2$ 2 mM was added to the media to induce its production. Cultures were diluted 1:1000 in 50 ml of TSB and incubated at 37°C under 200 rpm shaking conditions in 250 ml flasks until they reached an OD_{600} of 0.8. Bacterial cultures were centrifuged and pellets were washed once with 1 ml of PBS and resuspended in 500 µl PBS. Bacteria were mechanically lysed with a Fastprep apparatus (MP Biomedicals) at speed 6.0 m s^{-1} during 45 s at 4ºC. Then, tubes were centrifuged for 10 min at 14 000 rpm. Supernatants were collected and kept at −20°C until used. Protein concentration was determined by colorimetric technique using the Bicinchoninic Acid Kit (Sigma Aldrich) and the Epoch apparatus (Biotek). Samples were adjusted to 20 µg ml^{-1} of total protein and a volume of Laemmli buffer was added. Total protein extracts were denatured by boiling at 100ºC for 5 min. Proteins were separated on TGX Stain-Free FastCast Acrylamide Gels (Bio-Rad). For western blotting, proteins were blotted onto Amersham™ Protran™ Premium 0.45 µm nitrocellulose blotting membranes (Cytiva) by electroblotting. Membranes were blocked overnight in PBS containing 5% skimmed milk and 0.1% Tween 20 (PBST) under shaking conditions and incubated with monoclonal anti-FLAG antibodies labelled with horseradish peroxidase (Sigma) for 3x-FLAG labelled proteins strains. For GFP expressing strains, Mouse anti-GFP antibody (Sigma) diluted 1:1000 in blocking solution for 2 h at room temperature was used as primary antibody. Then, a three PBS 10 min washing in shaking conditions at room temperature was performed. Next, monoclonal anti-Mouse antibody labelled with horseradish peroxidase (Sigma) diluted 1:1000 in blocking solution was used as secondary antibody, with a 2 h at room temperature incubation conditions. 3xFLAG labelled proteins and GFP were detected with ECL Prime western blotting detection reagents (Cytiva) in a ChemiDoc MP Imaging System (Bio-Rad).

Oxygen consumption rate measurements

The Seahorse XF HS Mini Analyser (Agilent) was used to quantify oxygen consumption following the protocol established by Lobritz et al. (2015), with some modifications. Strains were grown in TSB overnight at 37°C. The cultures were then diluted 1:200 in TSB and cultured at 37°C until they reached an OD_{600} of 0.3. Cultures were then diluted 1:50 in PBS and 90 µl of each diluted culture was inoculated into Seahorse XF HS miniplates (Agilent). Plates were incubated at 37°C for 2 h. Media were renewed after this step and before Seahorse measurement. It is important to note that the maximum oxygen consumption rate (OCR) recorded on the Seahorse analyser is ∼700–800 pmol min^{-1}. When this maximum is reached, the OCR curve starts to fall because the oxygen consumption of the cells is higher than the oxygen replenishment of the system. The experiments were therefore stopped at this point.

After the Seahorse analysis, a dispersing treatment with Dispersing B enzyme (DspB; 40 µg ml^{-1}) was performed during 1 h and 30 min at 37°C in order to dissagregate bacterial cells (Valle et al. 2007). Serial dilutions up to 10^{-6} of the resuspended cells were prepared and cultured on TSB plates, overnight at 37°C. CFU were counted in order to normalize OCR values to the number of cells in each well. Three biological replicates were used for each strain.

Inmunofluorescence

Detection of 01914 and GFP proteins by immunofluorescence was performed as follows. Strains containing plasmids pCN51::*men* operon_*gfp* 01914_3XFlag or pCN51::*men* operon_*gfp* 01914 invert_3XFlag were grown overnight in 5 ml TSB at 37°C under shaking conditions at 200 rpm. Cultures were diluted 1:1000 in 20 ml TSB with erythromycin 10 µg ml^{-1} and $CdCl_2$ 5 µM and incubated in 100 ml flasks at 37°C under shaking conditions at 200 rpm until an OD_{600} of 0.8 was reached. Cultures were centrifuged and washed three times with PBS. Fixation was performed with 4% paraformaldehyde for 30 min at RT. A second fixation step was performed with EtOH 70%, which also aids in permeabilization. Permeabilization was performed with 0.1 mg ml^{-1} lysostaphin (10 min, RT). BSA 0.5% was used to block nonspecific antibody binding (overnight, 4ºC). The primary antibodies used were mouse anti-Flag M2 monoclonal antibody, diluted 1:1000 (Sigma Aldrich) for the 01914 protein, and rabbit anti-GFP polyclonal ab290, diluted 1:1000 (Abcam) for the GFP. Alexa Fluor 647-conjugated goat antimouse IgG (Invitrogen) and Alexa Fluor 488-conjugated goat antirabbit IgG (Invitrogen) were used as secondary antibodies, respectively. Hoechst was used to visualize DNA and localize bacteria. Samples were visualized using a Zeiss LSM 800 confocal scanning laser microscope.

Image analysis

Detection and quantification of bacteria in fluorescence images was performed using a plugin developed for Fiji/ImageJ, an open-source Java-based image processing software (Schindelin et al. 2012). The plugin was developed by the Imaging Platform of the Centre for Applied Medical Research

(CIMA, Pamplona) (https://cima.cun.es/en/research/technology-platforms/image-platforms).

First of all, individual bacteria were segmented from the DAPI channel using the following automated image processing pipeline: a background correction algorithm was applied, followed by an intensity threshold to create a mask for DAPI signal. A maxima detection algorithm was then applied to the original DAPI signal, and the detected maxima were used as seeds for a marker-driven watershed algorithm limited by the previously calculated segmentation mask (Legland et al. 2016). The resulting segmentation mask was used to measure the expression of red and green fluorescent signals in each individual bacterium, and the signal ratio between green and red was calculated. Bacteria with an average image intensity <1 for the red or green channel—very low expression, which can be considered as noise—were not included in the final quantification (Counts, WT: $n = 2198 \pm 212$; 01914 invert: $n = 2461 \pm 423$).

Menaquinone extraction and HPLC analysis

Each *S. aureus* strain was grown overnight in 5 ml TSB supplemented with antibiotics as required in a shaking incubator at 37°C. The culture was then diluted 100-fold in 5 ml TSB medium without antibiotics and incubated at 37°C until mid-log phase (A_{600} ~0.3–0.7). A volume of 4 ml of the culture broth was centrifuged at 8000 × *g* for 5 min at 4°C, washed with PBS, and suspended in 625 µl of PBS supplemented with 25 mM EDTA and 5 µM α-tocopherol. A volume of 3 ml of hexane–ethanol (5:2) was added to the suspension and mixed vigorously for 2 min, and the upper layer was collected after centrifugation at 4000 × *g* for 3 min. The bottom layer was rinsed with 1 ml of hexane–ethanol, and the top layer was pooled, dried *in-vacuo*, dissolved in 100 µl of ethanol, and 20 µl of the sample were analysed using Waters Alliance high-performance liquid chromatography (HPLC) system equipped with a CAPCELL CORE C18 2.7 µm column (3.0 φ × 150 mm) (SHISEIDO, Tokyo, Japan). HPLC analysis was performed with isocratic elution of ethanol : methanol (1:1) at the flow rate of 0.4 min ml^{-1} and column temperature at 40 °C. Detection was performed using a fluorescence detector (Waters, wavelengths 320 and 430 nm for excitation and emission, respectively, after post-column reduction using a platinum column (4.0 mm × 20 mm, GL Sciences, Tokyo, Japan)). The total amount of MKs was calculated as the sum of the peak areas of MK7, MK8, and MK9 (RT 6.9, 9.1, and 12.4 min, respectively) and normalized to culture A_{600} (Hamamoto et al. 2015).

Results

Direct RNA sequencing identification of NcOs in the *S. aureus* genome

We used Oxford nanopore technology to directly sequence native RNA molecules from strains *S. aureus* MW2 and *S. aureus* RN10359. We produced three independent libraries (one from *S. aureus* MW2 and two from *S. aureus* RN10359) without any amplification process on separate occasions. Nanopore sequencing libraries yielded a total of 356 600, 922 620, and 1 902 936 reads that mapped to the *S. aureus* NCTC8325 genome with an average mapped read length of 733 bp, 749 bp, and 689 bp, respectively (Table S4 and Fig. S1, Supporting Information). To identify NcOs (Fig. 1A), we first calculated all possible theoretical NcOs present in the *S. aureus* genome by identifying all codirectional gene pairs that were separated by one to three genes in the opposite direction, regardless of the intergenic distance or functional relationship between the genes. Among these, we needed to select those sets of codirectional genes that were transcribed together despite being separated by at least one antisense gene. To make this selection, we determined that NcOs had to meet two conditions: first, there should be a minimun of one long transcript that overlaps the two cooriented genes and second, the RNA-seq read coverage signal should remain relatively stable within the intergenic region where the antisense gene/s is located. The first condition was analysed by visual inspection of the mapped reads using Geneious (see the section 'Materials and methods'). To evaluate the second condition, we calculated the mean signal coverage in the 20 nucleotides upstream of the end of each annotated gene. The resulting value served to determine the threshold for the mean coverage that had to be maintained in the intergenic region to establish gene cotranscription. If the average coverage at three consecutive nucleotides in the intergenic region dropped below 0.1 reads per nucleotide, we did not consider the two cooriented genes as cotranscribed. Conversely, if the average coverage remained above this threshold across the entire intergenic region, we considered the cooriented genes to be transcribed together. Our analysis led to the identification of a total of 18 operons scattered throughout the genome that meet both criteria (Fig. 1B). It is important to note that when the libraries were mapped against the genome of strain MW2, some NcOs were not detected because genes encoding small peptides are not annotated in the genome of this strain (data not shown). In addition, four more NcOs were located within the genome of the temperate 80α bacteriophage (Fig. 1B). The majority of NcOs have a single gene that is oriented in the opposite direction to the rest of the operon. The only exception to this pattern is observed in the *02997* NcO, where three genes are transcribed in the opposite direction to the rest of the operon (Figs 1B and 2).

To enhance the reliability of the identified NcOs, we performed RT-PCR experiments using oligonucleotides designed to bind to sequences adjacent to the gene(s) transcribed in the opposite direction to the rest of the operon. This design ensured that the PCR reaction would only yield a product if the transcripts completely covered both the codirectional genes and the region occupied by the gene(s) in the opposite direction (Fig. S2, Supporting Information). Our results confirmed the cotranscription of genes that were separated by an antisense gene in 16 out of the 18 identified NcOs in the bacterial genome. The two NcOs that could not be confirmed by RT-PCR were the *01073* and *02997* NcOs. In summary, these results indicate that the *S. aureus* genome harbours a minimun of 16 NcOs that meet the criteria of having two sets of cotranscribed genes, and one or more interspersed genes that are transcribed in the opposite direction. In addition, the 80α phage genome contains four NcOs that also fulfill these criteria.

Presence of NcOs across taxonomic clades

For each of the 18 NcOs identified in the *S. aureus* genome, we conducted a thorough search to determine whether they are present in the 1719 bacterial species included in the OMA database (Altenhoff et al. 2020). We assigned each species to a specific taxonomic category based on its most recent common ancestor with *S. aureus*, resulting in 32 *Staphylococcus*, 1 *Staphylococcaceae*, 77 *Bacillales*, 136 *Bacilli*, 109 *Bacillota*, and 1364 *Bacteria*. As the *Staphylococcaceae* category was underrepresented, it was merged with the *Bacillales* category. We considered an NcO to be present when all of its constituent genes were identified with no more than two genes in between them. Interestingly, we discovered that 16 out of the 18 *S. aureus* NcOs were extensively distributed among *Staphylococcus* species, being present in over 70% of the strains. On the

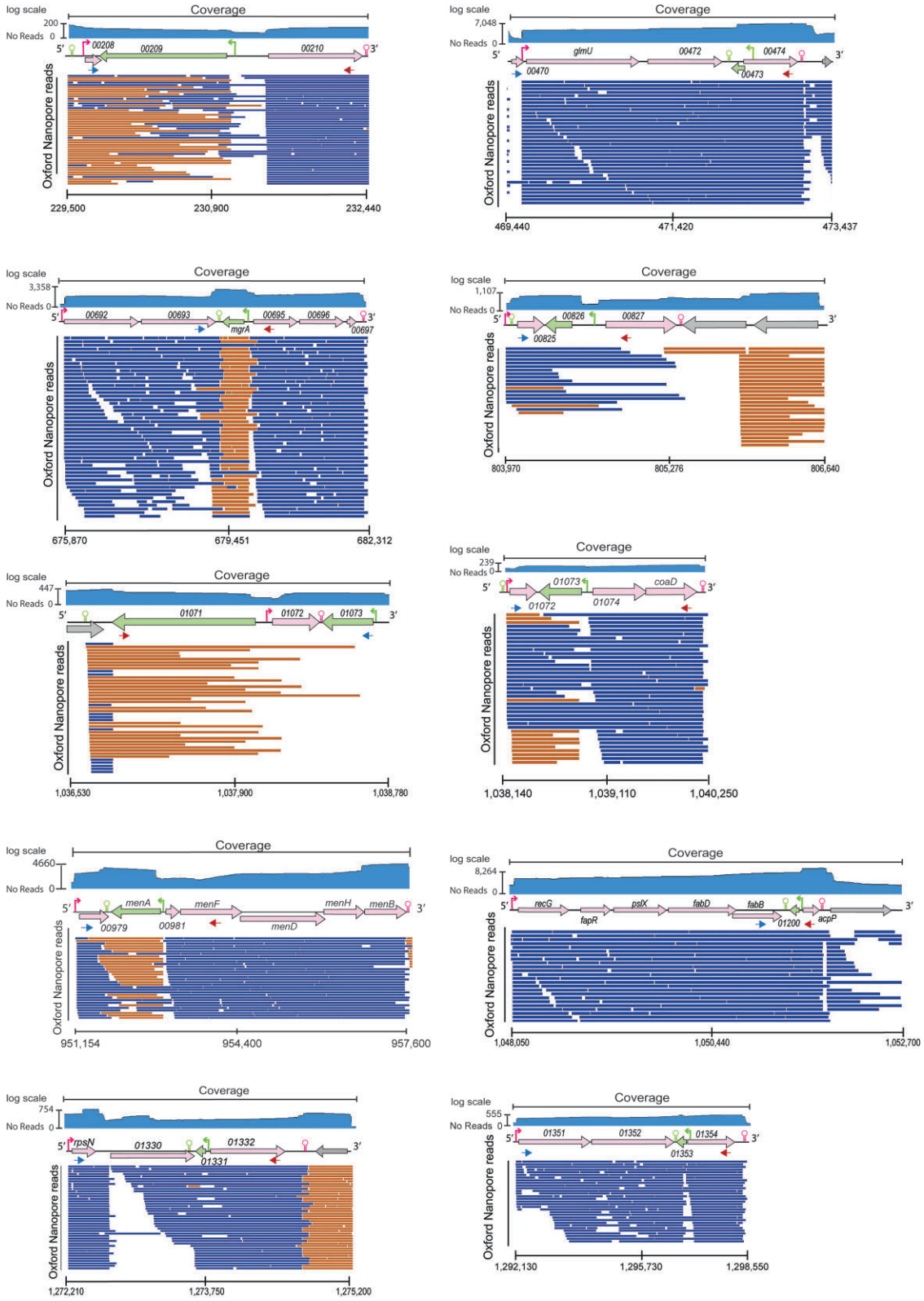


Figure 2. Oxford Nanopore mapping for the 18 NcOs found in *S. aureus* RN10359. Images from the Geneious software showing a fraction of the uniquely mapped long RNAs. Genes located on the forward strand are shown in pink. Genes on the reverse strand are shown in green. Each read mapped is shown as a blue line (sense reads) or an ochre line (antisense reads). Genomic positions in the *S. aureus* NCTC8325 genome are indicated along the x-axis. Gene annotation corresponds to the *S. aureus* NCTC8325 strain. Read coverage per nucleotide is shown on a logarithmic scale. Promoters (arrow) and terminators (hairpin) are shown. Thin arrows indicate the position of the oligonucleotides used for the RT-PCR reaction.

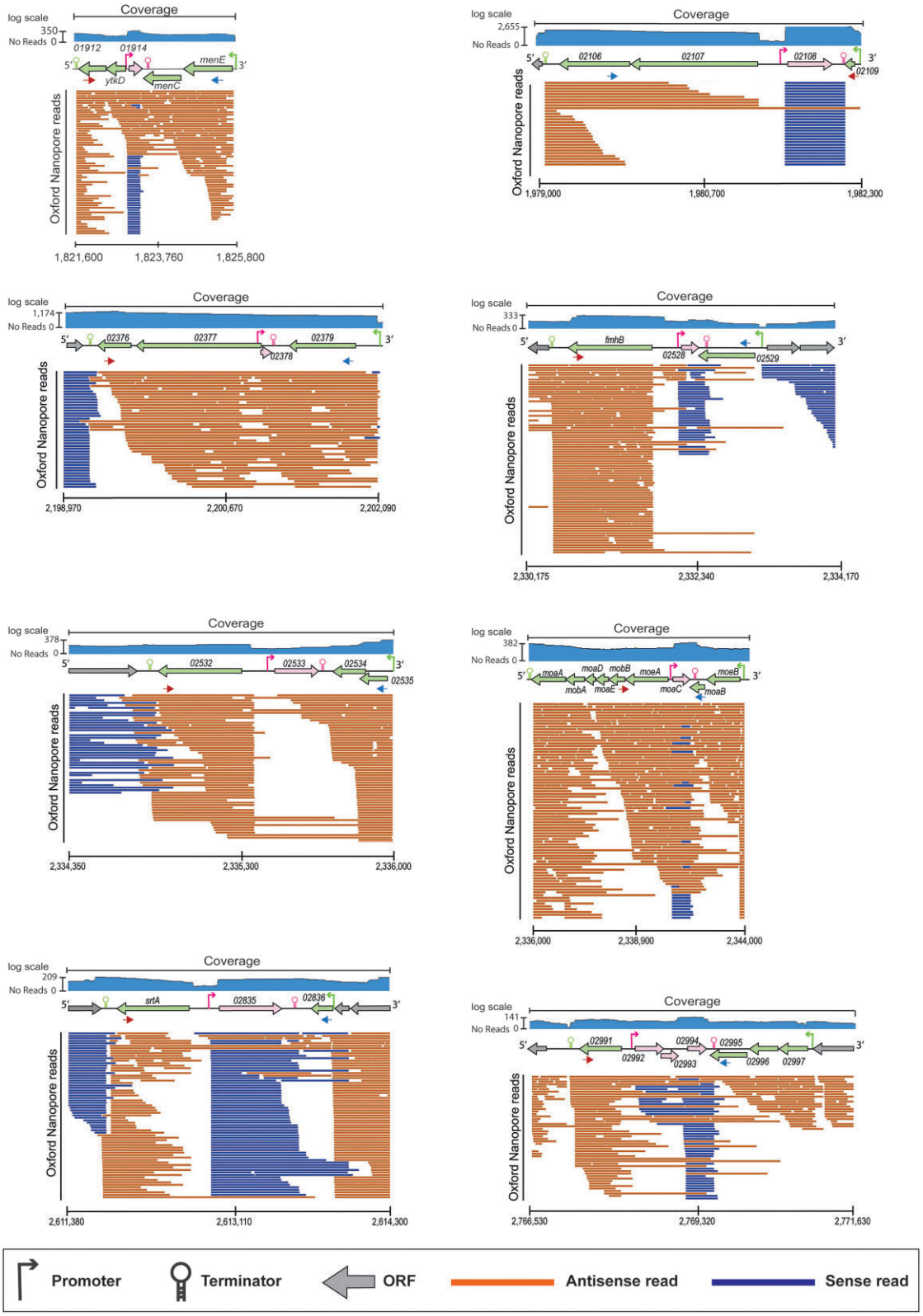


Figure 2. Continued.

other hand, *01072*, *01073*, and *moeB* NcO conservation of gene distribution extended to the *Bacillales* group (Fig. 3A).

To obtain a more comprehensive understanding of the conservation patterns of these NcOs, we decided to calculate a conservation score instead of solely determining strict operon presence (see the section 'Materials and methods'). This score was calculated for cases where at least two genes within an operon were found adjacent to one another in a given species. In order to compare the conservation degree of NcOs we built three different sets of gene clusters. The first contained the group of 18 NcOs of *S.*

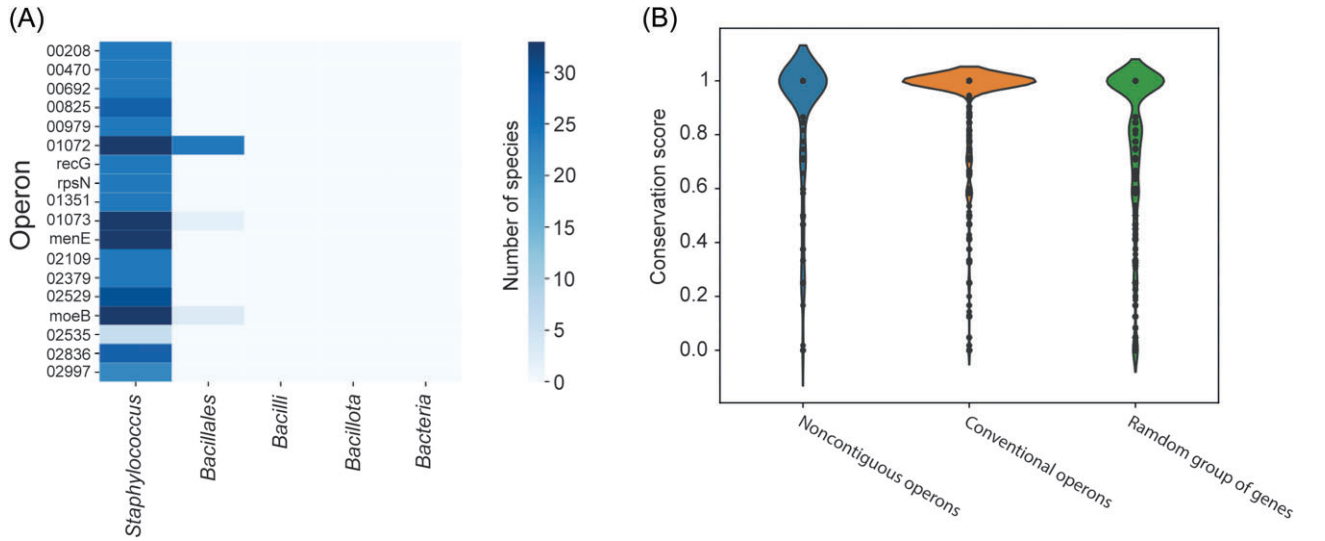


Figure 3. Conservation of gene neighbourhoods in NcOs. (A) Heatmap showing the number of species within each taxonomic group within the *Staphylococcus* lineage that contain a given NcO. Each NcO is named after the first gene that is cotranscribed with other genes in the operon. (B) Violin plot showing the distribution of conservation scores across three different datasets. Conventional operons are more conserved than NcOs (*P*-value NcOs vs conventional operons: 1.66 e^{-11}, NcOs versus random: 1.03 e^{-16}, and conventional operons versus random: 0.0).

aureus. The second set comprised 273 conventional operons that are present in the *S. aureus* genome and are composed of three or more genes. Finally, the last set comprised 700 randomly selected groups of neighbouring genes of varying sizes. To avoid biases due to gene prediction issues, specially regarding the presence of small proteins, we reannotated the 32 *Staphylococcus* genomes found in the OMA database ensuring that all proteins found in the reference were also found in the strains when present (see the section 'Materials and methods'). As expected, both the set of NcOs and conventional operons exhibited a higher degree of gene order conservation when compared to the random dataset. Notably, our results also indicated that conventional operons display a higher level of conservation when compared to NcOs (Fig. 3B).

NcOs protein–protein interaction network

We then conducted an extensive search in the String database to discover connections between the proteins encoded by the genes within the NcOs (Szklarczyk et al. 2022). Out of the 18 *S. aureus* NcOs, String revealed relationships between the proteins encoded on both sides of the antisense gene in five cases (Fig. 4; Fig. S3, Supporting Information). In the *moeB* operon, all nine proteins are associated with the biosynthesis of the Mo-molybdopterin cofactor, as indicated by GO term enrichment (Fig. 4). Within the *01072* operon, proteins encoded by genes from both sides of the antisense gene are associated with Coenzyme A biosynthesis and phosphopantothenoylcysteine decarboxylase activity, while *coaD* is linked to the others by common gene neighbourhood (Fig. 4). In the recG operon, all genes upstream and downstream of the antisense gene encode for proteins related to fatty acid and phospholipid metabolism (Fig. 4). In the 02997 NcO, proteins encoded by genes upstream and downstream of the antisense genes are associated with ROS, while the function of the three proteins encoded by genes on the opposite strand is unknown (Fig. 4). Regarding the *00470* NcO, there are no known functional relationships between *00474* and other genes of the NcO, but experimental evidence indicate that *00474* is coexpressed with all the genes located upstream of the antisense gene in *B. subtilis* (Hilden et al. 1995).

Importantly, in seven NcOs (*00692*, *00825*, *00979*, *01073*, *menE*, *02109*, and *02836*), String analyses showed that the function of the antisense gene is related to the other proteins encoded in the NcO (Fig. S3, Supporting Information). On the other hand, there are three examples (operons *rpsN*, *01351* and *02379*) for which we could not establish a functional relationship between their encoded proteins, though String associated them due to conserved gene neighborhood (Fig. S3, Supporting Information). Lastly, there are only three NcOs (*00208*, *02529*, and *02535*) whose encoded proteins are either associated just due to the proximity of conserved genes or not functionally associated.

Taken together, these results indicated that neighbourhood of genes from NcOs is conserved at least across the *Staphylococcus* genus, and in some cases they remain closely associated even within the *Bacillales* order. The presence of functional associations between the proteins encoded by the genes flanking the antisense gene strongly reinforces the idea that NcOs constitute a transcriptional unit connecting the expression of genes responsible for functionally related proteins.

Genes responsible for synthesizing menaquinone are arranged into two NcOs

To investigate the biological significance of NcOs, we focused on menaquinone synthesis because two of the identified NcOs are associated with this process (*menE* and *00979* NcOs) (Fig. 1). Nanopore sequencing yielded a total of 1510 and 9168 reads mapped to the *menE* and *00979* NcOs with an average mapped read length of 874 nts and 718 nts, respectively. Importantly, 35 single reads of the menE NcO and three reads of the *00979* NcO completely spanned the *01914* and *menA* antisense genes, respectively. The String analysis unveiled a significant co-occurrence of the genes from these two NcOs across numerous *Staphylococcus* species, suggesting the potential importance of their configurations in the regulation of menaquinone synthesis (Fig. S4, Supporting Information). It is worth noting that the *00979* NcO contains a gene encoding a small peptide (SAOUHSC_00981) whose presence has only been annotated in a few strains. An examination of the functional associations between the proteins en-

Figure 4. Functional associations between proteins encoded in *moeB*, *01072*, *recG*, *00470*, and *02 997* NcOs as indicated by the String database (A) Co-occurrence of the genes within the NcO on the phylogenetic tree obtained by String. Genes are represented as arrows at the top of the figure. The red colour scale represents the degree of co-occurrence of genes in the genomes analyzed. (B) String analysis showing the functional associations between the proteins encoded by the genes within the NcOs. Protein IDs are provided. The strand indicates the orientation of the gene: sense (+) and antisense (–). In MoeB and 01072, the function of the antisense gene is related to the other proteins encoded in the NcO, whereas in recG, 00470 and 02997, the proteins encoded by genes from both sides of the antisense gene are associated according to String.

coded in the *menE* NcO divides them into two groups: MenEC proteins are associated to menaquinone synthesis, while the functions of 01914, YtkD, and 01912 remain unidentified. In the case of the *00979* NcO, the antisense gene (*menA*) and the *menFDHB* genes are involved in menaquinone synthesis, while the function of *00979* itself remains unknown.

In a previous study using the *menE* NcO as a model, we observed that overexpression or deletion of the operon promoter affected the expression of the *01914* antisense gene in a process dependent on RNaseIII activity and transcriptional interference (Sáenz-Lahoya et al. 2019). To further explore the impact of NcO architecture on gene expression, we first generated a mutant in which the chromosomal copy of the *01914* gene was inverted (*S. aureus 01914* invert) (Fig. 5A). This inversion allowed all genes in the *menE* NcO to be transcribed as a single polycistronic transcript. Secondly, we constructed another mutant in which the promoter region upstream the *01914* gene was deleted and the *01914* gene and its promoter were inserted next to the innocuous *attB* site of the lipase gene (*S. aureus* Δ*P01914*-trans) (Fig. 5A). Thus, in this strain, expression of *01914* from another location of the chromosome prevents the transcriptional interference mechanism of gene regulation in the *menE* NcO. It is also important to note that both strains encode for 3xFlag tagged 01912 and 01914 proteins, so that their expression levels could be monitored by Western Blot analysis and compared to those in a wild type tagged strain (*S. aureus* Wt) (Fig. 5A). Western blot analysis revealed that 01914 levels decreased when the *01914* gene was inverted and transcribed along with the rest of the genes in the operon (Fig. 5A). In contrast, 01912 levels increased when the *01914* gene was inverted or expressed from a distinct chromosomal location. To further confirm that the presence of the antisense gene (*01914*) influences the expression of the NcO genes upstream of it, we cloned the entire wild-type operon and the complete *01914*-invert operon, fused to a *gfp* gene downstream of the last gene (*01912*), in a plasmid under the control of a cadmium-inducible promoter (Fig. 5B). Hence, these plasmids allowed the expression of wild type or *01914* inverted operons that contain a last *gfp* gene as part of the NcO. When NcO expression was induced with cadmium, GFP levels significantly increased only in the case of the *01914* inverted operon (Fig. 5B). Next, we analysed GFP and 01914 protein expression at the individual cell level using confocal microscopy. Consistent with western blot results, confocal microscopy revealed that 01914 and GFP levels were inversely related, with increasing levels of GFP expression coinciding with decreasing levels of 01914 (Fig. 6B and C), confirming that GFP production was significantly higher in bacteria containing the *01914* inverted form of the NcO compared to bacteria containing the WT operon. Taken together, these results suggest that expression of the antisense gene (*01914*) in an NcO uncouples the expression levels of upstream and downstream genes.

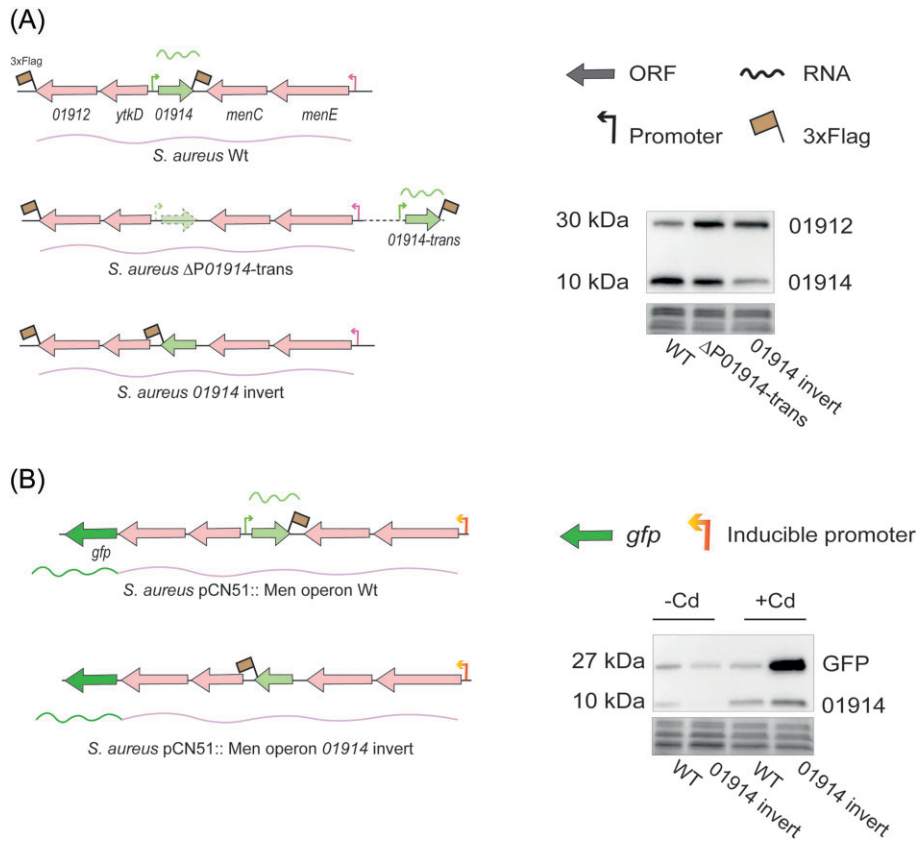


Figure 5. Analysis of the impact of NcO architecture on gene expression. (A) Schematic representation of the constructions used in this study (left) and western blot analysis of 01914 and 01912 proteins produced from the chromosomal copy of WT, Δ*P01914*-trans and *01914*-invert strains (right). The 3xFlag-tagged proteins were detected with commercial anti-3xFlag antibodies. The molecular weight of the proteins is indicated. (B) Schematic representation of the constructions used in this study (left) and western blot analysis of GFP and 01914 proteins produced from plasmid pCN51 containing the *menE* NcO in its WT or *01914* inverted version in the absence (−Cd) or presence (+Cd) of cadmium (right). The molecular weight of the proteins is indicated. For panels (A) and (B), Coomassie-stained gels were used as a loading controls. The data shown are representative of experiments performed in biological triplicates.

Disruption of the *menE* NcO architecture increases menaquinone synthesis and reduces oxygen consumption

The expression of enzymes that control menaquinone synthesis must be tightly controlled since elevated levels of menaquinone are known to be toxic to bacteria due to increased oxidative stress (Johnston and Bulloch 2020). We decided to investigate whether disrupting the architecture of the *menE* NcO had any effect on menaquinone synthesis. To do this, we compared menaquinone accumulation levels in *S. aureus* WT, Δ*P01914*-trans and *01914*-invert strains. As a control, we included a *S. aureus* Δ*Pmen* mutant, which has a partial deletion of the operon promoter, leading to a significant reduction in operon expression (Sáenz-Lahoya et al. 2019). The results showed that disruption of the NcO structure led to an increase in menaquinone accumulation (Fig. 7A), without affecting the ability to produce different menaquinone isoforms (MK7, MK8, and MK9) (Fig. 7C).

Within the *menE* NcO, only the *menEC* genes are known to encode enzymes involved in the synthesis of menaquinone. The function of the YtkD, 01912 and 01914 proteins is unknown. However, the coexpression of *ytkD* and *01912* with *menEC* suggests that the encoded proteins may be involved in menaquinone synthesis. To investigate this hypothesis, we overexpressed the *ytkD*, *01912* or *01914* genes in the wild type strain, and quantified the accumulation of menaquinone (Fig. 7B). Overexpression led to a significant increase in menaquinone levels compared to those in the WT strain. Although the understanding of the mechanism behind the rise in menaquinone levels caused by overproduction of these proteins requires further analysis, these results strongly suggest that the proteins produced by the genes within an NcO operon, similar to conventional operons, are involved in the same biological processes.

Menaquinone is an essential component of the electron transport chain, and thus of the bacterial respiratory process. Under aerobic conditions, oxygen is the final electron acceptor. To investigate whether the increased accumulation of menaquinone caused by NcO disarrangement has any impact on the efficacy of the electron transport chain, we compared the OCR of the WT strain and the Δ*P01914*-trans and *01914*-invert strains. The results demonstrated that alterations of the NcO structure resulted in a reduction in the rate of oxygen consumption (Fig. 7D), reinforcing the notion that the proteins encoded in an NcO are functionally interconnected to optimize the efficiency of the metabolic pathway they regulate. Finally, we compared the growth rate of bacteria overexpressing the WT or *01914* inverted versions of the operon. Bacteria producing the *01914*-invert operon showed a reduced growth rate compared to the same strain overexpressing the regular operon, indicating that uncoordinated expression of operon genes has detrimental effects on bacterial fitness (Fig. 7E). Taken together, these findings suggest that all the proteins within the *menE* NcO operon play a role in controlling menaquinone levels to ensure their ideal balance for an efficient respiratory chain.

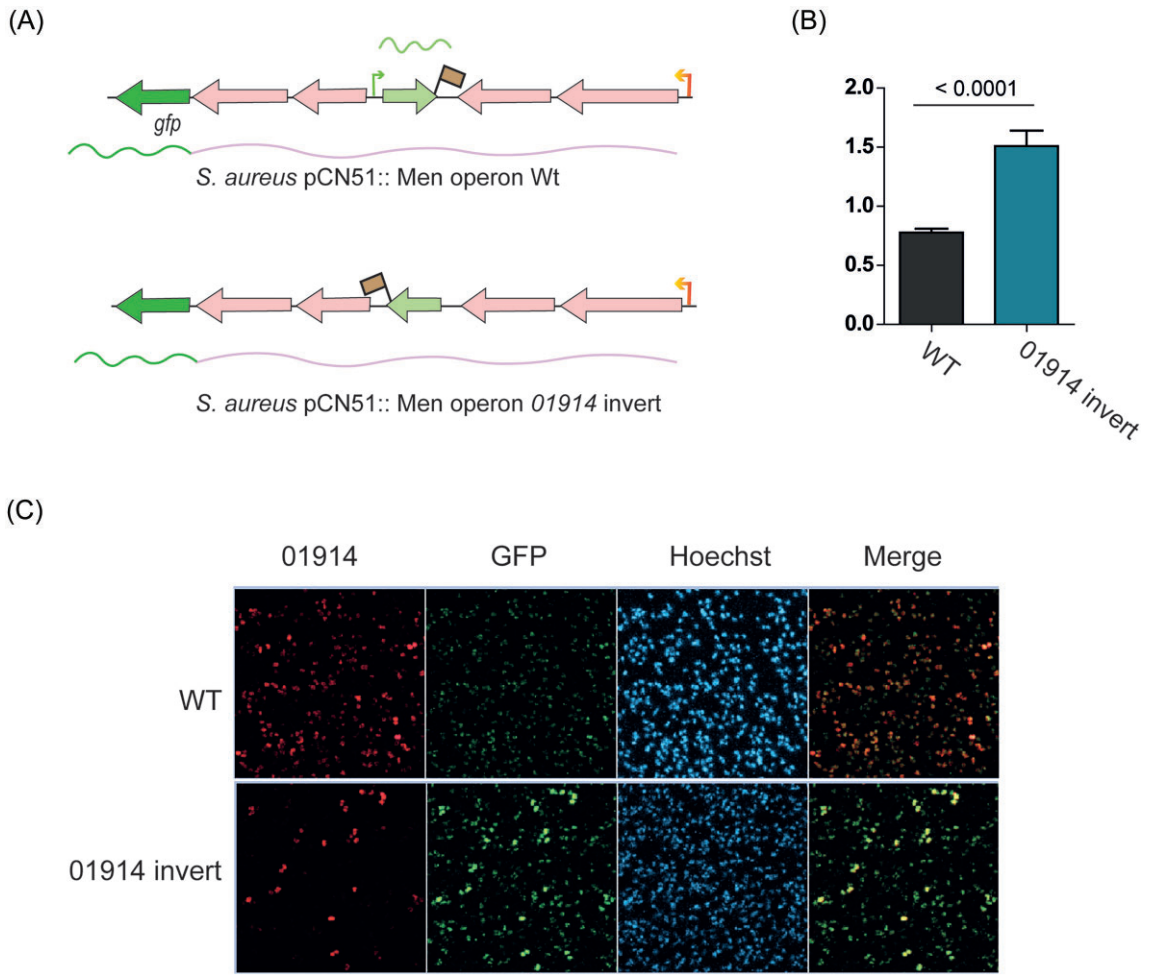


Figure 6. Analysis of the impact of NcO architecture on gene expression at single cell level. (A) Schematic representation of the strains used in this study. (B) Graphical representation of fluorescence quantification of strains containing pCN51::*menE* NcO WT or pCN51::*menE* NcO 01914 invert. The ratio of GFP to 01914 protein levels was calculated from the fluorescence intensity. 01914 was stained with anti-3xFLAG antibody conjugated to Alexa Fluor 647. GFP was immunostained with anti-GFP antibody conjugated with Alexa Fluor 488. Numerical data represent the mean ± SD of three independent experiments (counts, WT: $n = 2198 \pm 212$; 01914 invert: $n = 2461 \pm 423$). (C) Representative immunofluorescence microscopy images of *S. aureus* complemented with the pCN51 plasmid carrying the *menE* NcO in its WT or 01914 inverted version, grown to an OD_{600} of 0.8 at 37°C in TSB media.

Discussion

The prediction of operons in bacterial genomes serves multiple purposes, aiding in the comprehension of gene expression regulation, the identification of essential proteins for specific biological functions, and facilitating precise genetic modifications in bacterial chromosomes. Advances in transcriptomic methodologies have revealed previously unrecognized genetic configurations that could not have been deduced solely from genomic information (Byrne et al. 2017, Depledge et al. 2019, Shi et al. 2023, Xu et al. 2023). In this report, we use direct long-read RNA sequencing with nanopore technology to identify full-length transcripts in *S. aureus*. Sample preparation for nanopore long-read sequencing does not require PCR amplification, thus reducing the biases associated with amplification procedures, albeit at the cost of reduced sequencing throughput (Garalde et al. 2018). Consistent with previous studies, our RNA-seq libraries contain reads of up to 8 kb. However, coverage by full length reads was low for the majority of transcripts larger than 4 kb (Soneson et al. 2019). The identification of NcOs within the *S. aureus* genome relied on two assumptions: first, the mean number of reads mapping in the intergenic region between two codirectional genes that are interrupted by one or more genes in the opposite direction should not drop below 0.1 reads nt^{-1} relative to the average coverage of the adjacent genes. Second, the presence of at least one read connecting the two codirectional genes. This threshold might be considered too cautious. However, it is important to consider that those NcOs with high expression of counter-transcribed genes can lead to significant reductions in read coverage due to the action of RNaseIII and transcriptional interference. Conversely, if the expression levels of the operon genes are elevated, they can completely switch off the expression of the antisense gene. Consequently, the map of NcOs may exhibit slight variations depending on factors such as the environmental conditions during bacterial growth for RNA purification, the chosen threshold for determining cotranscription, and even genome annotations (Mäder et al. 2016). Of the 22 identified NcOs (including four from the phage 80α), eight have a gene encoding a peptide of less than 50 amino acids. In most cases (seven out of eight), the gene responsible for this short peptide aligns with the gene transcribed in the opposite direction. Since genome annotations frequently omit genes encoding peptides of less than 50 amino acids, NcOs containing such genes may escape detection.

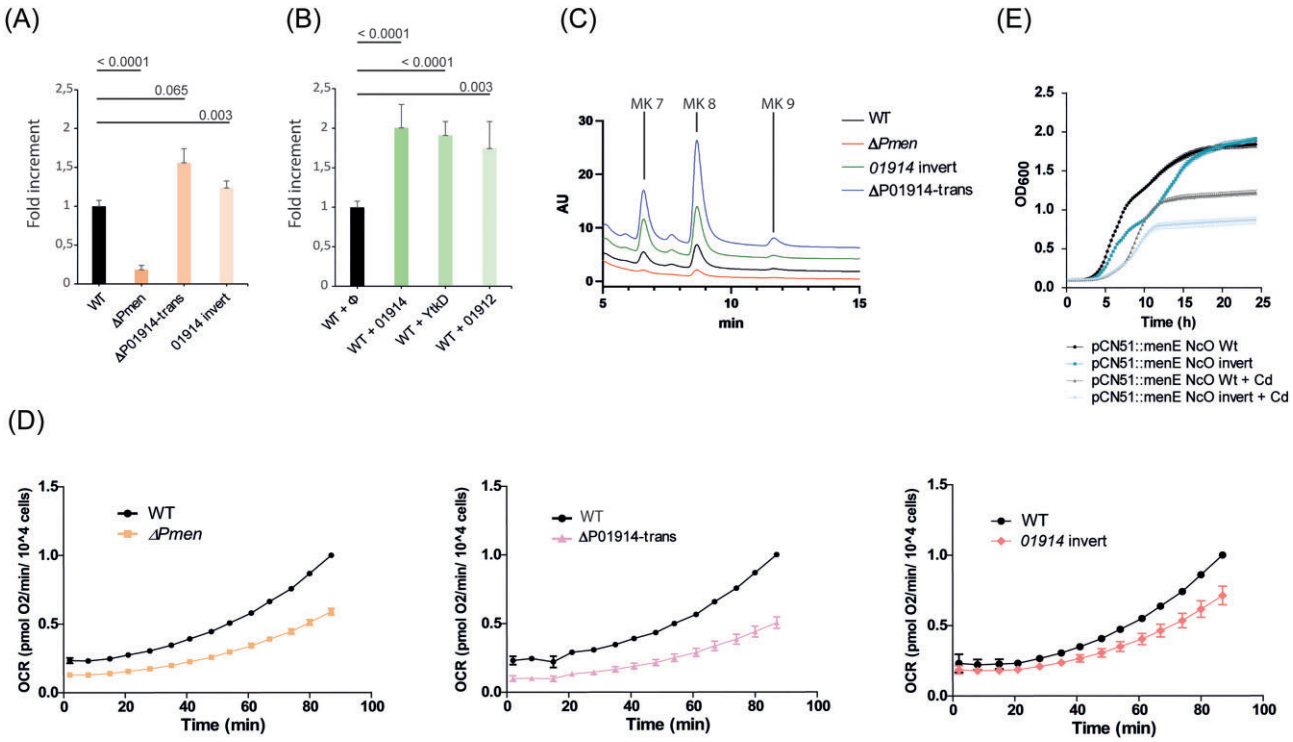


Figure 7. Effect of *menE* NcO architecture on menaquinone levels and OCR. (A) Quantification of total menaquinone levels in *S. aureus* WT, Δ*Pmen*, Δ*P01914*-trans, and *01914*-invert strains by HPLC. (B) Quantification of total menaquinone levels in *S. aureus* overproducing YtkD, 01914 or 01912 proteins. Control strains carry an empty plasmid. (C) Evaluation of the influence of *menE* NcO disruption on the accumulation of different menaquinone species. (D) Assessment of the OCR using the Seahorse XF (Agilent) analyzer by *S. aureus* WT, Δ*Pmen*, Δ*P01914*-trans, and *01914*-invert strains. (E) Growth rate of *S. aureus* complemented with the pCN51 plasmid carrying the *menE* NcO in its WT or *01914* inverted version, both in the presence or absence of cadmium, for a 24-h period at 37°C with shaking.

An intriguing aspect of the transcriptional architecture in NcOs is that transcription has to persist across extended regions that do not undergo translation since they correspond to the region where the antisense gene is located. In such cases, the coordination between RNA polymerase and ribosome movement is disrupted because the RNA polymerase needs to maintain its activity even when the ribosome halts at the conclusion of the upstream gene (Adhya and Gottesman 1978, Richardson 1991, Das 1993, Henkin 1996). This scenario may appear to contradict the long-accepted dogma in bacterial genetics that transcription and translation are intricately linked. However, a recent investigation analysing this coupling within the bacterium *B. subtilis* has revealed that RNA polymerase activity in this bacterium remains unaffected by the translation process (Johnson et al. 2020). These findings imply that the coupling between transcription and translation described in *E. coli* does not apply to Gram-positive bacteria and that the ribosomes progress independently of the RNA polymerase in a phenomenon termed 'runaway transcription'. Hence, the transcription of the extensive intergenic regions within the NcO remains compatible with the absence of translation.

In bacteria, the relationship between gene expression and genome organization is facilitated through operons, which assemble genes with related functions into a transcriptional unit. NcOs might have arisen from the insertion of a gene into the middle of an existing operon, connecting through antisense RNA mechanisms. The analysis of the conservation of NcOs identified in this study within *S. aureus* unveiled that among the 18 operons examined, 16 are consistently present in various *Staphylococcus* species, encompassing more than 70% of the strains investigated. However, only a limited number of operons exhibit a broader distribution, with *01073*, *01072*, and *moeB* operons being the most notable examples. The *moeB* operon is particularly noteworthy as all its constituent genes are intricately associated with molybdenum metabolism. Within the *moeB* NcO, the MoaC enzyme is encoded in the opposite direction. This enzyme plays a pivotal role as one of the initial enzymes in the molybdenum metabolism pathway, responsible for generating the precursor pyranopterin monophosphate (cPMP) from GTP, which subsequently forms the molybdenum cofactor (Iobbi-Nivol and Leimkühler 2013). It appears that the bacterium deliberately avoids transcribing this gene alongside the other operon genes and instead connects its expression via antisense regulation. One plausible explanation for this organizational choice is that the activities of the MoaC enzyme may either conflict with those of the other metabolic pathway enzymes or are required when the other enzymes in the metabolic pathway are inactive. According to this explanation, functional analysis of NcO cI from phage 80α (Fig. 1) has shown that this genetic organization allows the expression of a protein (ORF3) that confers immunity to host cell infection by other phages to be completely inhibited when the phage initiates its own replication, thus preventing an autoimmune response (Rostöl et al. 2024).

The results from String analysis also reveal that in the *00470*, *recG*, *01072*, and *02997* NcOs, proteins from the two gene groups, separated by the gene in the opposite direction, share a related function. This finding underscores the significance of creating NcO maps for a bacterium, as it enables the discovery of functional relationships between genes that might otherwise remain undetected. Conversely, the fact that these genes are functionally associated reinforces the idea that NcOs, much like conventional operons, group genes participating in the same metabolic path-

way. The integration of data obtained from NcO maps, in conjunction with the information available in newly developed databases that predict protein function based on their structural attributes (such as Foldseek) may be of great help in the future in deciphering the function of the large proportion of bacterial genomes whose function is still unknown (Barrio-Hernandez et al. 2023).

To investigate the significance of the NcO structure in the physiology of *S. aureus* bacteria, we employed the NcO comprising *menE-menC-01914-yktD-01912*, which plays a role in menaquinone synthesis. Menaquinones are lipid-soluble compounds that serve as electron carriers, facilitating the transfer of electrons between membrane-bound protein complexes within the electron transport chain. The energy released during this electron transfer process is used to generate an electrochemical gradient, ultimately leading to the production of ATP (Kurosu and Begari 2010). Menaquinones have also been implicated in sensing alterations in redox state and oxidative stress in *E. coli* (Bekker et al. 2010) and with persistance and biofilm formation in *S. aureus* (Mashruwala et al. 2017). The process of synthesizing naphthoquinone moiety of menaquinone from chorismite involves the activity of at least eight enzymes (MenB to MenH) (Johnston and Bulloch 2020). Regulation of menaquinone can take place at protein level, through sublocalization of the Men enzymes to ensure that menaquinone is produced exclusively where it is needed. Additionally, the Men enzymes can also be allosterically inhibited by binding of final metabolites from the pathway to their catalytic domains, preventing the excessive accumulation of menaquinone (Bashiri et al. 2020). In a previous study, we analyzed the effect of RNaseIII and transcriptional interference on the coordination of expression of the *menEC-MW1733-ytkD-MW1731* operon genes (*menEC-01914-ytkD-01912* according to NCTC8325 annotation) and observed that overexpression of MW1733 inhibited *menEC* expression and caused the appearance of small colony variants (Sáenz-Lahoya et al. 2019). In this work, we have analyzed the consequences of this transcriptional architecture on the biology of the bacterium. Our results suggest that NcO architecture contributes to the transcriptional control of the enzymes involved in menaquinone synthesis. Specifically, the *menEC* and *menAFDHB* genes are organized in two separate NcOs. In the case of *00979* NcO, it includes a gene of unknown function (*00979*) that cotranscribes with the *00981-menFDHB* genes, forming a long polycistronic transcript that overlaps with the mRNA of the *menA* gene. This *menA* gene is located between *00979* and *00981* genes, and it is transcribed in the opposite direction. It is striking that *menA*, which is involved in the penultimate step of a cascade of reactions for menaquinone synthesis (Boersch et al. 2018), is not part of the same operon as other enzymes in the pathway, but is located contiguous to the operon but transcribed in the opposite direction (something similar to the *moaC* case). MenA serves as the rate-limiting enzyme responsible for linking the naphthoquinone moiety and the isoprenyl chain in the final stage of menaquinone synthesis. In situations where naphthoquinone accumulates, it may be necessary to upregulate MenA, while MenFDH, which is involved in the initial stage of naphthoquinone synthesis, may need to be slightly downregulated. The NcO architecture offers an intricate feedback mechanism for precise regulation of menaquinone production, thereby preventing the buildup of toxic intermediates.

An important aspect of transcriptomic analyses performed on purified RNA from bacterial populations is that the presence of sense and antisense RNAs at the same genomic position does not necessarily mean that both transcripts are present simultaneously in the same bacterium. The result of transcriptomes obtained with RNA purified from bacterial populations is the same whether the sense and antisense transcripts are produced by different groups of bacteria or whether both transcripts are produced by the same bacterium at the same time. The only way to resolve this uncertainty is to perform transcriptomic analyses at the level of individual cells. However, the efficiency of the technique remains low in *S. aureus*, probably due to problems with the efficiency of bacterial lysis at the single cell level (Blattman et al. 2020).

The strategy of clustering functionally related genes is a phenomenon observed across all domains of life (Komili and Silver 2008, Nützmann et al. 2018). In eukaryotes, this clustering does not involve cotranscription but rather clustering in regions of the genome where chromatin has similar levels of accessibility to the transcriptional machinery. In bacteria, clustering of genes for simultaneous expression is a fundamental aspect of gene regulation. This study shows the need to expand the operon concept to include variations that do not conform to the traditional definition. NcOs enable both the cotranscription of genes and the regulation of gene expression in the opposite direction, revealing previously overlooked connections between proteins participating in the same biological process. More generally, our findings demonstrate that the combined utilization of cotranscription and antisense transcription can be used to coordinate the expression of different biological processes.

Acknowledgement

We express our gratitude to Alvaro G. Hernandez, Chris Field, and Lindsay V. Clark from the University of Illinois at Urbana-Champaign for providing technical support and guidance with the RNA direct sequencing and bioinformatic analysis using the Nanopore platform. Special thanks to Eva Santamaria-Monasterio from CIMA-Universidad de Navarra for providing technical support with Seahorse XF HS Mini Analyser.

Author contributions

P. I-S: investigation, visualization, methodology, validation, writing first draft. A.SM: investigation; H.H: investigation, methodology. M.M., T.G. investigation, methodology; C. S.: supervision; I.L.: conceptualization, supervision, writing, project administration and funding acquisition. All authors reviewed the results and contributed to the editing of the manuscript. All authors read and approved the manuscript.

Supplementary data

Supplementary data is available at *FEMSML Journal* online.

Conflict of interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

This work was financially supported by the Spanish Ministry of Science, Innovation and Universities grant PID2020-113494RB-I00/AEI/10.13039/501100011033. (Agencia Espanola de Investigación/Fondo Europeo de Desarrollo Regional, European Union) to I.L., and JSPS KAKENHI grant number JP23H02719 and the Institute for Fermentation, Osaka to H.H, respectively. P. I-S. was supported by a F.P.I. (PRE2021-097385) contract from the Spanish Ministry of Science, Innovation and Universities. A.S.M. was sup-

ported by a contract from the University Department of Navarra government (Res. 229E/2020 Microbiomics grant). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication. Open access funding provided by Universidad Publica de Navarra.

References

Adhya S, Gottesman M. Control of transcription termination. *Annu Rev Biochem* 1978;**47**:967–96.

Altenhoff AM, Train C-M, Gilbert KJ *et al.* OMA orthology in 2021: website overhaul, conserved isoforms, ancestral gene order and more. *Nucleic Acids Res* 2020;**49**:gkaa1007.

Arnaud M, Chastanet A, Débarbouillé M. New vector for efficient allelic replacement in naturally nontransformable, low-GC-content, gram-positive bacteria. *Appl Environ Microb* 2004;**70**:6887–91.

Baba T, Bae T, Schneewind O *et al.* Genome sequence of *Staphylococcus aureus* strain Newman and comparative analysis of staphylococcal genomes: polymorphism and evolution of two major pathogenicity islands. *J Bacteriol* 2008;**190**:300–10.

Baba T, Takeuchi F, Kuroda M *et al.* Genome and virulence determinants of high virulence community-acquired MRSA. *Lancet* 2002;**359**:1819–27.

Ballouz S, Francis AR, Lan R *et al.* Conditions for the evolution of gene clusters in bacterial genomes. *Plos Comput Biol* 2010;**6**:e1000672.

Barrio-Hernandez I, Yeo J, Jänes J *et al.* Clustering predicted structures at the scale of the known protein universe. *Nature* 2023;**622**:1–9.

Bashiri G, Nigon LV, Jirgis ENM *et al.* Allosteric regulation of menaquinone (vitamin K2) biosynthesis in the human pathogen. *Mycobacter Tuber J Biol Chem* 2020;**295**:3759–70.

Bekker M, Alexeeva S, Laan W *et al.* The ArcBA two-component system of *Escherichia coli* is regulated by the redox state of both the ubiquinone and the menaquinone pool. *J Bacteriol* 2010;**192**:746–54.

Blattman SB, Jiang W, Oikonomou P *et al.* Prokaryotic single-cell RNA sequencing by in situ combinatorial indexing. *Nat Microbiol* 2020;**5**:1192–201.

Boersch M, Rudrawar S, Grant G *et al.* Menaquinone biosynthesis inhibition: a review of advancements toward a new antibiotic mechanism. *Rsc Adv* 2018;**8**:5099–105.

Byrne A, Beaudin AE, Olsen HE *et al.* Nanopore long-read RNAseq reveals widespread transcriptional variation among the surface receptors of individual B cells. *Nat Commun* 2017;**8**:16027.

Charpentier E, Anton AI, Barry P *et al.* Novel cassette-based shuttle vector system for Gram-positive bacteria. *Appl Environ Microb* 2004;**70**:6076–85.

Danecek P, Bonfield JK, Liddle J *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* 2021;**10**:giab008.

Das A. Control of transcription termination by RNA-binding proteins. *Annu Rev Biochem* 1993;**62**:893–930.

Depledge DP, Srinivas KP, Sadaoka T *et al.* Direct RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen. *Nat Commun* 2019;**10**:754.

Feklístov A, Sharon BD, Darst SA *et al.* Bacterial sigma factors: a historical, structural, and genomic perspective. *Annu Rev Microbiol* 2014;**68**:357–76.

Forquet R, Jiang X, Nasser W *et al.* Mapping the complex transcriptional landscape of the phytopathogenic bacterium *Dickeya dadantii*. *mBio* 2022;**13**:e00524–22.

Fülöp Á, Torma G, Moldován N *et al.* Integrative profiling of Epstein–Barr virus transcriptome using a multiplatform approach. *Virol J* 2022;**19**:7.

Garalde DR, Snell EA, Jachimowicz D *et al.* Highly parallel direct RNA sequencing on an array of nanopores. *Nat Methods* 2018;**15**:201–6.

Gilchrist CLM, Booth TJ, van Wersch B *et al.* cblaster: a remote search tool for rapid identification and visualization of homologous gene clusters. *Bioinform Adv* 2021;**1**:vbab016.

Gottesman S. Bacterial regulation: global regulatory networks. *Annu Rev Genet* 1984;**18**:415–41.

Hamamoto H, Urai M, Ishii K *et al.* Lysocin E is a new antibiotic that targets menaquinone in the bacterial membrane. *Nat Chem Biol* 2015;**11**:127–33.

Henkin TM. Control of transcription termination in prokaryotes. *Annu Rev Genet* 1996;**30**:35–57.

Hilden I, Krath BN, Hove-Jensen B. Tricistronic operon expression of the genes gcaD (tms), which encodes N-acetylglucosamine 1-phosphate uridyltransferase, prs, which encodes phosphoribosyl diphosphate synthetase, and ctc in vegetative cells of *Bacillus subtilis*. *J Bacteriol* 1995;**177**:7280–4.

Iobbi-Nivol C, Leimkühler S. Molybdenum enzymes, their maturation and molybdenum cofactor biosynthesis in *Escherichia coli*. *Biochim Biophys Acta Bioenerg* 2013;**1827**:1086–101.

Jacob F, Monod J. Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol* 1961;**3**:318–56.

Jain M, Abu-Shumays R, Olsen HE *et al.* Advances in nanopore direct RNA sequencing. *Nat Methods* 2022;**19**:1160–4.

Johnson GE, Lalanne J-B, Peters ML *et al.* Functionally uncoupled transcription–translation in *Bacillus subtilis*. *Nature* 2020;**585**:124–8.

Johnston JM, Bulloch EM. Advances in menaquinone biosynthesis: sublocalisation and allosteric regulation. *Curr Opin Struc Biol* 2020;**65**:33–41.

Komili S, Silver PA. Coupling and coordination in gene expression processes: a systems biology view. *Nat Rev Genet* 2008;**9**:38–48.

Krishnakumar R, Ruffing AM. OperonSEQer: a set of machine-learning algorithms with threshold voting for detection of operon pairs using short-read RNA-sequencing data. *Plos Comput Biol* 2022;**18**:e1009731.

Kurosu M, Begari E. Vitamin K2 in electron transport system: are enzymes involved in vitamin K2 biosynthesis promising drug targets?. *Molecules* 2010;**15**:1531–53.

Lasa I, Toledo-Arana A, Dobin A *et al.* Genome-wide antisense transcription drives mRNA processing in bacteria. *Proc Natl Acad Sci* 2011;**108**:20172–7.

Lasa I, Toledo-Arana A, Gingeras TR. An effort to make sense of antisense transcription in bacteria. *RNA Biol* 2012;**9**:1039–44.

Lawrence M, Huber W, Pagès H *et al.* Software for computing and annotating genomic ranges. *PLoS Comput Biol* 2013;**9**:e1003118.

Legland D, Arganda-Carreras I, Andrey P. MorphoLibJ: integrated library and plugins for mathematical morphology with ImageJ. *Bioinformatics* 2016;**32**:btw413.

Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018;**34**:3094–100.

Lobritz MA, Belenky P, Porter CBM *et al.* Antibiotic efficacy is linked to bacterial cellular respiration. *Proc Natl Acad Sci* 2015;**112**:8173–80.

Mäder U, Nicolas P, Depke M *et al.* *Staphylococcus aureus* transcriptome architecture: from laboratory to infection-mimicking conditions. *PLoS Genet* 2016;**12**:e1005962.

Mao F, Dam P, Chou J *et al.* DOOR: a database for prokaryotic operons. *Nucleic Acids Res* 2009;**37**:D459–63.

Mao X, Ma Q, Zhou C *et al.* DOOR 2.0: presenting operons and their functions through dynamic and integrated views. *Nucleic Acids Res* 2014;**42**:D654–9.

Marcet-Houben M, Collado-Cala I, Fuentes-Palacios D *et al.* EvolClustDB: exploring eukaryotic gene clusters with evolutionarily conserved genomic neighbourhoods. *J Mol Biol* 2023;**435**:168013.

Marcet-Houben M, Gabaldón T. EvolClust: automated inference of evolutionary conserved gene clusters in eukaryotes. *Bioinformatics* 2019;**36**:1265–6.

Mashruwala AA, Guchte Avd, Boyd JM. Impaired respiration elicits SrrAB- dependent programmed cell lysis and biofilm formation in *Staphylococcus aureus*. *eLife* 2017;**6**:1–29.

Mihelčić M, Šmuc T, Supek F. Patterns of diverse gene functions in genomic neighborhoods predict gene function and phenotype. *Sci Rep* 2019;**9**:19537.

Nützmann H-W, Scazzocchio C, Osbourn A. Metabolic gene clusters in eukaryotes. *Annu Rev Genet* 2018;**52**:1–25.

Prazsák I, Moldován N, Balázs Z *et al.* Long-read sequencing uncovers a complex transcriptome topology in varicella zoster virus. *BMC Genomics* 2018;**19**:873.

Pust M-M, Davenport CF, Wiehlmann L *et al.* Direct RNA nanopore sequencing of *Pseudomonas aeruginosa* clone C transcriptomes. *J Bacteriol* 2021;**204**:e00418–21.

Richardson JP. Preventing the synthesis of unused transcripts by rho factor. *Cell* 1991;**64**:1047–9.

Rogozin IB, Makarova KS, Murvai J *et al.* Connected gene neighborhoods in prokaryotic genomes. *Nucleic Acids Res* 2002;**30**:2212–23.

Rohmer C, Dobritz R, Tuncbilek-Dere D *et al.* Influence of *Staphylococcus aureus* strain background on Sa3int phage life cycle switches. *Viruses* 2022;**14**:2471.

Rostöl J, Quiles-Purchalt N, Iturbe-Sanz P *et al.* Bacteriophages avoid autoimmunity from cognate immune systems as an intrinsic part of their life cycles. *Nat Microbiol* 2024;38565896. https://doi.org/10.1038/s41564-024-01661-6

Sáenz-Lahoya S, Bitarte N, García B *et al.* Noncontiguous operon is a genetic organization for coordinating bacterial gene expression. *Proc Natl Acad Sci* 2019;**116**:201812746.

Schenk S, Laddaga R. Improved method for electroporation of *Staphylococcus aureus*. *FEMS Microbiol Lett* 1992;**73**:133–8.

Schindelin J, Arganda-Carreras I, Frise E *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* 2012;**9**:676–82.

Sesto N, Wurtzel O, Archambaud C *et al.* The excludon: a new concept in bacterial antisense RNA-mediated gene regulation. *Nat Rev Micro* 2013;**11**:75–82.

Shi J, Yan S, Li W *et al.* PacBio full-length transcriptome analysis provides new insights into transcription of chloroplast genomes. *RNA Biol* 2023;**20**:248–56.

Slater GSC, Birney E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinform* 2005;**6**:31.

Soneson C, Yao Y, Bratus-Neuenschwander A *et al.* A comprehensive examination of Nanopore native RNA sequencing for characterization of complex transcriptomes. *Nat Commun* 2019;**10**:3359.

Szklarczyk D, Kirsch R, Koutrouli M *et al.* The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res* 2022;**51**:D638–46.

Taboada B, Ciria R, Martinez-Guerrero CE *et al.* ProOpDB: prokaryotic operon DataBase. *Nucleic Acids Res* 2012;**40**:D627–31.

Taboada B, Estrada K, Ciria R *et al.* Operon-mapper: a web server for precise operon identification in bacterial and archaeal genomes. *Bioinformatics* 2018;**34**:4118–20.

Tanizawa Y, Fujisawa T, Nakamura Y. DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication. *Bioinformatics* 2018;**34**:1037–9.

Toledo-Arana A, Dussurget O, Nikitas G *et al.* The *Listeria* transcriptional landscape from saprophytism to virulence. *Nature* 2009;**459**:950–6.

Toledo-Arana A, Lasa I. Advances in bacterial transcriptome understanding: from overlapping transcription to the excludon concept. *Mol Microbiol* 2020;**113**:593–602.

Torma G, Tombácz D, Moldován N *et al.* Dual isoform sequencing reveals complex transcriptomic and epitranscriptomic landscapes of a prototype baculovirus. *Sci Rep* 2022;**12**:1291.

Ubeda C, Olivarez NP, Barry P *et al.* Specificity of staphylococcal phage and SaPI DNA packaging as revealed by integrase and terminase mutations. *Mol Microbiol* 2009;**72**:98–108.

Valle J, Toledo-Arana A, Berasain C *et al.* SarA and not sigmaB is essential for biofilm development by *Staphylococcus aureus*. *Mol Microbiol* 2003;**48**:1075–87.

Valle J, Vergara-Irigaray M, Merino N *et al.* σB regulates IS256-mediated *Staphylococcus aureus* biofilm phenotypic variation. *J Bacteriol* 2007;**189**:2886–96.

Wurtzel O, Sesto N, Mellin JR *et al.* Comparative transcriptomics of pathogenic and non-pathogenic *Listeria* species. *Mol Syst Biol* 2012;**8**:583.

Xu R, Prakoso D, Salvador LCM *et al.* *Leptospira* transcriptome sequencing using long-read technology reveals unannotated transcripts and potential polyadenylation of RNA molecules. *Microbiol Spectr* 2023:e02234–23.

Received 11 November 2023; **revised** 20 March 2024; **accepted** 8 April 2024