



Global Annotation, Expression Analysis, and Stability of Candidate sRNAs in Group B *Streptococcus*

 Rebecca A. Keogh,^a  Brady L. Spencer,^a Hailee M. Sorensen,^b Rachel L. Zapf,^b Paul Briaud,^b Abigail E. Bonsall,^b  Kelly S. Doran,^a  Ronan K. Carroll^b

^aDepartment of Immunology and Microbiology, University of Colorado Anschutz, Aurora, Colorado, USA

^bDepartment of Biological Sciences, Ohio University, Athens, Ohio, USA

Rebecca A. Keogh and Brady L. Spencer contributed equally to this work. Order of names was decided by rock, paper, scissors. Kelly S. Doran and Ronan K. Carroll contributed equally to this work.

ABSTRACT Small, noncoding RNAs (sRNAs) are being increasingly identified as important regulatory molecules in prokaryotes. Due to the prevalence of next-generation sequencing-based techniques, such as RNA sequencing (RNA-seq), there is potential for increased discovery of sRNAs within bacterial genomes; however, these elements are rarely included in annotation files. Consequently, expression values for sRNAs are omitted from most transcriptomic analyses, and mechanistic studies have lagged behind those of protein regulators in numerous bacteria. Two previous studies have identified sRNAs in the human pathogen group B *Streptococcus* (GBS). Here, we utilize the data from these studies to create updated genome annotation files for the model GBS strains NEM316 and COH1. Using the updated COH1 annotation file, we reanalyze publicly available GBS RNA-seq whole-transcriptome data from GenBank to monitor GBS sRNA expression under a variety of conditions and genetic backgrounds. This analysis generated expression values for 232 putative sRNAs that were overlooked in previous transcriptomic analyses in 21 unique comparisons. To demonstrate the utility of these data, we identify an sRNA that is upregulated during vaginal colonization and demonstrate that overexpression of this sRNA leads to increased bacterial invasion into host epithelial cells. Finally, to monitor RNA degradation, we perform a transcript stability assay to identify highly stable sRNAs and compare stability profiles of sRNA- and protein-coding genes. Collectively, these data provide a wealth of transcriptomic data for putative sRNAs in GBS and a platform for future mechanistic studies.

IMPORTANCE In recent years, sRNAs have emerged as potent regulatory molecules in bacteria, including numerous streptococcal species, and contribute to diverse processes, including stress response, metabolism, housekeeping, and virulence regulation. Improvements in sequencing technologies and *in silico* analyses have facilitated identification of these regulatory molecules as well as improved attempts to determine the location of sRNA genes on the genome. However, despite these advancements, sRNAs are rarely included in genome annotation files. Consequently, these molecules are often omitted from transcriptomic data analyses and are commonly repeat identified across multiple studies. Updating current genomes to include sRNA genes is therefore critical for better understanding bacterial regulation.

KEYWORDS RNA stability, *Streptococcus agalactiae*, group B streptococcus, regulatory RNA, sRNA, virulence

Streptococcus agalactiae, or group B *Streptococcus* (GBS), is a Gram-positive bacterium that colonizes the urogenital and gastrointestinal tract of 20% to 30% of healthy individuals (1, 2). However, as an opportunistic pathogen, GBS also causes

Editor Nancy E. Freitag, University of Illinois at Chicago

Copyright © 2021 Keogh et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Ronan K. Carroll, carrolr3@ohio.edu.

Received 22 September 2021

Accepted 29 September 2021

Published 2 November 2021

diverse diseases, including meningitis, sepsis, skin and soft tissue infections, and pneumonia (3, 4). Studies on GBS gene regulation have focused on protein regulators (such as two-component systems), which sense environment signals and respond by regulating bacterial gene expression (5–7). However, nonprotein regulators, such as small, noncoding RNAs (sRNAs), have been understudied in GBS thus far.

Two previous studies identified putative sRNAs in GBS strain NEM316 (8, 9). Pichon et al. utilized *in silico* analysis to identify 197 novel sRNA candidates (8), and subsequent work by Rosinski-Chupin et al. utilized differential RNA sequencing (dRNA-seq) and strand-specific RNA sequencing (RNA-seq) to identify 120 putative sRNAs (9). These sRNAs were later classified by Wolf et al. into conserved RNA families across 27 GBS genomes (10). However, because only those with high structural similarity to known bacterial sRNAs were analyzed, only 30 GBS sRNAs were classified in this study, potentially excluding many legitimate GBS sRNAs. Here, we utilize these previous studies to update the genomes of the clinically relevant strains NEM316 and COH1 to include sRNA annotations and utilize these updated genomes to generate both sRNA expression and transcript stability data.

ANNOTATION OF sRNAs ON THE GBS GENOME

To facilitate the study of sRNAs in GBS, we updated two genomes from clinically relevant GBS strains (NEM316 and COH1) to include annotations for all putative sRNA candidates identified by Pichon et al. and Rosinski-Chupin et al. Supplemental files indicating the location of the sRNAs from each study were combined, and all repeat-identified sRNAs were eliminated. The remaining sRNAs were then annotated on the GBS genomes, resulting in the addition of 272 sRNAs to the NEM316 genome and 232 to the COH1 genome (Fig. 1A; see File S1 in the supplemental material, genome annotation files available at https://figshare.com/projects/Global_annotation_expression_analysis_and_stability_of_sRNAs_in_Group_B_Streptococcus/117768). A total of 40 sRNAs were not added to the COH1 genome due to reduced sequence homology (<80%) or absence from the genome. sRNA-encoding genes were annotated sequentially starting at the origin of replication as was performed for GBS coding sequence annotations (see Text S1).

GLOBAL sRNA EXPRESSION ANALYSIS

All previously published GBS RNA-seq studies have overlooked sRNA gene reads. To recover and analyze these overlooked data, we downloaded publicly available GBS RNA-seq data sets and generated expression data for 232 sRNAs in the COH1 background (see Text S1 for selection criteria used). A total of 70 RNA-seq data sets fulfilled the criteria for reanalysis (see File S2 in the supplemental material) (7, 11–15). Using these studies, we performed 21 unique differential gene expression analyses (DEAs) in which sRNA expression was analyzed across two different conditions (see Table S1 in the supplemental material; Fig. 1B). The expression values of all sRNAs in each DEA are included as File S3 in the supplemental material, and the number of differentially expressed (>3-fold with a minimum expression of 10 in at least 1 condition) sRNAs in each comparison is shown in Fig. 1B and File S4 in the supplemental material.

Of these 21 comparisons, we first examined DEA 9 (A909 chemically defined medium [CDM] versus A909 vaginal tract) to evaluate sRNAs that have altered expression *in vivo* (compared with laboratory conditions), as these sRNAs may affect GBS host persistence. A total of 85 sRNAs demonstrated >3-fold variation in expression (which met our cutoff criteria), with 28 being downregulated and 57 being upregulated in the vaginal tract (Fig. 1B; File S4). Of those 57 sRNAs, s1160 was the most highly upregulated in the vagina (88-fold) (Fig. 1C) and exhibited the second highest expression in the GBS cell (average reads per kilobase per million [RPKM] expression level of >32,000 across all conditions examined) (File S3). Since sRNA function is often related directly to abundance, we hypothesized that s1160 may play a role during GBS vaginal colonization. To examine this hypothesis, we overexpressed s1160 in GBS and assessed its

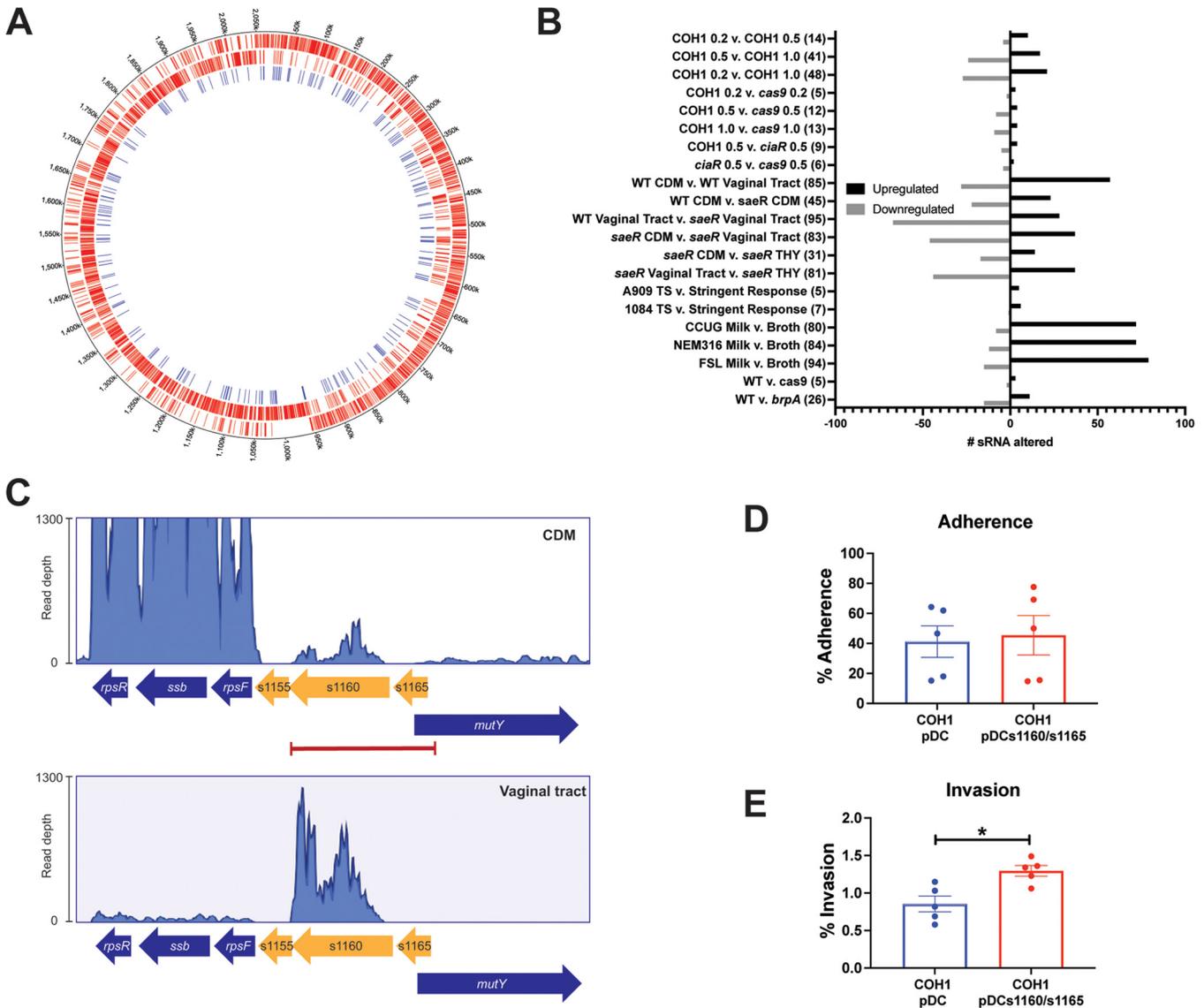


FIG 1 Global sRNA identification and expression analysis. (A) Diagram showing the distribution of sRNA genes (blue lines, inner circle) throughout the GBS genome. Coding sequence genes on the forward strand (outer circle) and reverse strand (middle circle) are depicted in red. Circos plot was created with Circa (<http://omgenomics.com/circa>). (B) Results of 21 DEAs performed. The conditions for each DEA are indicated as well as the number of sRNA genes upregulated and downregulated (>3-fold). The total number of sRNA genes altered is shown in parentheses. (C). The s1160/s1165 locus and transcript read mapping in GBS strain COH1. Coding sequence genes are illustrated as dark-blue arrows and sRNA genes as yellow arrows. RNA-seq read mappings from DEA9 (CDM top, vaginal tract bottom) are displayed for the locus. Red bar indicates the region cloned into pDC to create pDCs1160/s1165. (D and E) GBS strain COH1 with an empty vector (pDC) compared with an s1160 overexpression strain (pDCs1160/s1165) were used to test adherence to (D) or invasion of (E) hVECs. All data are represented as percent CFU recovered of the initial inoculum and were performed at least in technical triplicates. Each plot is representative of five individual experiments. Significance determined by paired *t* test; *, *P* < 0.05.

interaction (adherence and invasion) with human vaginal epithelial cells (hVECs). The s1160 overexpression construct included the upstream sequence as to include its putative promoter. Although an adjacent sRNA, s1165, was also included in this upstream region, s1165 had an expression value of 0 in the vaginal tract and little to no expression across the other DEAs; therefore, it is unlikely to influence any vaginal colonization phenotypes. s1160 overexpression did not impact GBS adherence to hVECs but did significantly increase GBS invasion compared with the vector control (Fig. 1D and E). These results demonstrate the utility of the analysis presented here and suggest a potential role for s1160 during GBS vaginal colonization. Future studies will examine the contribution of s1160 to GBS colonization *in vivo*, as well as investigate the mechanism of action of this sRNA. Quantitative PCR (qPCR) analysis of the s1160 and s1165

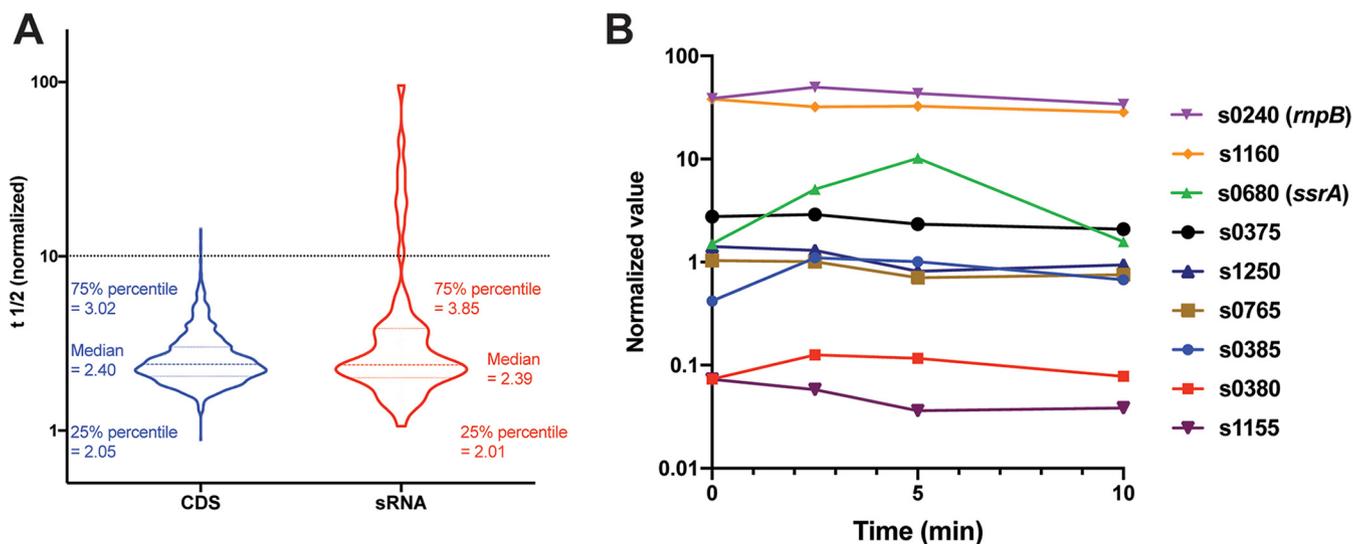


FIG 2 RNA stability analysis. (A) Violin plot of normalized RNA stability data showing (i) the range and (ii) the median half-life of GBS sRNA and coding sequence (CDS) RNA. The 75% and 25% range are also indicated. The dashed line indicates a half-life of 10 min. (B) Normalized stability profiles of select sRNAs over time. Normalized stability plots over time (from 0 to 10 min) are shown for all sRNAs with a half-life exceeding 10 min, including s0385 which had a half-life greater than that of *ssrA*. Plots show abundance relative to *ssrA*.

genes in COH1 at an optical density at 600 nm (OD_{600}) of 0.2, 0.5, and 1.0 (corresponding to the time points used in DEAs 1, 2, and 3) confirmed the expression patterns observed in the RNA-seq data analysis (see Fig. S1 in the supplemental material).

GLOBAL RNA STABILITY ANALYSIS

Cellular RNA levels are regulated by the control of RNA synthesis (i.e., transcription) but also by the rate of RNA stability/degradation. The combined application of rifampicin RNA stability experiments with RNA-seq (here referred to as stability RNA-seq) allows a global analysis of the stability/rate of degradation of all cellular transcripts simultaneously (16). We performed stability RNA-seq using our updated GBS genome reference file and analyzed the stability of both coding DNA sequence (CDS) and sRNA transcripts. Gene expression values were normalized against the value for *ssrA* at the corresponding time point, and the half-life of each RNA was calculated (see Text S1 for details and cut off criteria applied). Half-lives were determined for 1,759 CDS transcripts and 72 sRNAs (see File S5 in the supplemental material). While the median normalized half-life for CDS and sRNA transcripts was very similar (2.4 min and 2.39 min, respectively), the mean half-life for sRNA transcripts was much larger (6.3 min compared with 2.7 min for CDS) (Fig. 2A). Only five CDS transcripts had half-lives exceeding 10 min, with the maximum half-life of 14.4 min for GBSCOH1_RS09545. In contrast, eight sRNA transcripts had half-lives longer than 10 min (Fig. 2B), including values of 45.7 min for *ssrA*, 32.8 min for *rnpB*, and a maximum value of 95.9 min for the uncharacterized sRNA s0380. Of note, s1160 (shown above as one of the most highly expressed sRNAs in GBS and being upregulated during vaginal colonization) was also among the most stable transcripts in the cell with a half-life value of 26.75 min. This result was confirmed by Northern blotting (see Fig. S2 in the supplemental material).

The half-life analysis returned negative values for three CDS transcripts and five sRNAs. In most cases, this result was because one (or more) of the normalized expression values for that transcript was zero at one (or more) time points. However, one CDS and one sRNA transcript, namely, GBSCOH1_RS10860 and GBSCOH1_s0385, generated negative values because their normalized expression values increased over time, indicating that the abundance of these transcripts increased relative to *ssrA*; therefore, they are more stable than *ssrA*. Interestingly, s0385 (which exhibited the highest sRNA half-life) (Fig. 2B) is encoded adjacent but antisense to another highly stable sRNA,

s0380. These two sRNAs are divergently transcribed, and the locus contains a gene potentially encoding a holin-like protein (see Fig. S3 in the supplemental material). The high degree of stability of antisense transcripts, of which one putatively encodes a toxin, is reminiscent of a toxin-antitoxin system and would be of interest to study in the future.

CONCLUSIONS

The data presented here demonstrate the utility of the updated genome annotation files, allowing us to determine the stability of 232 sRNAs in COH1 as well as their expression in 21 unique comparisons. These results will inform future phenotypic studies and likely identify new sRNA regulators in GBS. Undoubtedly, many sRNAs have yet to be discovered in GBS. Our newly annotated genomes will be an invaluable tool for the further identification of sRNAs, and these updated annotation files will prevent repeat sRNA identification in future studies. Collectively, the data generated highlight the importance of updating genome annotation files as new regulatory elements are identified in bacteria.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

FILE S1, XLSX file, 0.03 MB.

FILE S2, XLSX file, 0.03 MB.

FILE S3, XLSX file, 0.1 MB.

FILE S4, XLSX file, 0.1 MB.

FILE S5, XLSX file, 0.2 MB.

TEXT S1, DOCX file, 0.03 MB.

FIG S1, TIF file, 1 MB.

FIG S2, TIF file, 1.8 MB.

FIG S3, TIF file, 0.2 MB.

TABLE S1, DOCX file, 0.02 MB.

ACKNOWLEDGMENTS

Funding for this work was provided by NIHT32DK120521 to R.A.K., NIH/NIAID F32 AI143203 to B.L.S., NIH/NINDS R01 NS116716 and NIH/NIAID R01 AI153332 to K.S.D., and AI143743 to R.K.C.

REFERENCES

- Wilkinson HW. 1978. Group B streptococcal infection in humans. *Annu Rev Microbiol* 32:41–57. <https://doi.org/10.1146/annurev.mi.32.100178.000353>.
- Regan JA, Klebanoff MA, Nugent RP. 1991. The epidemiology of group B streptococcal colonization in pregnancy. *Vaginal Infections and Prematurity Study Group. Obstet Gynecol* 77:604–610.
- Schuchat A. 1998. Epidemiology of group B streptococcal disease in the United States: shifting paradigms. *Clin Microbiol Rev* 11:497–513. <https://doi.org/10.1128/CMR.11.3.497>.
- Doran KS, Nizet V. 2004. Molecular pathogenesis of neonatal group B streptococcal infection: no longer in its infancy. *Mol Microbiol* 54:23–31. <https://doi.org/10.1111/j.1365-2958.2004.04266.x>.
- Patras KA, Wang NY, Fletcher EM, Cavaco CK, Jimenez A, Garg M, Fierer J, Sheen TR, Rajagopal L, Doran KS. 2013. Group B Streptococcus CovR regulation modulates host immune signalling pathways to promote vaginal colonization. *Cell Microbiol* 15:1154–1167. <https://doi.org/10.1111/cmi.12105>.
- Quach D, Van Sorge NM, Kristian SA, Bryan JD, Shelver DW, Doran KS. 2009. The ciar response regulator in group b streptococcus promotes intracellular survival and resistance to innate immune defenses. *J Bacteriol* 191:2023–2032. <https://doi.org/10.1128/JB.01216-08>.
- Cook LCC, Hu H, Maienschein-Cline M, Federle J. 2018. A vaginal tract signal detected by the group B Streptococcus SaeRS system elicits transcriptional changes and enhances. *Infect Immun* 86:e00762-17. <https://doi.org/10.1128/IAI.00762-17>.
- Pichon C, Du Merle L, Caliot ME, Trieu-Cuot P, Le Bouguéne C. 2012. An in silico model for identification of small RNAs in whole bacterial genomes: characterization of antisense RNAs in pathogenic Escherichia coli and Streptococcus agalactiae strains. *Nucleic Acids Res* 40:2846–2861. <https://doi.org/10.1093/nar/gkr1141>.
- Rosinski-Chupin I, Sauvage E, Sismeiro O, Villain A, Da Cunha V, Caliot ME, Dillies MA, Trieu-Cuot P, Bouloc P, Lartigue MF, Glaser P. 2015. Single nucleotide resolution RNA-seq uncovers new regulatory mechanisms in the opportunistic pathogen Streptococcus agalactiae. *BMC Genomics* 16:419. <https://doi.org/10.1186/s12864-015-1583-4>.
- Wolf IR, Paschoal AR, Quiroga C, Domingues DS, de Souza RF, Pretto-Giordano LG, Vilas-Boas LA. 2018. Functional annotation and distribution overview of RNA families in 27 Streptococcus agalactiae genomes. *BMC Genomics* 19:556. <https://doi.org/10.1186/s12864-018-4951-z>.
- Spencer BL, Deng L, Patras KA, Burcham ZM, Sanches GF, Nagao PE, Doran KS. 2019. Cas9 contributes to group B streptococcal colonization and disease. *Front Microbiol* 10:1930. <https://doi.org/10.3389/fmicb.2019.01930>.
- Hooven TA, Catomeris AJ, Bonakdar M, Tallon LJ, Santana-Cruz I, Ott S, Daugherty SC, Tettelin H, Ratner J. 2018. The Streptococcus agalactiae stringent response enhances virulence and persistence in human blood. *Infect Immun* 86:e00612-17. <https://doi.org/10.1128/IAI.00612-17>.
- Richards VP, Choi SC, Bitar PDP, Gurjar AA, Stanhope MJ. 2013. Transcriptional and genomic evidence for Streptococcus agalactiae adaptation to the bovine environment. *BMC Genomics* 14:920. <https://doi.org/10.1186/1471-2164-14-920>.

14. Ma K, Cao Q, Luo S, Wang Z, Liu G, Lu C, Liu Y. 2018. cas9 enhances bacterial virulence by repressing the regR transcriptional regulator in *Streptococcus agalactiae*. *Infect Immun* 86:e00552-17. <https://doi.org/10.1128/IAI.00552-17>.
15. Patras KA, Derieux J, Al-Bassam MM, Adiletta N, Vrbanac A, Lapek JD, Zengler K, Gonzalez DJ, Nizet V. 2018. Group B streptococcus biofilm regulatory protein A contributes to bacterial physiology and innate immune resistance. *J Infect Dis* 218:1641–1652. <https://doi.org/10.1093/infdis/jiy341>.
16. Sorensen HM, Keogh RA, Wittekind MA, Caillet AR, Wiemels RE, Laner EA, Carroll RK. 2020. Reading between the lines: using RNA-seq data for global analysis of sRNAs in *Staphylococcus aureus*. *mSphere* 5:e00439-20. <https://doi.org/10.1128/mSphere.00439-20>.