



A Conserved *Histophilus somni* 23S Intervening Sequence Yields Functional, Fragmented 23S rRNA

 Gregory P. Harhay,^a Dayna M. Harhay,^a Kerry D. Brader,^a Timothy P. L. Smith^a

^aUSDA-ARS-U.S. Meat Animal Research Center, Clay Center, Nebraska, USA

ABSTRACT *Histophilus somni* is a Gram-negative bacterial organism that acts as an opportunistic pathogen and is a fastidious member of the *Pasteurellaceae* family associated with diseases of respiratory, reproductive, cardiac, and other tissues of ruminants. We identified an intervening sequence (IVS) embedded in all five copies of the 23S rRNA gene in the closed genome sequence of the *H. somni* isolate USDA-ARS-USMARC-63250 that may play an important role in affecting the biology of the organism. Sequencing the RNA from this isolate shows that most of the IVS is cleaved from the transcript, resulting in independent fragments of this structural rRNA that remain functional within the bacterial ribosome. The IVS lies between positions 1170 and 1278 bp of the 3,017-bp gene and exhibits self-complementarity between its 5' and 3' ends that predicts a stem-loop structure interrupting helix-45 in the transcribed 23S rRNA. Excision removes a 94-nucleotide (nt) stem-loop structure that displays an unusual 1-nt 3' end overhang instead of the more typical 2-nt overhang commonly observed at the ends of other excised IVS stem-loops. A comparison with genomes of other *H. somni* isolates indicates that this IVS is highly conserved, with 31 of 32 complete genomes having similar interruptions of canonical 23S rRNA genes. The potential biological effects of either the released IVS or the fragmentation of the functional 23S rRNA are unknown, but fragmentation may enhance rRNA degradation in ways that contribute to the regulation of gene expression.

IMPORTANCE The genome biology underlying *H. somni* virulence, pathogenicity, environmental adaptability, and broad tissue tropism is understood poorly. We identified a novel *H. somni* 109-nt IVS stem-loop structure, of which the central portion is excised from the 23S rRNA transcript, resulting in the fragmentation of this rRNA in the *H. somni* isolate USDA-ARS-USMARC-63250 and the release of a 94-nt structured RNA of unknown function. We determined that this peculiar rRNA biology is widespread among sequenced *H. somni* isolates, suggesting it has importance to organism biology. The fragmented 23S rRNA molecules remain functional in the ribosome, given that the isolate grows in culture. The structured excised portion of the IVS, presumably due to the action of the endoribonuclease III, has an unusual 1-nt 3' end overhang. This newly discovered *H. somni* 23S rRNA fragmentation may enhance rRNA degradation providing a previously unrecognized avenue for regulating *H. somni* biological processes.

KEYWORDS 23S RNA, DNA sequencing, *Pasteurellaceae*, RNA stability, RNA structure, cattle, genomics, Gram-negative bacteria, intervening sequence, mucosal pathogens, transcription, bovine respiratory disease

The assembly of complete, closed bacterial genomes based on long sequence reads is routine and supports novel studies of bacterial genome function based on high-depth sequencing of RNA transcripts under various culture conditions. Long-read transcript sequencing can produce reads that span, and thus define, bacterial operons and

Citation Harhay GP, Harhay DM, Brader KD, Smith TPL. 2021. A conserved *Histophilus somni* 23S intervening sequence yields functional, fragmented 23S rRNA. *Microbiol Spectr* 9: e01431-21. <https://doi.org/10.1128/Spectrum.01431-21>.

Editor Swaine L. Chen, The National University of Singapore and the Genome Institute of Singapore

This is a work of the U.S. Government and is not subject to copyright protection in the United States. Foreign copyrights may apply.

Address correspondence to Gregory P. Harhay, gregory.harhay@usda.gov.

Received 6 October 2021

Accepted 12 October 2021

Published 1 December 2021

may provide new insight into transcriptional regulation. We published closed *Histophilus somni* genome sequences (1) isolated from a 2013 field study to perform the transcriptomic studies reported here.

Attempts to sequence the mRNA from protein-coding genes were confounded by high abundance RNA mapping to the 23S gene, even after hybridization-based rRNA depletion using kits from two different manufacturers. Rosenow et al. reported (2) that rRNA constitutes 95% to 97% of total RNA in bacterial cells. An examination of reads mapping to each of the five 23S genes in *H. somni* USDA-ARS-USMARC-63250 (GenBank accession [CP018802](#)) revealed significantly reduced read coverage in a 1×10^2 -bp region at 1.2×10^3 bp downstream of the 5' end of each 23S gene suggesting a diminished abundance of rRNA mapping to this region. Therefore, in this study, we set out to understand the biochemical and genomic factors underlying this observation to mitigate this effect, and in the process, we discovered a phenomenon previously unrecognized to occur in *H. somni*.

Intervening sequences (IVSs) have been reported in 23S rRNA genes of multiple bacterial species in 12 families (3). IVSs in helix-45, as defined by Evguenieva-Hackenberg (4) for the predicted secondary structure of 23S rRNA in *Escherichia coli*, have been reported to occur at position 1170 of the gene in *Salmonella* spp. (5), *Proteus* and *Providencia* strains (6), and *Campylobacter jejuni* (7), as well as at position 1171 in *Haemophilus* spp. (8). IVSs have also been observed in helices 9 and 23 (3). The length of IVSs in helix-45 varies widely, from 37 to 240 bp within the *Campylobacter* genus (9) up to 759 bp within the spirochete genus *Leptospira* (10). Overall, IVSs are distributed sporadically among the bacteria (6) and can occur in both 16S and 23S rRNA genes (3). The fragmented 23S rRNA appears to be functional in the ribosome (4).

Here, we present comparative genomic, electrophoretic, and transcriptomic evidence consistent with a 109-nucleotide (nt) IVS stem-loop structure in *H. somni* USDA-ARS-63250 23S rRNA and confirm the viability of the strain in culture. The functional role or roles of the excised 23S rRNA, if any, or what remains of it after degradation in bacterial or infected host cells, have not been identified definitively. While the examination of the total RNA size distribution profile and transcriptome provides evidence for highly abundant 3' and 5' ends of the 23S gene, the 94-nt excised region is not present in high abundance proportional to the size fractions putatively associated with the ends of the 23S gene. Therefore, we infer that the 94-nt excised region is degraded or sequestered within or exported out of the cell. Finally, we conclude with a discussion of the potential effects of fragmented 23S rRNA on regulation.

RESULTS

The electrophoretic profile of total RNA extracted from an *H. somni* USDA-ARS-USMARC-63250 culture did not display a prominent peak corresponding to highly abundant 23S rRNA that is typical for bacterial RNA profiles (Fig. 1a). The profile lacked a prominent peak near the 3017-nt full-length 23S rRNA position but displayed a strong peak at approximately 1341 nt along with a peak at 1840 nt. After ribodepletion (Fig. 1b) using the default rRNA capture oligonucleotide mix (see "RNA isolation and ribodepletion combination 1 [RNA I&D 1]" in the Materials and Methods), the peak centered at 1840 nt and lower mass shoulder was greatly diminished; however, the peak at 1350 nt persisted. Two additional custom-designed capture oligonucleotides targeting the 5' end of the 23S gene were added to the capture mixture for ribodepletion resulting in a significant reduction in the abundance of molecules at 1350 nt, suggesting that this peak is due to a 5'-end fragment of the 23S gene. Given that the capillary electrophoresis instrument vendor reports the sizing accuracy of the assay used to be $\pm 20\%$, it was not possible to confidently associate particular RNA sequences (16S, 23S, or their fragments) with the observed peaks using these results alone. Multiple attempts to extract RNA from *H. somni* USDA-ARS-USMARC-63250 cultured in brain heart infusion 0.1% Trizma base and 0.01% thiamine monophosphate (BHI-TT) never yielded a capillary electrophoretic separation pattern with a prominent 23S rRNA peak

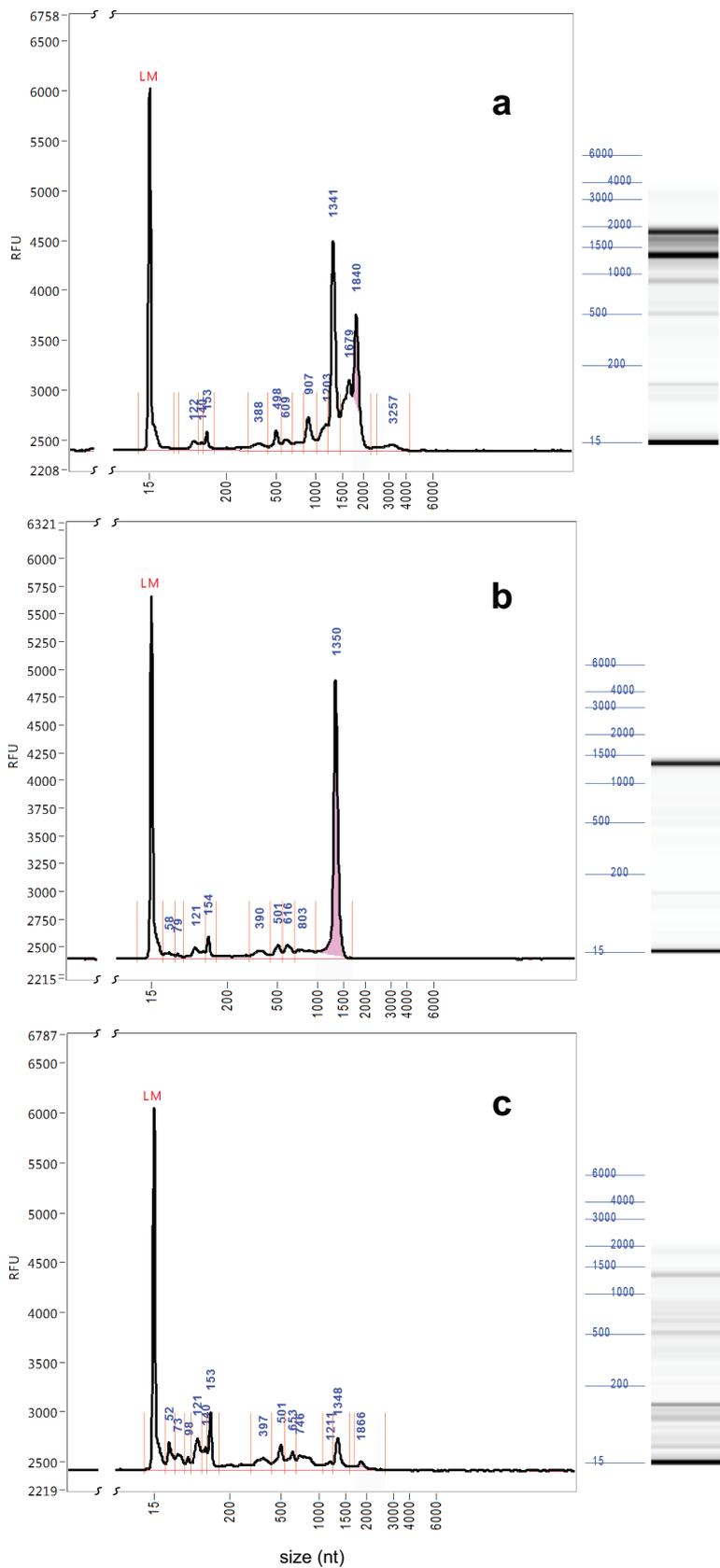


FIG 1 Analysis of the size distributions of *H. somni* RNA molecules as produced with the RNA isolation and ribodepletion combination 1 method. (a) Capillary electrophoretic separation pattern of total RNA
(Continued on next page)

at or around 3017 nt. However, the low abundance population of molecules (relative to the other populations) at 3257 nt in total RNA (Fig. 1a) is eliminated with ribodepletion in Fig. 1c. This elimination suggests that this population may be associated with intact 23S although at an atypically low level. Using different rRNA extraction and ribodepletion procedures (see “RNA isolation and ribodepletion combination 2 [RNA I&D 2]” in the Materials and Methods) yielded results (see Fig. S1 in the supplemental material) like those in Fig. 1; there was no significant peak associated with intact 23S rRNA at or near 3017 nt in total RNA, suggesting the absence of a peak near 3017 nt is not an experimental artifact. The reasons for these observations were explored using a comparative genomics approach.

Identification of the *H. somni* IVS. The sequences of the five copies of 23S rRNA genes present in the genome of *H. somni* isolate USDA-ARS-USMARC-63250 were first examined by aligning them to each other, which established that they are 100% identical. The representative sequence used subsequently for an interisolate and interspecies comparison was taken from locus_tag BTV18_00080 located at 14280 to 17296 bp (designated as CP018802-23S-1). A discontinuous megablast (E value, 1×10^{-100}) comparison was performed with this sequence against all RNA sequences from bacteria belonging to the *Gammaproteobacteria* class in the NCBI RefSeq RNA database to identify sequence differences. The top 40 out of 452 alignments with the highest identity are shown in descending order of similarity in Fig. 2a. The two subject sequences immediately below the self-alignment with the highest pairwise identity (>99%) with the query sequence are sequences from *H. somni* isolates 2336 (NR_103965) and 129PT (NR_076444). These alignments show a typically low level of intraspecies variation in the 23S rRNA gene sequence (dark black vertical lines in Fig. 2a). In contrast, the alignments to 23S rRNA genes belonging to other species suggest the presence of an IVS within *H. somni* 23S genes, whose position is revealed by the interruption of the blue coverage rectangle at the top of Fig. 2a.

The non-*H. somni* 23S rRNA genes with the highest similarity outside the IVSs include sequences from other *Pasteurellaceae* and a few *Salmonella* genes. The pairwise identity of *Pasteurellaceae* 23S rRNA genes to *H. somni* is lower in the ~1170-bp portion of the gene upstream of the insertion than in the ~1730-bp downstream portion. Species with lower pairwise identity include members of the *Enterobacteriaceae* family, such as *Raoultella ornithinolytica*, *Shimwellia blattae*, *Enterobacter asburiae*, and *Salmonella enterica* subsp. *enterica* serovar Paratyphi. Among the 37 non-*H. somni* alignments shown, only *Mannheimia succiniciproducens* (NR_076083) includes a sequence spanning the *H. somni* IVS. The alignment of this 23S rRNA gene across the IVS exhibits 39 differences relative to the query sequence and reveals an additional 22-bp insertion within *M. succiniciproducens* (Fig. 2b); consequently, we do not consider NR_076083 to be sufficiently similar to the IVS to be considered a biologically relevant match. We note that if megablast results are sorted in descending order by Geneious (4) grade score instead of overall pairwise identity, additional alignments that span the IVS region were identified. However, these alignments have even lower similarities to the *H. somni* IVS and are not considered biologically relevant matches to the IVS within the *H. somni* query sequence (Fig. S2 and S3).

The 3' end of the gap is well defined by the alignment of the subject sequence beginning at 1279 bp (Fig. 2b). The 5' end of the gap is less well defined, but we estimate

FIG 1 Legend (Continued)

from *H. somni* USDA-ARS-USMARC-63250 grown in BHI-TT. The electropherogram is on the left, and its corresponding digital gel image is on the right. LM designates the lane marker standard at 15 nt; these runs were calibrated with known standards according to vendors directions. A prominent 23S band in the vicinity of 3.1 kb was not observed; instead, 1.34- and 1.84-kb peaks were observed, consistent with 23S rRNA fragmentation. (b) After ribodepletion using the standard oligonucleotide capture mix showing a prominent peak at 1350 nt. (c) Ribodepletion using the standard oligonucleotide capture mix + two additional custom capture oligonucleotides targeting the 5' end of the 23S rRNA greatly reduces the abundance of molecules giving rise to the 1350-nt peak. A large region with no detectable signal to the left of 15-nt lane marker was clipped out of each electropherogram in a, b, and c as indicated by breaks in the axes.

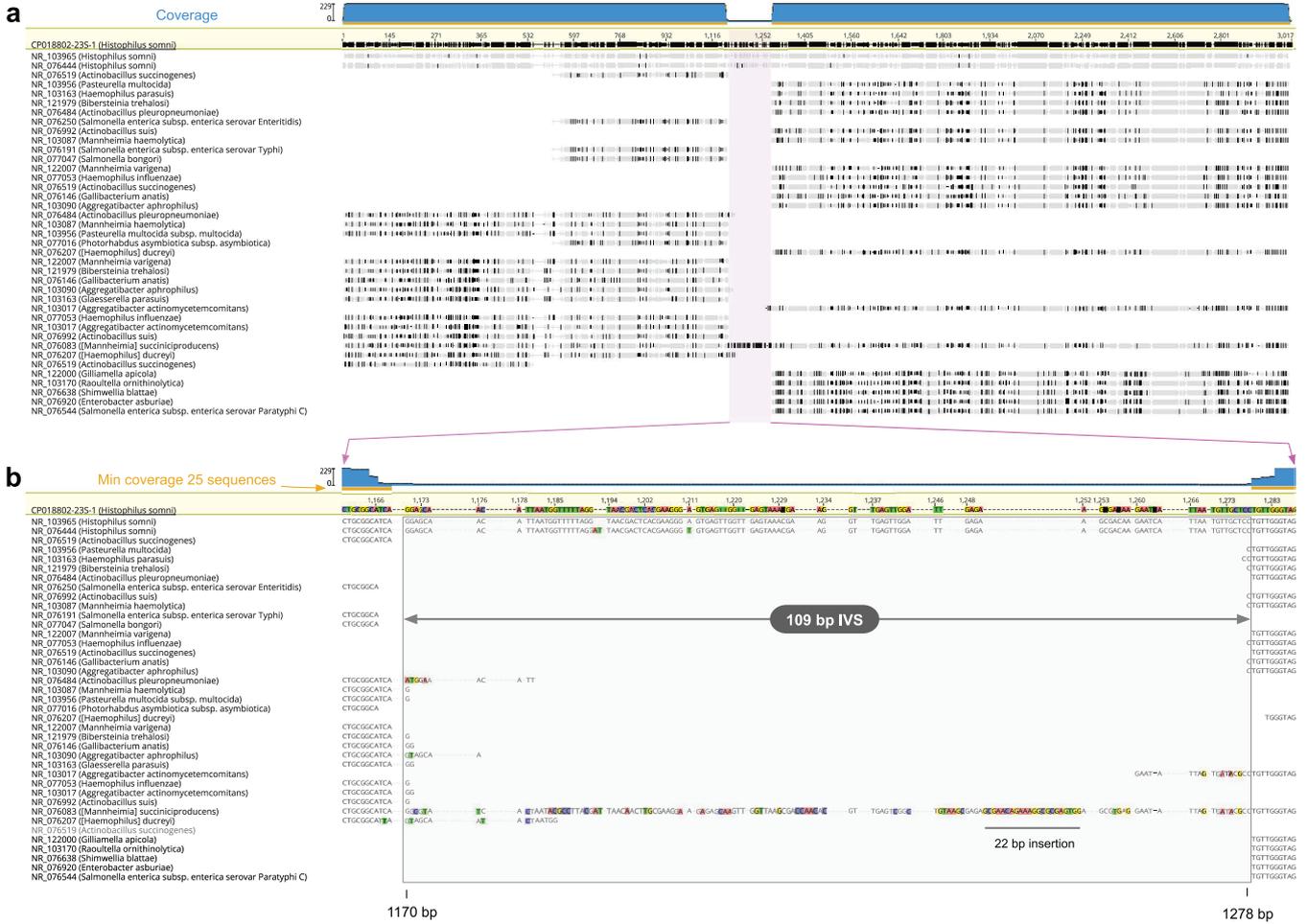


FIG 2 Identifying the putative 23S intervening sequence. (a) Discontiguous megablast results using USDA-ARS-USMARC-63250 CP018802-23S-1 query sequence against all *Gammaproteobacteria* (taxid 1236) in the RefSeq RNA database. The results are sorted by best pairwise % identity with query in descending order, showing only the top 39 subject sequences out of 452 total. We note that the 3' ends of the *Pasteurellaceae* members generally have the highest pairwise % identity with the query sequence, followed by the 5' ends of the same sequences. The blue coverage plot at the top reveals even coverage of the query sequence by the subject sequences, except for a gap near the center of the query sequence, that is, a gap between the 5' and 3' ends of the subject sequences. The gold bar below the blue coverage plot defines the region covered by at least 25 matching subject sequences. Grayed blocks or letters in the subject sequence indicate matches to the query sequence, e.g., NR_103965 and NR_076444 *H. somni* 23S sequences. Dark-black vertical lines indicate discrepancies, e.g., NR_076083 *Mannheimia succiniciproducens*, within the gap region. (b) The central gap region as demarcated by the light-magenta box is expanded to single-base resolution. The 109-bp IVS region is defined to fall 1170 to 1279 bp upstream from the 5' end. Bases in the subject sequences that are discordant from the query sequence are colored, while those that are the same are gray. The single NR_076083 alignment spanning the IVS is not sufficiently concordant with the IVS to be considered a biologically relevant match.

that the 5' end resides somewhere between 1166 and 1173 bp with a midpoint at 1170 bp. Given these boundaries, we define a putative 109-nt IVS sequence spanning between 1170 and 1278 bp within the 23S gene. A blastn similarity search (default parameters with an E value threshold of 1×10^{-20}) was performed using the CP018802-23S-1 109-nt IVS query sequence against the entire RefSeq RNA database, which only returned the two previously mentioned *H. somni* subject sequences, namely, NR_103965 (*H. somni* isolate 2336) and NR_076444 (*H. somni* isolate 129PT). Evidently, this putative IVS sequence is unique to *H. somni* within the RefSeq RNA database.

Defining *H. somni* IVS variation within closed genomes. A blastn similarity search (E value, 1×10^{-5}) was conducted with the 109-nt putative IVS as the query sequence against all 32 *H. somni* closed genomes as the subject database. This search returned 160 subject sequences that represented 7 unique sequences. The MAFFT alignment of these seven unique sequences is shown in Fig. 3a.

The query IVS sequence is 100% identical to the isolate 2336 sequence. Except for ASc-MMNZ-VFA-073, each isolate yielded multiple identical matches to the query

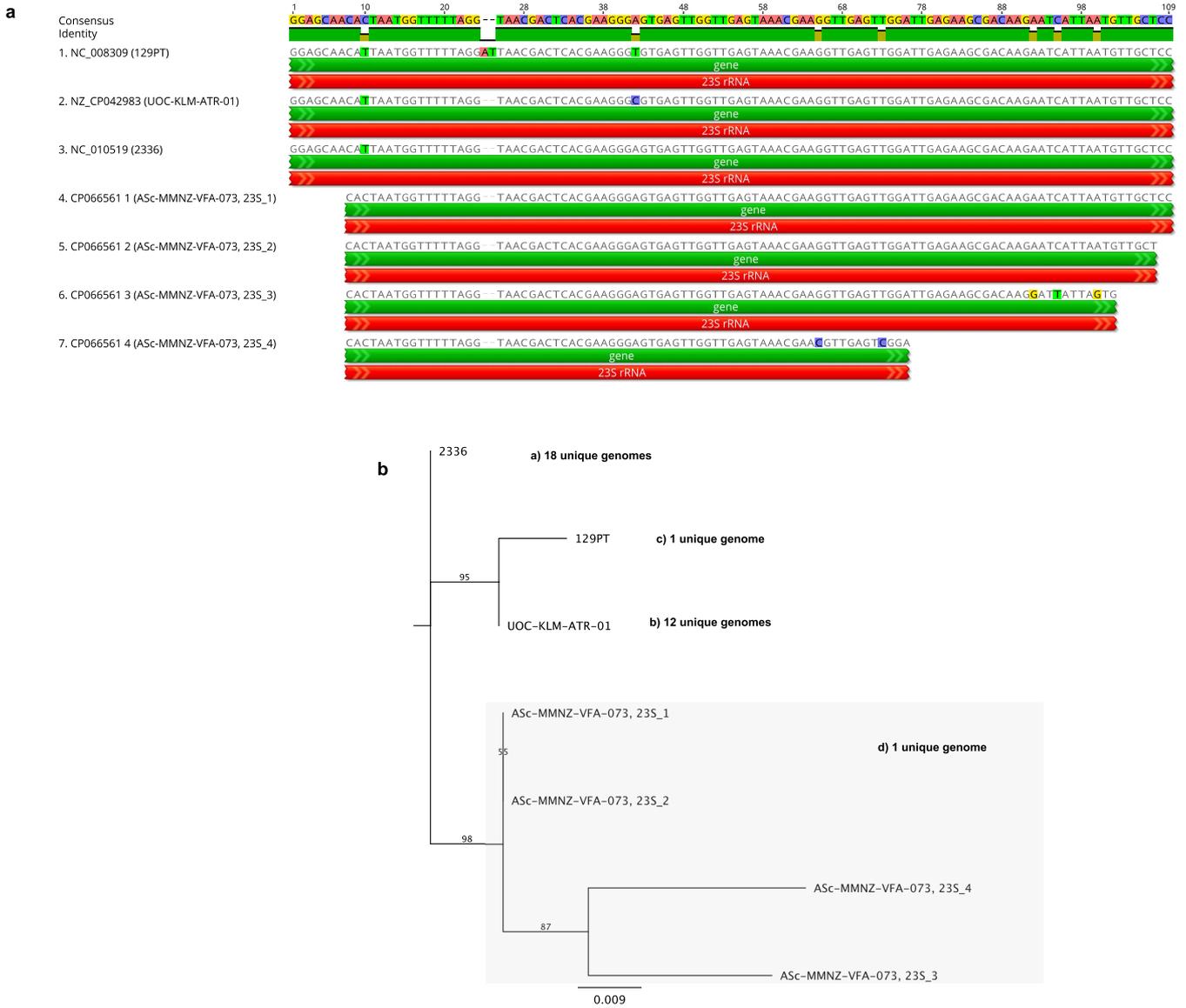


FIG 3 Phylogenetic analysis of the putative *H. somni* IVS region. (a) MAFFT (G-INS-i, 1PAM/k = 2) alignment of 7 unique subject sequences resulting from a blastn (E value, 1×10^{-3}) search of the 109-nt putative IVS against the 32 *H. somni* closed genomes. (b) The rooted RAxML tree (GTRGAMMA, -N 1000, -f a, -# autoMRE) of the alignments in a, with leaves identified by isolate and bootstrap values visible on the branches. The leaves of the tree are annotated with the number of closed genomes it represents. The query IVS sequence is identical to isolate 2336.

corresponding to identical copies within each of the 23S genes of the isolate. The single-nucleotide polymorphism (SNP) at position 10 in the consensus is due to a T-to-C transition with C dominating in the four unique sequences from ASc-MMNZ-VFA-073. Only the genome assembly of isolate ASc-MMNZ-VFA-073 exhibited intragenome IVS sequence variation among the copies of the 23S rRNA gene (Fig. 3a and b), which may represent actual variation or possibly genome assembly error. A rooted RAxML tree of the IVS sequences found in the 32 closed genomes is shown in Fig. 3b. Isolate ASc-MMNZ-VFA-073 was omitted from further analyses because of the unique characteristics demonstrated in Fig. 3a and b. Two IVS sequences (genotypes) dominate in the non-ASc-MMNZ-VFA-073 isolates, with the isolate 2336 genotype found in 18 unique closed genomes and isolate UOC-KLM-ATR-01 genotype found in 12 unique closed genomes, with a single SNP at position 42 constituting an A-to-C transversion found in isolate UOC-KLM-ATR-01 separating 30 of 32 of the closed genomes studied. As described previously, isolate 129PT exhibits both an AT insertion between bases 24 and 25 and an A-to-T transversion at position 42 in the consensus (Fig. 3a) and is the

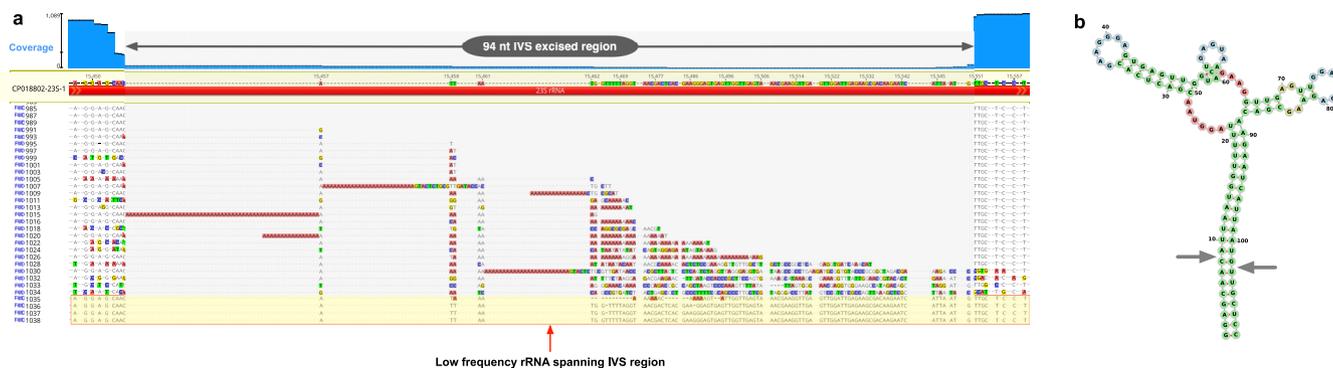


FIG 4 Identifying rRNA excised from USDA-ARS-USMARC-63250 CP018802-23S-1 query 23S rRNA. (a) The CP018802-23S-1 reference sequence is shown at the top of the alignment. Below this reference is shown a small sample of transcriptomic (cDNA) reads mapping to the IVS region, with the blue coverage bar graph at the top of the figure showing poor coverage within the 94-nt excised region. The orientation of the reference sequence is 5' to 3', left to right as indicated by the white arrowheads in the red track. Bases within reads perfectly concordant with the reference are gray; discrepant bases are colored according to the base identity. Read gaps are shown as dashes. Large gaps (>≈10 bases) in the alignment are caused by homopolymer repeats that are most likely indel artifacts of PacBio IsoSeq utilizing circular consensus sequencing (18). Most reads extending into the IVS region exhibit discrepant base alignment with the reference at the colored bases. The stair-step appearance of these discrepant reads mapping within the IVS contributes to the uneven coverage within the IVS. In most cases, if a read extends into the IVS with discrepant bases, then the bases in the region flanking the excised IVS exhibit discrepant alignments to the reference. These discrepancies suggest strongly that these reads are poor matches to the IVS region. In contrast, the raggedness within flanking regions covered by hundreds of reads is primarily due to reads, concordant with the reference, terminating in the flanking region. This behavior suggests that these reads are mapped robustly. At the bottom of the figure, outlined with a red box and highlighted in yellow are low abundance cDNA reads concordant with the reference sequence in the IVS region. These reads span the excised IVS and show perfect concordance with its flanking regions. However, read 1035 exhibits 19 deletions and 8 incorrect base calls, read 1036 shows 2 deletions, and 1037 shows a single deletion, while read 1038 perfectly matches the reference. (b) RNAfold predicted minimum free energy (MFE) rRNA stem-loop structure of the 109-nt IVS region. The MFE is predicted to be -42.68 kcal/mol constituting 14.77% of the ensemble of structures predicted. The apparent endoribonuclease cleavage sites shown at the gray arrows inferred by the 94-nt gap shown in a results in a 1-nt 3' overhang in the excised region.

sole genome exhibiting this genotype (Fig. 3b). These data show that the putative IVS region is conserved among 31 of 32 *H. somni* isolates with closed genomes with a single SNP separating 30 genomes.

Characterizing the IVS excision. The comparative genomic data reveal that the 23S genes of most sequenced *H. somni* isolates include a conserved IVS. The fate of this IVS was examined using USDA-ARS-USMARC-63250 transcriptomic data, which confirms that the central portion of the insertion was excised from the mature rRNA transcript, consistent with an IVS that leads to rRNA fragmentation. The excision boundaries were defined by mapping long transcriptomic reads (PacBio IsoSeq), generated from ribodepleted total RNA, back to the genome. An illustration of a small number of these mapped reads is shown in Fig. 4a. The IsoSeq (cDNA) coverage gap is shown to be 94 nt wide within the center of the IVS region. A read spanning the gap is shown at the bottom of the figure, but such reads were of low abundance in the IsoSeq data. A quantitative estimate of the occurrence of these low-frequency reads was not warranted due to the distortion introduced by the ribodepletion step, which would preferentially remove transcripts that include the IVS. However, the diminished occurrence of full-length IVS-containing reads is consistent with the observed qualitative reduction of the low abundance band at 3257 nt (Fig. 1a) after ribodepletion (Fig. 1b).

The details of the excision revealed in Fig. 4a were explored using the RNA secondary structure prediction program RNAfold (6, 8) that computed the IVS minimum free energy (MFE) structure (at 37°C) shown in Fig. 4b. The stable stem structure shown was based on the contiguous 20-nt complementarity between the 5' and 3' ends of the IVS, as RNAfold classified the pairing of base 21 (U) with 89 (A) as a member of a multi-loop structure and not the stem. Additional loops and stems are predicted in the center of the IVS. The predicted structure of the excised *H. somni* IVS formed by base pair complementarity between the 5' and 3' ends of the IVS and loop(s) in its center is consistent with structures predicted for IVS in other species (7). The mapped transcriptomic reads (Fig. 4a) define the boundary of extant rRNA and excised rRNA at the gray arrows on the stem shown in Fig. 4b. The most parsimonious explanation for this break in the stem, with a 1-nt 3' overhang in the excised RNA, is that it marks the cleavage

site of a double-stranded endoribonuclease. RNase III has been cited as the enzyme responsible for this cleavage in IVS of other species. However, their excised regions possess 2-nt overhangs on their 3' ends. The reasons for this discrepancy are unknown. *H. somni* USDA-ARS-USMARC-63250 includes an endoribonuclease III gene (BTV18_06945) whose gene product is the primary candidate responsible for this cleavage. A similar RNAfold analysis was performed for the IVS in NR_076444 (*H. somni* isolate 129PT) with the predicted structure (Fig. S4) having a slightly lower MFE of -43.05 kcal/mol but less than one-half of the predicted ensemble frequency at 5.81% relative to the 14.77% frequency of the MFE structure in Fig. 4b. The AT insertion at consensus positions 25 to 26 (Fig. 3a, Fig. S4) resulted in new base pairings, starting at positions 26:57 expanding into a longer imperfect stem structure than that present in Fig. 4b. The other SNP (A-to-T transversion) in isolate 129PT, position 42 in the consensus (Fig. 3a, Fig. S4) results in additional base pairing, namely, a GU wobble base pair (positions 38:44). The single SNP at position 42 (relative to isolate 2336) constituting an A-to-C transversion found in isolate UOC-KLM-ATR-01 establishes an additional GC base pairing at positions 36:42 in Fig. 4b with an MFE of -44.39 kcal/mol and ensemble frequency of 14.53%.

DISCUSSION

The combined comparative genomic, transcriptomic, and RNA capillary electrophoresis evidence presented support the presence of an IVS within 23S rRNA genes in most closed *H. somni* genomes. We used two different RNA isolation and ribodepletion combinations, namely, RNA I&D combinations 1 and 2, to investigate their effects on intact 23S rRNA abundance. Both combinations yielded negligible intact 23S rRNA but only fragments suggesting that the fragmentation was not a method artifact. Given the $\pm 20\%$ accuracy of the lowly abundant 3,257-nt electrophoretic RNA band in (Fig. 1a), it is reasonable to attribute the molecules responsible for it to intact 3017 nt 23S rRNA molecules deduced from genome sequence data. Upon ribodepletion (Fig. 1b), the lowly abundant 3,257-nt electrophoretic band (intact 23S rRNA) is dramatically reduced, consistent with the near absence of unbroken 23S cDNA reads spanning the 94-nt excised region of the IVS (Fig. 4a and b) in ribodepleted RNA. The staggered coverage of robustly mapped cDNA reads on the 3' end of the IVS flanking region on the 5' fragment (Fig. 4a) contrasts with the tighter alignment of reads defining the 5' end of the IVS flanking region on the 3' fragment. This observation suggests the possibility that 3'-to-5' exoribonucleases were more active or abundant than 5'-to-3' exoribonucleases that act on the 23S sequence fragments. This observation is consistent with the presence of multiple 3'-to-5' exoribonucleases, including exoribonucleases T (locus_tag BTV18_03270), R (locus_tag BTV18_00830), and D (locus_tag BTV18_05730), and a single 5'-to-3' exoribonuclease (exoribonuclease II [locus_tag BTV18_06455]) in the genome. The increased number of distinct 3'-to-5' exoribonucleases relative to 5'-to-3' exoribonucleases suggests that the genome biology of USDA-ARS-USMARC-63250 provides an increased opportunity for 3'-to-5' exoribonuclease activity; this conclusion is borne out in Fig. 4a. Different physical measurements mutually support the existence of an IVS in the 23S gene in USDA-ARS-USMARC-63250. A single SNP distinguishes the IVS of 30 of 32 closed genomes indicating high conservation in this IVS, suggesting the IVS may provide an important function. At this SNP, an A-to-C transversion found in isolate UOC-KLM-ATR-01 IVS relative to isolate 2336 results in an additional base pairing at positions 36:42 in the MFE structure (Fig. 4b). Although isolate 129PT exhibits a unique IVS genotype (Fig. 3b), it maintains the predicted stem-loop structure (Fig. S4) of other IVSs (Fig. 4b). Thus, genomic variation in this IVS tends to preserve the predicted RNA secondary structure. The unique IVS genotype of isolate 129PT, an isolate recovered from a healthy prepuce and reported to be nonpathogenic (5), suggests that broad field surveys with sufficient statistical power may reveal associations of the IVS genotype with tissue tropism, pathogenesis, or virulence.

The presence of IVS has been reported to be sporadic among species and isolates of most genera where it has been documented, including *Salmonella*, *Helicobacter*,

Proteus, *Providencia*, *Leptospira*, *Yersinia*, and *Campylobacter* (7). To the best of our knowledge, the widespread presence of a conserved IVS within a 23S gene has been reported previously only for *Coxiella burnetii* (3), the causative agent of Q fever. The predicted cleavage site on the *H. somni* IVS stem-loop structure has an unusual 1-nt 3'-end overhang in the excised product, as opposed to the 2-nt overhang observed in the IVS excised product of other species; this cleavage site has been attributed to the action of RNase III (9). Isolate USDA-ARS-USMARC-63250 possesses a gene annotated as endoribonuclease III (BTV18_06945), and future studies may be able to determine if this enzyme is responsible for excision of the 23S rRNA IVS. Other processes may also affect the 5' or 3' ends of the rRNA fragments unique to USDA-ARS-USMARC-63250 or *H. somni* in general.

The previously proposed biological roles for IVS and the resulting rRNA fragmentation fall into two broad categories, as follows: (i) biological roles associated with the IVS sequence or (ii) biological roles of the fragmented rRNA, where the IVSs are a means to generate such fragmented rRNA and have no other role(s). Hypothesized roles for IVSs range from protection mechanisms from bacteriocins (10, 11) to facilitating communication between endosymbiotic bacteria and their host(s) (12). We did not detect the excised IVS sequence in RNA sequence libraries, although the methods employed may bias against the detection of very short and highly structured transcripts. However, the absence of the IVSs in the transcriptomic library and in the capillary electropherogram suggests that the IVS does not directly play a regulatory role after excision, although the development of specific directed assays would be required to confirm this conclusion. Nevertheless, we believe that further study of the functions of the 23S fragments will likely yield new biological insight into *H. somni* genome biology. For example, regulating concentrations of ribosomes through degradation of fragmented rRNA has been shown to provide a selective advantage to *Salmonella* strains in response to environmental stress (13). This concept was first proposed by Hsu et al., who demonstrated that *Salmonella* isolates with fragmented rRNA exhibited an increased rate of rRNA degradation and ribosome depletion when entering stationary-phase growth, relative to isolates without fragmented rRNA. Hsu et al. posited that stationary-phase growth more closely reflects the bacterial response to nutrient limitation or stress encountered in the hostile environment of the host. Additionally, these authors further hypothesized that enhanced rRNA degradation in the stationary phase is a posttranslational regulatory mechanism allowing the organism to adapt more quickly to environmental changes. Testing similar hypotheses may lead to new insights into *H. somni* genome biology. Removing the IVSs with gene editing would provide bacterial variants where the contributions of 23S rRNA fragmentation to strain virulence, pathogenicity, or environmental adaptability could be assessed directly.

CONCLUSION

The ubiquitous presence of an IVS in the 23S genes of *H. somni* suggests that 23S rRNA fragmentation plays an important role in the biology of this organism. The high conservation of this IVS in *H. somni* distinguishes it from *Mannheimia haemolytica* and *Pasteurella multocida*, which are other *Pasteurellaceae* members associated with bovine respiratory disease. In these species, we could not find reports of IVSs in their rRNA genes. Therefore, we postulate that the IVS provides access to biological pathways dependent on 23S rRNA fragmentation that are unavailable to *M. haemolytica* or *P. multocida*.

MATERIALS AND METHODS

USDA-ARS-USMARC-63250 culture conditions. Colonies of this isolate were revived from a glycerol stock (stored at -80°C) by streaking them onto chocolate II agar (Becton, Dickinson Co. Sparks, MD, USA) in 5% CO_2 at 37°C for 48 h (14). Multiple isolated colonies from a single plate were used to inoculate 6 ml of brain heart infusion broth (Hardy Diagnostics, Santa Maria, CA, USA) supplemented with 0.1% Trizma base (Sigma-Aldrich, St. Louis, MO, USA) and 0.01% thiamine monophosphate (Sigma-Aldrich; this medium is referred to as BHI-TT [15]) in a 50-ml conical tube. The inoculated BHI-TT was vortexed, and a target optical density at 600 nm (OD_{600}) of 0.2 was verified. The total volume of the

suspension was then brought to 45 ml with fresh BHI-TT (resulting in a new OD_{600} of 0.02) and incubated at 37°C with shaking at ~ 190 rpm for 6 to 8 hours until an OD_{600} of 0.3 (mid-log-phase growth) was reached (VCN measured as $\sim 1 \times 10^9$ CFU/ml) where the cultures reach a maximum OD_{600} of 0.6 in 24 hours.

RNA isolation and ribodepletion (RNA I&D). Because of the unusual total RNA size distribution profile of the isolate, two different RNA extraction and ribodepletion method combinations were used to assess their effects on the total RNA size distribution profile. Both approaches exhibited similar RNA size distribution profiles, each without a significant 23S band at 3.1 kb in total RNA (Fig. 1, Fig. S1), suggesting that these results are not artifactual.

RNA I&D combination 1. All reagents and materials used in this protocol were certified as RNase and DNase free. From the 45-ml culture, 22 ml was placed in a fresh 50-ml conical tube and centrifuged for 3 minutes at $5,000 \times g$ at 4°C to pellet bacteria. The supernatant was decanted, and the pellet was resuspended in the remaining medium ($\sim 100 \mu\text{l}$). To the resulting pellet, 5 ml of RNAprotect (product number 76506; Qiagen, Germantown, MD, USA) was added to resuspend the cells. The cells were vortexed, incubated at room temperature for 5 minutes, and then centrifuged at $5,000 \times g$ at 4°C for 10 minutes; afterward, the supernatant was decanted. Total RNA was extracted from the resulting cell pellet using the Qiagen RNeasy midi kit protocol (product number 75144; protocol available online at www.qiagen.com/HB-2684) as per the manufacturer's instructions. Total RNA was purified from the resulting 16-ml lysate using four midi columns, and the RNA yield per column ranged from 53 to 65 μg . Total RNA (five reactions, using a maximum of 10 μg per reaction, for a total of 50 μg) was DNase treated using the Turbo DNA-free kit (product number AM1970; Invitrogen/ThermoFisher, Waltham, MA, USA) as per the manufacturer's protocol. DNase-treated RNA was pooled in a 15-ml conical tube (product number 0030 122.160; Eppendorf, Hamburg, Germany), and ethanol precipitated by adding a $0.1 \times$ volume of 3 M sodium acetate and a $3 \times$ volume of 100% ethanol. The tube was mixed by inversion and then incubated at -80°C for 1 hour. RNA was collected by centrifugation at $12,000 \times g$ for 30 minutes at 4°C. The supernatant was removed, and the pellet was washed with 2 ml of ice-cold 80% ethanol, followed by centrifugation at $12,000 \times g$ for 10 min at 4°C. The supernatant was removed, and the pellet was allowed to slightly air dry, with the tube inverted for ~ 3 to 5 min. Finally, the pellet was resuspended in 200 μl of nuclease-free water for 15 minutes at room temperature. The resulting RNA was ribodepleted using the MicrobeExpress bacterial mRNA enrichment kit (product number AM1905; Invitrogen) and two custom capture oligonucleotides, namely, oligo 1 that is 99 bp downstream from the 5' end of the 23S rRNA with sequence 5'-A(18)-CGGATATCTCGGATTATTCGCCT-3' and oligo 4 that is 356 bp downstream from the 5' end of 23S rRNA with sequence 5'-A(18)-CCC GCCCTACTTCTCGTTAGC-3' (IDT Technologies, IA). They were designed to bind selectively to the 5'-end 23S fragment. These oligonucleotides were each diluted in RNase-free water to create 5 μM stock solutions. For ribodepletion, a solution of 200 μl of binding buffer and 4 μl of capture OligoMix and each of the additional capture oligonucleotides, oligo 1 and 4, were added to an 8- μl solution of 2.5- μg total RNA in RNase-free water and processed according to the manufacturer's instructions. The size distribution of the resulting RNA fragments was analyzed using the Fragment Analyzer Automated capillary electrophoresis (CE) System (model FSv2-CE2, Advanced Analytical, now Agilent, Santa Clara, CA) with the DNF-472 total RNA kit, with typical results shown in Fig. 1a (total RNA) and Fig. 1b (ribodepleted). The population of RNA molecules responsible for the 1350-nt peak (Fig. 1b) was reduced greatly, as seen in Fig. 1c, by the addition of the custom capture oligos 1 and 4.

RNA I&D combination 2. The Ambion RiboPure bacterial kit (product number AM1925; ThermoFisher Scientific) RNA isolation protocol was modified as described here. As above, *H. somni* was cultured in 45 ml of BHI-TT, to an OD_{600} of approximately 0.3. Cells were pelleted by centrifugation for 5 minutes at $5,000 \times g$ and 20°C. The supernatant was discarded, and the pellet was resuspended in 1 ml of RNAwiz by vortexing vigorously for 10 to 15 seconds. The sample was split into thirds, with each aliquot added to ice-cold zirconia beads in a 0.5-ml screw cap tube. Subsequent cell lysis, RNA isolation, purification, and DNase treatment steps were performed as directed by the manufacturer. The Ribo-Zero rRNA removal kit for bacteria (Illumina, San Diego, CA, USA) was used for ribodepletion according to the manufacturer's directions. The size distribution of the resulting RNA fragments was analyzed using the FSv2-CE2 fragment analyzer with the DNF-472 total RNA kit, with typical results shown in Fig. S1, with Fig. S1a showing total RNA and Fig. S1b ribodepleted, showing substantially similar RNA length distribution profiles as those from RNA I&D combination 1 (Fig. 1), especially the absence of a peak in the vicinity of 3017-nt 23S rRNA. RNA isolated using this combination of RiboPure and Ribo-Zero was used for PacBio IsoSeq sequencing.

PacBio IsoSeq library preparation and sequencing. Next, poly(A) tailing of approximately 10- μg rRNA-depleted bacterial RNA was performed using a kit (PAP5104H) as recommended by the manufacturer (Lucigen, Middleton, WI, USA). Tailed RNA was purified using the RNeasy MinElute cleanup kit as recommended by the manufacturer (Qiagen, Hilden, Germany). Conversion of the tailed RNA to cDNA was performed using the SMARTer kit and tailed polyT primer as recommended by the manufacturer (Clontech, Mountain View, CA, USA). Test amplifications of cDNA were performed to determine the optimal number of cycles for amplification retaining the input RNA size distribution. The cDNA was then amplified for 12 cycles using the primers in the SMARTer kit and size selected on an ELF instrument (Sage Science, Beverly, MA, USA) to target fragments in the 1- to 6-kb range. Fractions falling within that range were pooled, and another round of 12 cycles of amplification was performed. Finally, the amplified cDNA was converted into SMRTbell libraries using the SMRTbell template prep kit 1.0 as directed by the manufacturer (Pacific BioSciences, Menlo Park, CA, USA) and sequenced on an RS II instrument using P6 chemistry.

Bioinformatics. BLAST searches were initiated within the Geneious Prime 2021 application (Biomatters Ltd., Auckland, New Zealand). The discontinuous megablast search of the USDA-ARS-USMARC-63250 23S gene (CP018802-23S-1) against the RefSeq-RNA database (Fig. 2) at NCBI (March 2021) was conducted with the following nondefault parameters: Max E-Value = 1e-100, Max Target Seqs = 100, Maximum Hits = 1000, Entrez Query = Gammaproteobacteria[Organism]. The blastn search of the IVS against all 32 closed *H. somni* available at GenBank as of February 2021 was conducted using a Max E value = 1e-5 and Maximum Hits = 1000 with all other parameters left to their default values. The resulting subject (hit) sequences were deduplicated within Geneious yielding a list of seven unique IVS sequences that were aligned with MAFFT v. 7.5 (16) using the G-INS-I algorithm, the 1PAM/k = 2 scoring matrix, gap open penalty = 1.53, and offset value of 0.123 (Fig. 3a). This alignment was used to create a rooted RaXML v. 8.2.11 (17) tree, using the command-line input parameters -m GTRGAMMA -f a -x 1 -N 1000 -p 2 -# autoMRE -x 1231 -b 1231 (Fig. 3b) to conduct a rapid Bootstrap analysis and search for the best-scoring maximum likelihood (ML) tree using extended majority-rule consensus tree criterion (autoMRE). IsoSeq reads were aligned to the USDA-ARS-USMARC-63250 closed genome sequence CP018802 using BowTie2 v. 2.3.2 with the following nondefault parameters: -l 0 -X 10000 -p 16 -very-sensitive -q. The resulting alignment is available in the Sequence Read Archive (SRA) under accession SRR15202864. The IVS sequence was analyzed at the RNAfold server website with default parameters to generate the RNA secondary structures and metrics.

Data availability. The PacBio IsoSeq reads and BAM alignment to CP018802 for the SRR1520286 SRA accession are available at the Trace Archive <https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=SRR15202864>.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 3.6 MB.

ACKNOWLEDGMENTS

We thank Kelsey McClure, William Thompson, and Kristen Kuhn, of the USMARC Corelab for their technical support in PacBio library creation and sequencing.

The mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendations or endorsement by the USDA.

Conceptualization, Formal Analysis, Investigation, Data Curation, Visualization, Writing - original draft, Writing - Review & Editing, G.P.H.; Supervision, Investigation, Methodology, Validation, Writing - Review & Editing, D.M.H.; Methodology, Investigation, Writing - Review & Editing, K.D.B.; Supervision, Investigation, Methodology, Validation, Writing - Review & Editing, T.P.L.S.

REFERENCES

- Harhay GP, Harhay DM, Bono JL, Smith TP, Capik SF, DeDonder KD, Apley MD, Lubbers BV, White BJ, Larson RL. 2017. Closed genome sequences of seven *Histophilus somni* isolates from beef calves with bovine respiratory disease complex. *Genome Announc* 5:e01099-17. <https://doi.org/10.1128/genomeA.01099-17>.
- Rosenow C, Saxena RM, Durst M, Gingeras TR. 2001. Prokaryotic RNA preparation methods useful for high density array analysis: comparison of two approaches. *Nucleic Acids Res* 29:e112. <https://doi.org/10.1093/nar/29.22.e112>.
- Warrier I, Walter MC, Frangoulidis D, Raghavan R, Hicks LD, Minnick MF. 2016. The intervening sequence of *Coxiella burnetii*: characterization and evolution. *Front Cell Infect Microbiol* 6:83. <https://doi.org/10.3389/fcimb.2016.00083>.
- Biomatters Ltd. 2021. Geneious Prime. Biomatters Ltd., Auckland, New Zealand.
- Challacombe JF, Duncan A, Brettin TS, Bruce D, Chertkov O, Detter CJ, Han CS, Misra M, Richardson P, Tapia R, Thayer N, Xie G, Inzana TJ. 2007. Complete genome sequence of *Haemophilus somni* (Histophilus somni) strain 129Pt and comparison to *Haemophilus ducreyi* 35000HP and *Haemophilus influenzae* Rd. *J Bacteriol* 189:1890–1898. <https://doi.org/10.1128/JB.01422-06>.
- Lorenz R, Bernhart SH, zu Siederdissen CH, Tafer H, Flamm C, Stadler PF, Hofacker IL. 2011. ViennaRNA package 2.0. *Algorithms Mol Biol* 6:26. <https://doi.org/10.1186/1748-7188-6-26>.
- Evguenieva-Hackenberg E. 2005. Bacterial ribosomal RNA in pieces. *Mol Microbiol* 57:318–325. <https://doi.org/10.1111/j.1365-2958.2005.04662.x>.
- University of Vienna. 2021. RNAfold WebServer. University of Vienna, Vienna, Austria. <http://ma.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi>.
- Arraiano C, Andrade J, Domingues S, Guinote I, Malecki M, Matos R, Moreira R, Pobre V, Reis F, Saramago M, Silva I, Viegas S. 2010. The critical role of RNA processing and degradation in the control of gene expression. *FEMS Microbiol Rev* 34:883–923. <https://doi.org/10.1111/j.1574-6976.2010.00242.x>.
- Skurnik M, Toivanen P. 1991. Intervening sequences (IVSs) in the 23S ribosomal RNA genes of pathogenic *Yersinia enterocolitica* strains. The IVSs in *Y. enterocolitica* and *Salmonella typhimurium* have a common origin. *Mol Microbiol* 5:585–593. <https://doi.org/10.1111/j.1365-2958.1991.tb00729.x>.
- Konkel ME, Marconi RT, Mead DJ, Cieplak W. 1994. Identification and characterization of an intervening sequence within the 23S ribosomal RNA genes of *Campylobacter jejuni*. *Mol Microbiol* 14:235–241. <https://doi.org/10.1111/j.1365-2958.1994.tb01284.x>.
- Baker BJ, Hugenholtz P, Dawson SC, Banfield JF. 2003. Extremely acidophilic protists from acid mine drainage host Rickettsiales-lineage endosymbionts that have intervening sequences in their 16S rRNA. *Appl Environ Microbiol* 69:5512–5518. <https://doi.org/10.1128/AEM.69.9.5512-5518.2003>.
- Hsu D, Shih LM, Zee YC. 1994. Degradation of rRNA in *Salmonella* strains: a novel mechanism to regulate the concentrations of rRNA and ribosomes. *J Bacteriol* 176:4761–4765. <https://doi.org/10.1128/jb.176.15.4761-4765.1994>.
- Harhay GP, Harhay DM, Bono JL, Capik SF, DeDonder KD, Apley MD, Lubbers BV, White BJ, Larson RL, Smith TPL. 2019. A computational method to quantify the effects of slipped strand mispairing on bacterial tetranucleotide repeats. *Sci Rep* 9:18087. <https://doi.org/10.1038/s41598-019-53866-z>.
- Pan Y, Tagawa Y, Champion A, Sandal I, Inzana TJ. 2018. *Histophilus somni* survives in bovine macrophages by interfering with phagosome-lysosome fusion,

- but requires IbpA for optimal serum resistance. *Infect Immun* 86:e00365-18. <https://doi.org/10.1128/IAI.00365-18>.
16. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>.
 17. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>.
 18. Wenger AM, Peluso P, Rowell WJ, Chang P-C, Hall RJ, Concepcion GT, Ebler J, Fungtammasan A, Kolesnikov A, Olson ND, Töpfer A, Alonge M, Mahmoud M, Qian Y, Chin C-S, Phillippy AM, Schatz MC, Myers G, DePristo MA, Ruan J, Marschall T, Sedlazeck FJ, Zook JM, Li H, Koren S, Carroll A, Rank DR, Hunkapiller MW. 2019. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat Biotechnol* 37:1155–1162. <https://doi.org/10.1038/s41587-019-0217-9>.