

RNA editing, RNA modifications, and transcriptional units in *Listeria monocytogenes*

Ilhan Cem Duru (✉ ilhan.duru@helsinki.fi)

University of Helsinki

Anne Ylinen

University of Helsinki

Leontina Grigore-Gurgu

Dunarea de Jos University of Galati

Christian U. Riedel

Ulm University

Lars Paulin

University of Helsinki

Petri Auvinen

University of Helsinki

Research Article

Keywords:

Posted Date: April 21st, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1530110/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: The Gram-positive foodborne pathogen *Listeria monocytogenes* can resist several stress conditions, and therefore, food processing facilities employ methods with high energy demand, such as high pressure processing (HPP), to eliminate *L. monocytogenes*. Detection of novel genomic and transcriptomic functions may enable food processing methods to better target pathogens. In this study, we aimed to reveal organization of the transcriptional units of barotolerant *L. monocytogenes* strain RO15 using novel sequencing methods such as Cappable-seq and direct RNA sequencing. The identified transcriptional units allowed prediction of transcription terminator, promoter, and operon structures. Moreover, RNA modification of the whole transcriptome by direct RNA-seq in *L. monocytogenes* strain RO15 and ScottA revealed post-transcriptional mechanisms in *L. monocytogenes*. RNA editing was investigated in both strain RO15 and ScottA to elucidate functions of RNA editing during HPP.

Results: We observed 1641 transcription start sites (TSSs) in *L. monocytogenes* strain RO15 based on Cappable-seq. Comparison of HPP-treated and control samples indicated that some prophage genes might alter their TSSs used under different conditions. Short Illumina RNA-seq reads indicate that RNA editing (A to I) occurs in *L. monocytogenes* on several genes. Some of the target genes, such as *hpf*, RNA editing happened only after HPP, indicating that RNA editing might be an important mechanism in *L. monocytogenes* when recovering of the HPP injury. TACG was the most common motif of RNA editing in both strains (RO15 and ScottA). We observed that RNA editing sites cannot be identified by long Nanopore direct RNA-seq reads with the current methodologies. The whole transcriptome RNA modification analysis showed that thousands of bases were modified in both strains. Motif analysis of the RNA modification sites detected via direct RNA sequencing indicated that a GATGA-like motif is used in *L. monocytogenes* during RNA modification. Lastly, we were able to construct the whole transcriptome annotation using the long reads, which revealed operon structures in both *L. monocytogenes* RO15 and ScottA.

Conclusion: Here, we employed two novel sequencing methods, Cappable-seq and direct RNA sequencing, for *L. monocytogenes*. The results enhanced our understanding of transcriptional units of *L. monocytogenes*. We revealed that RNA editing is used in *L. monocytogenes* during HPP recovery, which brings a new perspective to developing novel methods to eliminate *L. monocytogenes* during food processes. Whole transcriptome RNA modification analysis was performed for the first time in bacteria, potentially providing a new basis for studying novel post-transcriptional functionalities in bacteria.

Introduction

Food contaminated with pathogens is causing significant hardship for societies due to food and economic losses [1]. It also increases waste, and thus, the carbon footprint of food production. Based on the alerts in the Rapid Alert System for Food and Feed (RASFF) that were followed by food recalls or withdrawals, *Listeria monocytogenes* is one of the main pathogens contributing to economic losses [2]. Moreover, *L. monocytogenes* can cause severe infections in humans like listeriosis [3], an illness that can lead to hospitalization and mostly affects persons with compromised immune response. Listeriosis has a mortality rate of 20–30% depending on the immune status of the patient and the *L. monocytogenes* strain causing the disease [4]. Listeriosis outbreaks have been reported for different types of foods such as milk, fish, cheese, ready-to-eat foods, vegetables, and meat [5, 6]. *L. monocytogenes* can survive in different stress conditions, e.g. low temperatures, high pressures,

and acidity, posing a challenge for the food industry [3, 7]. Food products should therefore be carefully treated to avoid *L. monocytogenes* contamination or survival. High pressure processing (HPP) is a non-thermal method to inactivate pathogenic microbes, including *Listeria sp.* [8]. Previously, we have studied the effect of HPP on *L. monocytogenes*, in particular its recovery after HPP stress [9–13].

In bacteria, RNA polymerase initiates transcription by recognizing and binding to specific sequence elements on the DNA. Sequences bound by RNA polymerase depend on the sigma factor subunits of the RNA polymerase complex. For example, the most investigated alternative sigma factor in *Listeria* is sigma B, which helps RNA polymerase to bind upstream regions of stress response-related genes [14, 15]. The first DNA nucleotide position that is transcribed into RNA by the RNA polymerase complex is called the transcription start site (TSS). Identification of TSS can be done with specific RNA-seq protocols named differential RNA-seq and Cappable-seq (cap-seq), both of which enrich the 5' end of transcripts [16, 17]. Previous studies have shown that identification of TSS is useful for investigating RNA polymerase binding sites, finding novel genes and small RNAs (sRNAs), and identifying global transcriptional units in several bacterial species [18–22].

Most of the regulation of gene expression occurs at the transcription level. An additional level of variation can be produced after transcription by a process called RNA editing [23]. After RNA molecules are produced by the RNA polymerase, their sequences can be edited, leading to a difference between the DNA template and transcript [23]. It has been reported that the *tadA* gene is responsible for adenosine (A) to inosine (I) RNA editing on both tRNA and mRNA in *Escherichia coli* [24] and *Pseudomonas putida* [25]. By RNA editing, *E. coli* regulates toxicity and toxin activity [24], and *P. putida* regulates stress adaptation and pathogenicity [25]. Moreover, the GACG motif has been shown to be the binding target for the *tadA* gene in mRNAs [25]. RNA editing, to our knowledge, has not been demonstrated in *L. monocytogenes* thus far.

RNA molecules have been classically studied by short read sequencing methods like Illumina [13, 26, 27]. Using Oxford Nanopore Technology (ONT), direct RNA sequencing (RNA-seq) can be achieved, covering long continuous RNA molecules. ONT is a recent method producing long RNA-seq reads from native RNA molecules without additional steps [28]. Since the method does not require amplification or reverse transcription steps, it allows identification of RNA modifications in addition to the sequence of the assayed molecules [28]. RNA modifications regulate several different biological processes [29], such as mRNA splicing [30], and cellular fate [31]. Previously, RNA modification identification using direct RNA-seq has been used successfully in eukaryotes [32, 33] and viruses [34, 35]. To our knowledge, the RNA modification pattern of the full transcriptome has not been widely evaluated before in bacteria. However, modifications of ribosomal RNA (rRNA) have been examined extensively in e.g. *E. coli*. rRNA modifications have several central roles in bacteria, including antibiotic resistance, quality control of the rRNA structure, and ribosomal subunits assembly [36].

We have been working on the *L. monocytogenes* response to HPP, aiming to improve inactivation of these bacteria. Our studies were performed at the level of comparing genomes and gene expression using DNA sequencing and RNA-seq approaches combined with bioinformatics [10–13]. In general, there are only limited datasets available on organization of the transcriptional units of *L. monocytogenes* [21]. Therefore, we sought to elucidate the organization of the transcriptional units in *L. monocytogenes* RO15, which is a barotolerant strain, using a cap-seq approach and direct RNA sequencing by ONT. The differential usage of TSS was also evaluated in light of the HPP treatment prepared for the samples. Moreover, we examined two *L. monocytogenes* strains,

RO15 and ScottA, for possible RNA editing events using our already published data [12, 13]. We used the direct RNA sequencing data also to identify RNA modifications in rRNA and mRNA of *L. monocytogenes*.

Results

Organization of transcription by prediction of TSSs in *Listeria monocytogenes* strain RO15

To identify TSSs, data from all samples were processed with the ANNOgesic bioinformatic toolbox (see Methods for details) [37]. In total, 1641 TSSs were identified across the genome of *L. monocytogenes* RO15 (Table 1). Among all TSS, 1233 had highest coverage within 300 bp upstream of an ORF and were classified as primary TSS. Secondary TSSs (i.e. TSSs with lower coverage within 300 bp upstream of an ORF) were detected with 90 genes. The number of TSSs located within the coding sequence of a gene (i.e. internal TSSs) was 233. Moreover, 199 TSSs located on the antisense strand of a protein coding gene (antisense TSSs) and 55 orphan TSSs not assigned to any coding DNA sequence (CDS) were detected.

Table 1

Number of detected TSSs in the RO15 strain. Analysis was performed using the ANNOgesic tool (see Methods for details).

	All	Primary	Secondary	Internal	Antisense	Orphan
Number of TSS detected in combined data:	1641	1233 (75%)	90 (5%)	233 (14%)	199 (12%)	55 (3%)
Number of TSS detected in 200 MPa 0 min:	1152	907 (79%)	75 (7%)	142 (12%)	97 (8%)	47 (4%)
Number of TSS detected in 200 MPa 0 min control:	1269	1001 (79%)	75 (6%)	162 (13%)	103 (8%)	47 (4%)
Number of TSS detected in 400 MPa 0 min:	814	655 (80%)	46 (6%)	95 (12%)	59 (7%)	37 (5%)
Number of TSS detected in 400 MPa 0 min control:	1275	1005 (79%)	75 (6%)	168 (13%)	102 (8%)	50 (4%)
Number of TSS detected in 400 MPa 24 h:	1236	979 (79%)	75 (6%)	155 (13%)	97 (8%)	48 (4%)
Number of TSS detected in 400 MPa 24 h control:	1188	945 (80%)	67 (6%)	147 (12%)	99 (8%)	45 (4%)

A separate secondary ANNOgesic analysis was performed using only condition-specific samples (control and HPP-treated) to identify TSSs related to different conditions (Table 1). However, based on manual inspection, using only one sample without replication caused several false-negatives. Therefore, we mainly used combined data.

The TSSs of strain EGD-e have been previously identified, and strain EGD-e was shown to contain 1576 TSSs [21]. To compare our TSS prediction in RO15 and EGD-e TSSs, [21] we lifted over (mapped) the EGD-e TSSs to the RO15 genome. For this, we first aligned these two genomes assigning their collinearity (Figure S1). Almost all EGD-e TSSs (1531 of 1576) were successfully mapped to the RO15 genome and the following analysis revealed that 876 of 1531 lifted TSS positions were the same as the TSS positions that we predicted in RO15 (Table S1).

Prediction of sRNAs

Prediction of sRNAs using ANNOgesic tool indicated that *L. monocytogenes* RO15 had 81 sRNAs (Table S2). Most of them (53 of 81) were identified as intergenic sRNA. In total, 20 sRNAs were recognized as untranslated region (UTR) -derived sRNA (four 3'UTR derived and 16 5'UTR derived). Only six of them were identified as antisense sRNA (asRNA). Lastly, two of the predicted sRNAs were recognized as InterCDS sRNA (located within the coding sequence). After the sRNA prediction, we performed a gene count to investigate sRNA up- and downregulation during HPP treatment. We have not seen any anticorrelation for fold changes between asRNA and the corresponding gene (Figure S2). Previously, 70 antisense TSSs were predicted in EGD-e [21], and nine asRNAs were verified by Northern blot [21]. Within the 70 antisense TSSs of EGD-e [21], 10 (10 of 70 antisense TSS) of the same antisense TSSs was also predicted in RO15 based on our TSS lift over data (Table S3). Moreover, we checked the nine verified asRNA [21] in our data and observed that antisense TSS can be seen in RO15 at the beginning of two of the verified asRNA in EGD-e (Figure S3). For one of them, an sRNA was predicted in the same location by ANNOgesic, but since it was not overlapping with the following gene, it is marked as intergenic sRNA (Figure S3).

Prediction of terminators, UTR length, and circ-RNAs

ANNOgesic tool allowed us to predict rho-independent transcription terminators, UTR length, and circular RNAs (circ-RNAs). The transcription terminator prediction resulted in the prediction of 79 terminators in the RO15 genome (Table S4). UTR prediction indicated that the median 5' UTR length is 33 bases in *L. monocytogenes* RO15 when we use primary TSSs. UTR length was also calculated using both primary and secondary TSSs, resulting in a median of 34 bases (Figure S4). The distribution indicates that most of the UTRs were shorter than 100 nucleotides (Figure S4). In addition, we were able to predict 10 circ-RNAs using ANNOgesic tool, all of which were localized on the forward strand (Table S5). We also performed differential expression, including circ-RNAs, and found that some of the circ-RNAs were upregulated after HPP (Figure S5).

Prediction of Promoters

When localizing the TSS, we were prompted to look for the promoter structures based on our data. Upstream region sequence (-50 bp to 1 bp) of all predicted TSSs was analysed using the MEME suite to identify promoter motifs in *L. monocytogenes*. For each TSS type, MEME suite reported motifs with low E-value. Predicted motifs showed that there was only one well-conserved region on the promoter regions, with a -10 position "TATAAT"-like motif. Promoter motif sequences of all types of TSSs were similar to each other. All types of TSSs were clearly observed to be either "A" or "G". The "TATAAT"-like motif was slightly different in secondary TSS promoter regions compared with other types of TSS regions (Fig. 1).

TSS of prophages

While analysing the genome of RO15 strain [12] and gene expression in response to HPP stress [13], we found that prophages were linked to the phenotypic characteristics of the strain [12]. We next analysed in detail the TSS and genome organization of these loci. RO15 had five different prophage regions with a size of 10.7 kb (prophage 1, 125824–136550), 33.2 kb (prophage 2, 673631–706883), 47.9 kb (prophage 3, 2175980–2223958), 42.7 kb (prophage 4, 2559206–2601984), and 41.3 kb (prophage 5, 2729417–2770759) (Table 2). Prophage 5 (Fig. 2) was also seen as a circular form, representing a free virus genome version [12]. We

specifically checked these regions to understand the prophage gene TSS structure. Prophage 5 had the largest number of TSSs across all prophage regions (Table 2). The ratio of antisense TSS was also higher in prophage 5 than in the other prophages (Table 2). It has previously it has been shown that 10 prophage genes (Table S6) might be specific to barotolerant strains [12]. Therefore, we first focused on these 10 prophage genes (Table S6). Only the gene OCPFDLNE_02580 (annotated as a hypothetical protein) had a clear TSS (Figure S6) based on our prediction. The TSS could be detected in all conditions, although direct RNA-seq reads were present only in the HPP treated sample (Figure S6).

Table 2

Prophage regions in R015 and the number of TSS. Phaster score > 90; intact, Phaster score 70–90; questionable.

Prophage Region	Number of genes	Primary TSS	Internal TSS	Anti sense TSS	Orphan TSS	Total
Prophage 1 (125824–136550 bp) Phaster score 70	17	1	1	0	0	2
Prophage 2 (673631–706883 bp) Phaster score 110	44	1	4	1	0	6
Prophage 3 (2175980–2223958 bp) Phaster score 70	79	7	4	3	1	15
Prophage 4 (2559206–2601984 bp) Phaster score 80	64	7	0	3	2	12
Prophage 5 (2729417–2770759 bp) Phaster score 84	67	10	1	7	0	18

Moreover, we investigated condition-specific prophage TSSs. We predicted an internal TSS (TSS:679897_f) for the gene OCPFDLNE_00657 (located in prophage 2, annotated as cytosine-specific methyltransferase). For this internal TSS, cap-seq reads were only seen for control samples (Figure S7). However, neither Illumina RNA-seq nor Nanopore direct RNA-seq allowed us to observe a clear expression signal for the gene OCPFDLNE_00657. Similarly, the predicted primary TSS (TSS:2178605_r) for gene OCPFDLNE_02163 (located in prophage 3, annotated as DUF4429 domain-containing protein) was only seen in control samples (Figure S7). The gene OCPFDLNE_02163 was also downregulated in HPP-treated samples at several time points. The predicted antisense TSS (TSS:2221338_f) for the gene OCPFDLNE_02235 (located in prophage 3, annotated as phage terminase large subunit) had only cap-seq reads in HPP-treated samples (Figure S7). The gene OCPFDLNE_02235 was mainly upregulated in treated samples, except for 400 MPa at the 48-hour time point [13].

TSS of anti-CRISPR genes

We were specifically interested in TSSs of anti-CRISPR genes due to the observed prophage activity related to HPP treatment seen earlier in this strain [13]. TSS predictions showed that both predicted anti-CRISPR genes (*acrIIA1* (OCPFDLNE_02770, OCPFDLNE_02583) and *acrIIA2* (OCPFDLNE_02582)) belonged to one operon that contains four genes and one TSS (Fig. 3). Nanopore direct RNA-seq long reads also supported these four genes expressed together (Fig. 3). On the *Cas* gene region, we did not observe any Illumina/Nanopore RNA-seq or cap-seq reads, therefore, no TSS was predicted in this region. *L. monocytogenes* RO15 has two long CRISPR array regions (region 1; 526524 bp – 527202 bp, region2; 543628 bp – 547169 bp), including 10 and 54 spacer sequences, respectively. Six of the total of 64 spacer sequences are self-spacers [12]. Interestingly, we observed both RNA-seq (short and long reads) and cap-seq reads at the beginning of both CRISPR region array regions and TSS was predicted for both CRISPR array regions (Figure S8).

TSS of genes/pathways relevant for recovery of after HPP treatment

There were several genes and pathways for which expression correlates with HPP treatment [13]. We examined the TSSs of selected target genes from these lists. A large operon, including cobalamin biosynthesis genes (OCPFDLNE_01234 - OCPFDLNE_01251), is upregulated during HPP treatment [13]. One primary TSS was predicted for this large operon (OCPFDLNE_01234 - OCPFDLNE_01251), but there was no difference between control and treated samples from the point of TSS prediction and usage of the sites (Figure S9). Similarly, we did not observe any difference for TSS prediction between control and treatment samples (Figure S9) for ethanolamine utilization genes (OCPFDLNE_01215 - OCPFDLNE_01231), which are downregulated with HPP [13].

Identification of RNA variants

We investigated RNA modifications of two different forms: first, the variants in which nucleotide identity is altered by the modification (RNA editing) and second, in a later section, more general RNA modifications, which are methylation or other modifications that do not alter the identity of the base, only its possible functionality are described. We analysed the already published RNA-seq data [13] in detail while working on the TSS analysis and identified sequence variants that were not sequencing mistakes and not observed in the gDNA-based PacBio or Illumina data [12]. We identified several A > G (T > C for reverse strand) variants in both strains, which have been described as RNA editing signals in other bacteria [24, 25]. The numbers of samples with variants were higher in HPP treatment conditions in several transcripts but not in all (Table 3, Table 4, Figure S10, Table S7). A fraction of the variants was only seen in HPP-treated samples. Nevertheless, some of the variants were also seen in control samples such as *pflA*, *mdxE*, *secA*, and *kdpD* in strain RO15 and *actA*, *ldh*, *dhaS*, LMOSA_26480, and *mdxE* in strain ScottA. Almost all variants were observed as partial variants, meaning around 50% of the RNA-seq reads had a variant base in the transcript. We also observed that the editing percentage increases with time of HPP (Figure S11, Figure S12). However, we did not observe a clear change in expression pattern for the genes with RNA variants; both upregulation and downregulation were detected for those genes (Figure S13).

Table 3

List of A > G variant positions and the corresponding gene in strain R015. Positions of A to G (T to C on reverse strand) variants in RNA of R015. Locus tag of corresponding gene and gene name also shown. Variants were detected in variable numbers of samples, some seen in nearly all samples (*pflA*). For the sake of clarity, variants detected in more than five samples (total of 104) are presented (Table S7).

Variant Pos.	Locus Tag	Gene Name	Ref.	Variant	# of treated sample variant seen (n = 51)	# of control sample variant seen (n = 53)	ref. codon	variant codon
204136	OCPFDLNE_00200	spoVG_1	A	G	9	1	TAT > Y	TGT > C
322606	OCPFDLNE_00309	walR	A	G	5	0	ACG > T	GCG > A
429958	OCPFDLNE_00419	mngB_1	A	G	8	0	TAC > Y	TGC > C
497744	OCPFDLNE_00478	dhaS	A	G	12	0	ACG > T	GCG > A
573660	OCPFDLNE_00547	pbpX	A	G	6	0	TAC > Y	TGC > C
1213571	OCPFDLNE_01217	pdtaS	A	G	9	3	TTA > L	TTG > L
1451307	OCPFDLNE_01455	pflA	A	G	51	47	TTA > L	TTG > L
1544500	OCPFDLNE_01549*		T	C	9	0	TAC > Y	TGC > C
1768810	OCPFDLNE_01761	yfhP	A	G	10	0	TTA > L	TTG > L
1801548	OCPFDLNE_01798*	iolU_2	T	C	7	0	ACG > T	GCG > A
2170275	OCPFDLNE_02153*	rimI	T	C	5	3	ACG > T	GCG > A
2269171	OCPFDLNE_02280*	mdxE	T	C	37	10	TAT > Y	TGT > C
2298917	OCPFDLNE_02315*	nrdE	T	C	5	0	TAT > Y	TGT > C
2428075	OCPFDLNE_02434	pepC	A	G	8	1	ATG > M	GTG > V
2467402	OCPFDLNE_02484	patB	A	G	7	0	ACG > T	GCG > A
2485029	OCPFDLNE_02503*		T	C	7	1	TTA > L	TTG > L

Variant Pos.	Locus Tag	Gene Name	Ref.	Variant	# of treated sample variant seen (n = 51)	# of control sample variant seen (n = 53)	ref. codon	variant codon
2646986	OCPFDLNE_02684*	secA	T	C	26	14	TTA > L	TTG > L
2649787	OCPFDLNE_02685*	hpf	T	C	21	1	CTA > L	CTG > L
2760956	OCPFDLNE_02809*		T	C	6	2	ACG > T	GCG > A
2857805	OCPFDLNE_02934*	kdpD	T	C	5	5	TTA > L	TTG > L
3016678	OCPFDLNE_03082*	garK_2	T	C	3	0	TTA > L	TTG > L
2288303	sRNA_92	rli47	A	G	8	1	-	-

Table 4

List of A > G variant positions and the corresponding gene in strain ScottA. Positions of A to G (T to C on reverse strand) variants in RNA of the ScottA genome. Locus tag of corresponding gene and gene name also shown. The numbers of samples in which the variants were detected are also shown. For the sake of clarity, variants detected in more than five samples (total of 107) are presented (Table S7).

Variant Pos.	Locus_tag	Gene name	Ref	Variant	# of treated sample variant seen (n = 53)	# of control sample variant seen (n = 54)	ref. codon	variant codon
242871	LMOSA_10890	spoVG	A	G	12	0	TAT > Y	TGT > C
252859	LMOSA_10970	actA	A	G	39	26	ACA > T	ACG > T
257308	LMOSA_11030	ldh	A	G	53	54	ATT > I	GTT > V
320780	LMOSA_11570	rpoB	A	G	5	1	CTA > L	CTG > L
403379	LMOSA_12240	-	A	G	2	6	ACG > T	GCG > A
568103	LMOSA_13830	dhaS	A	G	13	21	ACG > T	GCG > A
720351	intergenic_region	intergenic_region	A	G	10	0	-	-
1282535	LMOSA_21080	-	T	G	11	4	TTT > F	TTG > L
1586671	LMOSA_24230	-	T	C	7	0	TAC > Y	TGC > C
1630529	LMOSA_24610	glpF	T	C	4	1	TAC > Y	TGC > C
1717344	LMOSA_25410	pepV	T	C	15	1	TAC > Y	TGC > C
1731486	LMOSA_25560	-	A	G	5	1	TAC > Y	TGC > C
1775078	LMOSA_25880	-	T	C	10	0	ACG > T	GCG > A
1843087	LMOSA_26480	-	T	C	53	52	GGA > G	GGG > G
2229297	LMOSA_630	rimI	T	C	4	2	ACG > T	GCG > A
2281668	LMOSA_1140	mdxE	T	C	14	4	TAT > Y	TGT > C

Variant Pos.	Locus_tag	Gene name	Ref	Variant	# of treated sample variant seen (n = 53)	# of control sample variant seen (n = 54)	ref. codon	variant codon
2301135	LMOSA_1310	-	A	G	8	0	rev reads	rev reads
2324765	LMOSA_1570	gloA	A	G	12	0	TAC > Y	TGC > C
2481510	LMOSA_2970	pepC	A	G	11	1	ATG > M	GTG > V
2520887	LMOSA_3290	patB	A	G	10	0	ACG > T	GCG > A
2548257	LMOSA_3580	-	T	C	7	0	TTA > L	TTG > L
2619490	LMOSA_4270	rpoN	T	C	6	3	CTA > L	CTG > L
2675586	LMOSA_4770	hpf	T	C	21	0	CTA > L	CTG > L
2715706	LMOSA_5200	-	T	C	9	2	ACG > T	GCG > A
2996208	LMOSA_7990	glxK	T	C	5	0	TTA > L	TTG > L

We analysed the localization of the RNA edited sites with MEME (Fig. 4). The modified sites have different preferences between R015 and Scott A stains. Although the predicted consensus motif by MEME was not similar between the two strains, the same two four-base motifs (TACG and GTAA) were the most common edited motifs in both strains (Fig. 4).

Direct RNA-seq with long reads

To use ONT direct RNA-seq on bacteria, we needed to develop a custom sequencing protocol since it requires poly-A tails. We added poly-A tails preferentially to mRNAs to produce direct RNA reads (Fig. 5, see Methods section).

We sequenced four samples that were used in our previous HPP experiment [13] with ONT to produce long direct RNA sequencing reads (Table 5). Long direct RNA reads allowed us to validate the predicted TSS and transcription terminators. Moreover, long read data revealed the operon-like structures due to the long continuous reads (Table 5). The annotation of transcripts using long reads revealed 940 transcripts in R015 and 925 transcripts in ScottA (Supplemental File 1, Supplemental File 2).

Table 5

Statistics of direct RNA-seq. In each run we had the yeast enolase transcript as control, which lowered the reads mapped to the target genome.

Strain and samples identity	Number of reads obtained	Number of reads mapped (% of total obtained)	Number of reads mapped to rRNA (% of total mapped)	Mean read length of mapped Reads	Longest mapped read length
RO15 400 MPa 24h (R173)	454540	198231 (43.61%)	153444 (77.41%)	692.2	7874
RO15 Control 24h (R312)	338926	180960 (53.39%)	135159 (74.69%)	765.3	4937
ScottA 400 MPa 24h (R158)	314044	200127 (63.73%)	179614 (89.75%)	903.3	6072
ScottA Control 24h (R317)	1613144	1500901 (93.04%)	1321491 (88.10%)	969.5	9121

De novo RNA modification site prediction

De novo detection of modified bases was performed using Tombo software [38]. The results indicated that 4415 modified bases were identified for the *L. monocytogenes* RO15 HPP-treated sample (400 MPa treated – 24 h), 2665 modified bases for *L. monocytogenes* RO15 control sample (0.1 MPa – 24 h), 2665 for the ScottA HPP-treated sample (400 MPa – 24 h), and 8558 for the ScottA control sample (0.1 MPa – 24 h). In all samples, the majority of the modified bases were located within protein coding genes (Table S8, Table S9, Table S10, Table S11), while modified bases were also observed in rRNA transcripts, transfer-messenger RNAs (tmRNAs), and intergenic regions. The 15 bp length sequence surrounding the modified base was used for further motif enrichment analysis, which revealed a significant GATGA-like motif adjacent to the modified bases (Figure S14).

We observed that in RO15 some portion of the RNA-modified bases were co-localized with the genome DNA-modified bases according to our recent PacBio data [12] (Table S8, Table S9). However, based on the binomial test, the same strand co-localization of RNA and DNA modifications was seen by chance ($p > 0.05$).

We used Tombo “level_sample_compare” to perform a two RNA sample comparison (control vs. treatment) for modified base detection. In RO15, we did not observe a significant difference between control and treatment samples. In ScottA, a significant difference (D-statistic > 0.8) was observed at the base located at nucleotide 1919280. The base position is located within the gene EGJ25327.1 (Locus tag; LMOSA_24140, annotated as a hypothetical protein).

We did not observe a clear RNA modification signal for the RNA edited sites in direct RNA-seq reads RO15 or ScottA. However, for some of the genes (*pfIA*, *iolU*, and *hpf* in RO15 and *hpf* in ScottA), a partial variant (single

nucleotide variation; SNV) was observed in direct RNA-seq reads at the position one nucleotide before (or after, depending on gene strand) the observed RNA editing sites (Figure S15, Figure S16).

Moreover, we noted that none of the partial variants that were identified with Illumina RNA-seq were captured by Nanopore direct RNA-seq (Figure S17).

Discussion

Sequencing technologies and methodologies are improving constantly, allowing investigations of features of both genome and transcriptome from different perspectives. In this study, we used state-of-the-art methods, such as cap-seq and direct RNA-seq, to obtain a genome-wide view of the transcriptional units and to evaluate RNA modifications in *L. monocytogenes*. Moreover, previous Illumina RNA-seq reads [13] were re-investigated with the aim of identifying RNA editing sites in *L. monocytogenes*.

TSSs of *L. monocytogenes* EGD-e and non-pathogenic *L. innocua* have been studied previously [21]. Here, we evaluated a new strain and possible alternative usage of TSSs when cells were treated with HPP, causing extreme stress. We report, for the first time, identification of TSSs in *L. monocytogenes* by the cap-seq approach. To compare our R015 TSS prediction results and previous studies' results on strain EGD-e [21], we lifted the identified EGD-e TSSs to the R015 genome. Most TSSs could be lifted to the R015 genome, but only 57% of the lifted TSSs (876/1531) were in the same position as our TSS prediction using cap-seq. To understand such differences, we obtained the EGD-e reads that were used previously [21] and mapped the R015 genome. Comparison of the EGD-e data [21] with our cap-seq reads indicated that EGD-e reads [21] are noisier than our cap-seq, which may cause false-positives. Moreover, these two studies ([21] and this study) use different sequencing, mapping, and TSS prediction methods, which might explain discrepant results. One might also speculate that TSSs across the genomes may differ between strains due to a large difference between strains such as a different clonal complex and multilocus sequence type.

Subsequently, we focused on the group of genes found to be relevant for the recovery process after HPP treatment [13], including prophages, anti-CRISPR genes, ethanolamine utilization, and cobalamin synthesis. We looked at TSSs in these regions more closely and compared control and treated samples' TSS predictions. The only clear difference between control and treated samples TSS predictions was found in the prophage regions. Here, two TSSs were predicted in control samples, and one TSS was predicted in treated samples. Therefore, we speculated that prophages can change TSS usage possibly due to stress-induced mechanisms.

Previously, we detected *acrIIA1* and *acrIIA2* anti-CRISPR genes within strain R015 genome using homology-based prediction based on known anti-CRISPR genes [12]. It is possible that new candidate anti-CRISPR genes can be identified using the 'guilt-by-association' approach [39]. Based on our TSS prediction, anti-CRISPR genes (*acrIIA1* and *acrIIA2*) were found in an operon consisting of four genes. Moreover, long RNA-seq reads also supported the predicted four-gene operon structure. We believe that the rest of the genes in the operon might be *acrIIA3* and *acrIIA4*, which were defined before in *L. monocytogenes* [40] but without any sequence homology. A four-gene operon can be identified in EGD-e, which suggests that in addition to lmo2274, lmo2275, and lmo2276 the gene lmo2273 might be an *acr* family gene without homology to known family members.

In *Listeria*, sRNAs play a role in several processes including pathogenicity and host interaction [41]. In total, 81 sRNA were predicted in *L. monocytogenes* R015 using our data. Previously, similar methods (cap-seq and

ANNOgesic analysis) were employed for *Bacteroides thetaiotaomicron* [18], and the study successfully validated several predicted sRNAs using Northern blot assays [18]. Therefore, we expected to obtain accurate predictions from the ANNOgesic sRNA predictions. Within the predicted 81 sRNA *L. monocytogenes* RO15, six of them were identified as asRNA. However, identified asRNAs and corresponding genes were not anticorrelated based on the log₂fold change (Figure S2). It is possible that the expression of asRNAs is difficult to identify with RNA-seq, and therefore, we cannot see a clear anticorrelation. We predicted 199 antisense TSS, but only six asRNA were identified by ANNOgesic in *L. monocytogenes* RO15. Previously, nine asRNA were verified in *L. monocytogenes* EGD-e using earlier TSS sequencing methods and Northern blot assays [21]. We mapped these nine asRNA [21] to *L. monocytogenes* RO15 and observed that antisense TSS were predicted at the start of only two out of the nine verified asRNAs. These two strains (EGD-e and RO15) may harbour different asRNAs.

The median 5' UTR length was calculated as 33 bases in *L. monocytogenes* RO15. Previously, it was predicted that both *L. monocytogenes* EGD-e and *L. innocua* BUG499 also have a median length of 33 nucleotides 5' UTR [21] and *B. thetaiotaomicron* a median length of 32 nucleotides 5' UTR [18]. Moreover, a very similar 5' UTR length distribution was also predicted in *E. coli* and *Klebsiella pneumoniae* [42]. Thus, our prediction of 5' UTR length of *L. monocytogenes* RO15 is in line with published data for *L. monocytogenes* and other bacteria. It should be also noted that variation in median 5' UTR length was reported in some species such as median 5' UTR length of 27 nucleotides in cyanobacterium *Prochlorococcus* MED4 [43] and 42 nucleotides in cyanobacterium *Synechocystis* sp. PCC6803 [44].

The ANNOgesic "circrna" module predicted 10 circ-RNAs in *L. monocytogenes* RO15. Interestingly, all 10 predicted circ-RNAs were on the positive strand. Transcriptional analysis indicated that circ-RNAs are differentially expressed after HPP. The function of circ-RNA in bacteria is unknown, but it has been shown that bacteria can translate circ-RNA [45].

Prokaryotic promoter regions generally include -35 and -10 elements, which affect promoter activity [46]. Promoter motif analysis indicated that only the TATAAT-like -10 region could be clearly identified in *L. monocytogenes* RO15, while no clear -35 consensus motifs were found. However, promoters of some genes appear to have a -35 motif starting with "TTG" similar to the "TTGACA" -35 region of *E. coli* [47]. Since no clear -35 motif was predicted, we believe that several genes in *L. monocytogenes* RO15 lack a -35 element in the promoter region. In addition, we observed an extended -10 element "TG" (also known as -15 element) [48] in RO15, which might compensate for the lack of a -35 element [49].

In re-analysing our published RNA-seq data [13], we observed RNA variants (A to G, or A to I) compared with the genome sequence for several genes in both RO15 and ScottA, especially after pressure treatment. A to I editing in the *hokB* transcript was shown to increase toxicity of *E. coli* [24]. More recently, it has been shown that A to I RNA editing in the *fliC* transcript increases tolerance of *P. putida* to oxidative stress [25]. In our study, as RNA editing was mostly seen in HPP-treated samples, we hypothesize that RNA editing may be related to HPP stress response of *L. monocytogenes*. Therefore, RNA edited transcripts might have an important role in HPP stress in *L. monocytogenes*, and these genes can be a target for future research. In total, 12 genes were RNA edited in both strains. Especially the gene *hpf* encoding a ribosome hibernation-promoting factor drew our attention since the difference between treated samples and control samples increased over time. The gene *hpf* has an important role in *Listeria*. When *Listeria* encounters a stress condition, *hpf* converts translationally active 70S into inactive 100S ribosomes, which contributes to stationary-phase survival [50]. During HPP recovery *L. monocytogenes*

might improve efficiency of ribosome hibernation by editing the *hpf* transcript. It has been also reported that the *hpf* gene can be found in phages [51]. However, in both RO15 and ScottA, the *hpf* gene was not present in any prophage region.

Direct RNA sequencing is a novel method that creates long reads from native RNA, which allow us to predict full-length transcripts and RNA modifications [28]. For direct RNA sequencing, we selected four samples in total from both *L. monocytogenes* RO15 and ScottA strains, which were used in our previous HPP experiment [13]. Samples were selected based on earlier observations of abundant differentially expressed genes and high numbers of RNA editing events. Direct RNA sequencing allowed us to predict transcripts, providing a general view of gene structures and operons. The number of predicted transcripts was lower than the number of genes in both strains, due to operon structures. However, some regions of the genome did not have sufficient coverage to predict a transcript from the reads. Hence, transcriptional inactivity was another reason for the lower number of predicted transcripts.

Analysing the RNA modifications using direct RNA-seq reads indicated that a GATGA-like motif is used in both RO15 and ScottA for RNA modification. RNA modification can regulate virulence and resistance in pathogens [52]. To our knowledge, RNA modification of whole transcripts in bacteria has not been reported to date; therefore, this is the first proposed whole transcript-based RNA modification motif in bacteria.

No RNA modification signal was observed in RNA editing sites. This might indicate that RNA modifications are caused by a different mechanism than RNA editing. Moreover, we have not observed partial variants in the RNA editing sites using the direct RNA-seq reads. Therefore, we believe that direct RNA-seq is not suitable for investigating A to I RNA editing sites. RNA editing site prediction may require other sequencing methods such as Illumina.

In summary, we identified several features of the transcriptional unit landscape of *L. monocytogenes*, including transcription start sites and terminators, promoters, operon, RNA transcripts, RNA editing events, and RNA modifications. We compared EGD-e [21] and RO15 TSSs. More experimental work is needed to fully elucidate these findings linked to functionalities of *L. monocytogenes* and the relevance of our observation of growth and possible virulence of these bacteria.

Methods

Library processing, sequencing, and mapping for TSS

TSSs were determined using the cap-seq method [17] and the protocol published by New England Biolabs (<https://international.neb.com/protocols/2018/01/19/cappable-seq-for-prokaryotic-transcription-start-site-determination>). Before RNA sequencing library preparation, 15 µl of decapped RNA was vacuum-dried in DNA Speed Vac DNA110 (Savant) at a low drying rate for about 40 min and then suspended in 2.5 µl of ultra-pure water. RNA sequencing libraries were prepared with TruSeq Small RNA Library Prep (Illumina) using half of the kit's volumes according to the manufacturer's instructions. Libraries (half of the reaction volume) were pooled and concentrated using Amicon Ultra-0.5 100K filter device (Millipore) (Table 6). Size selection of 140–250 bp fragments was performed using BluePippin and 3% agarose gel cassette (Sage Science). NextSeq 500 (Illumina) was used to sequence the RNA-seq libraries.

Table 6
Sequences, lengths, and melting temperatures of rRNA blocking oligos.

Blocking oligo	Length (bases)	Sequence	T _m (°C)
16S_block	28	AGG TGA TCC AGC CGC AGG TTC TCC TAC G	66,5
23S_block	33	TTG GTT AAG TCC TCG ACC GAT TAG TAC TAG TCC	61,1
5S_block	26	TGC GTG GCA ACG TCC TAT CCT CGC AG	66,2
16S_HT107_block	34	CGG GTC CAT CCT AAA GTG ATA GCC GAA ACC ATC T	64,6
16S_HT682_block	30	TCC TGT TTG CTA CCC ATG CTT TCG AGC CTC	65,0
16S_HT1241_block	36	CCG CGG CAT GCT GAT CCG CGA TTA CTA GCG ATT CCG	69,9
23S_HT375_block	33	CAC CTT TCC CTC ACG GTA CTG GTT CAC TAT CGG	65,0
23S_HT1421_block	38	ACA TCA GGA ACT TCG CTA CTT AAT TTC GCT CCC CAT CA	65,0
23S_HT1641_block	34	TTG TTT GGG CCT ATT CAC TGC GGC TGA CCT GAC G	68,2

Read processing and mapping for TSS

The same preprocessing was performed for TSS-seq reads as described previously for RNA-seq reads [13], except using “TGGAATTCTCGGGTGCCAAGG” as an adapter during the adapter filtering step. Both RNA-seq and TSS-seq reads were mapped to the reference genome using the READemption v0.5.0 pipeline [53] with default options. Briefly, the pipeline uses Segemehl v0.2 [54] with minimal accuracy of 95%. Coverage and normalization of coverage was calculated using READemption v0.5.0 “coverage” function. The coverage was normalized by the total number of aligned reads and multiplied by the lowest number of aligned reads of all input libraries.

Identification of TSSs in RO15

We used two approaches: 1) each condition’s reads were used separately and 2) all reads were merged to one file. TSS identification was done by comparing the relative coverage of reads between TEX-treated samples and untreated samples. For this purpose, we used ANNOgesic v1.0.16 pipeline [37] with TSSpredator v1.06 [55]. To get the optimal parameters, we first manually annotated the first 50 kbp region. Manually annotated file was employed to obtain the optimal parameters using ‘annogesic optimize_tss_ps’. Then, “annogesic tss_ps” command was run using ‘-c normPercentile = 0.8, texNormPercentile = 0.2, allowedCompareShift = 4, allowedRepCompareShift = 4’.

Identification of UTR, transcription terminator, sRNA, sORF, and circ-RNA in RO15

We used ANNOgesic v1.0.16 pipeline [37] to predict additional UTR, transcription terminators, sORF, and circ-RNA from combined data. “annogesic terminator” was used with ‘-replicate_tex all_1’ to predict transcription terminators. “annogesic utr” with default options was used for UTR length prediction. “annogesic srna” and “annogesic sorf” with ‘-tex_notex 1’ were used for sRNA prediction and sORF prediction. “annogesic circrna” was used with default options to predict circ-RNAs in RO15.

EGD-e TSS lift over to R015

The genome sequences of EGD-e and R015 were aligned using LASTZ v1.04.15 [56] with “-chain -format = axt” options. We then chained the “axt” alignment files using axtChain [57] and generated chain format output. Chain format output was used to lift EGD-e TSS gff to R015 genome using CrossMap v0.5.4 [58].

Variant calling

We used the RNA-seq data obtained in our HPP experiment study [13]. RNA-seq preprocessing was described previously [13]. RNA-seq reads were mapped to the reference genome using Bowtie2 v2.3.4.3 [59] default options. The output was sorted and converted to BAM format using Samtools v1.9 [60]. Bcftools v1.9 [61] “mpileup” function with default options was used to generate genotype likelihoods. Bcftools v1.9 “call” function with “-mv -Ob” options was used for variant calling.

Direct RNA sequencing

Four RNA samples were additionally analysed with Direct RNA sequencing. Since the protocol assumes mRNAs to have poly-A tails, they were first added by treating 14.5 µl of total RNA (3-7.8 µg) with 7.5 units of *E. coli* poly(A) polymerase (New England Biolabs) and 2 mM ATP in 1x reaction buffer at 37°C for 30 min. The reaction also included 2.5 µM mix of nine rRNA blocking oligos (Table 6). The mix was made using 100 µM oligos otherwise in equal volumes except for 16S_block oligo, which was added in 1.5x volume. Blocking of the 3' ends of 16S, 23S, and 5S rRNA leads to preferential addition of poly-A tails to mRNA molecules [62]. The reactions were purified with 1.8 vol RNA Clean XP beads and eluted in RNase free water. The rRNA-depleted RNAs with poly-A tails were then used as starting material for the Direct RNA sequencing protocol SQK-RNA002 (Oxford Nanopore Technologies) according to the manufacturer's instructions, except for RNA Control Strand (RCS), which was diluted 1:20 prior to use. About 190 ng of the reverse-transcribed and adapted RNA was sequenced with FLO-MIN106 flow cells using MinKNOW software v19.06.8 or 21.06.0 (Oxford Nanopore Technologies).

RNA modification prediction

Direct RNA sequencing data were basecalled with Guppy v3.0.7 (<https://github.com/nanoporetech>). The sequence reads were aligned to *L. monocytogenes* R015 and ScottA genomes (GenBank assembly accession: R015; GCA_902827145.1 ScottA; GCA_000212455.1) using minimap2 [63] with “-MD -ax splice -uf -k14” options.

Modified base detection analysis was performed using Tombo [38]. Since Tombo does not support multi-read FAST5 format, we converted multi-read FAST5 to single-read FAST5 format using multi_to_single_fast5 command from ont_fast5_api tool (https://github.com/nanoporetech/ont_fast5_api). Tombo resquiggle command was then used to align the raw signal to the *L. monocytogenes* genome. With the Tombo “detect_modifications de_novo” command, de novo non-canonical base detection was performed. Threshold of dampened fraction > 0.9 was used for modified base detection.

Two-sample comparisons for modified base detection were performed using Tombo [38] “level_sample_compare” with default options. For base fast5 files, we used control samples, and for alternate fast5 files we used HPP-treated samples. D-statistic > 0.7 was used for significance threshold.

For motif prediction, we exported the 15 bp length sequence surrounding the modified bases using Tombo “text_output signif_sequence_context”. The fasta sequence was then used for MEME [64] run with “-dna -mod zoops” options.

Annotation of transcripts using direct RNA reads

To predict transcripts from direct RNA reads, we used wf-isoforms (<https://github.com/epi2me-labs/wf-isoforms>) pipeline with ‘-use_pychopper false’ options. To increase the number of the reads that were used in prediction, we combined both control and treated samples’ reads and used it as one input.

Abbreviations

small RNA (sRNA), antisense sRNA (asRNA), Cappable-seq (cap-seq), high pressure processing (HPP), transcription start site (TSS), Oxford Nanopore Technology (ONT), untranslated region (UTR), open reading frame (ORF), circular RNAs (circ-RNA), coding DNA sequence (CDS), transfer-messenger RNAs (tmRNAs)

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and material

The data for this study have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB51821 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB51821>). The data can be also directly downloaded using <https://zenodo.org/record/6463240> (DOI: 10.5281/zenodo.6463240)

Competing interests

The authors have no competing interests to declare.

Authors' contributions

PA conceived and designed the study. AY performed cap-seq and direct RNA-seq preparations. LP organized NGS assays. ICD performed the bioinformatics analyses. LGG was involved in the preparation of the samples. PA and ICD drafted the manuscript. LGG and CUR performed a critical review of the manuscript. All authors have read, commented on, and approved the final manuscript.

Funding

This study was supported by grants from the Academy of Finland to PA (311717, 307856, and 323576).

Acknowledgements

We thank Eeva-Marja Turkki for performing the NGS procedures for this project. We acknowledge the CSC – IT Center for Science, Finland, for computational resources. Daniela Borda and Anca Ioana Nicolau are thanked for useful comments, and the University of Helsinki Language Services for English language revision.

References

1. Bondi M, Messi P, Halami PM, Papadopoulou C, de Niederhausern S. Emerging microbial concerns in food safety and new control measures. *BioMed Res Int.* 2014;2014:251512.
2. Pięłowski M. Food hazards on the European Union market: The data analysis of the rapid alert system for food and feed. *Food Sci Nutr.* 2020;8:1603.
3. Bucur FI, Grigore-Gurgu L, Crauwels P, Riedel CU, Nicolau AI. Resistance of *Listeria monocytogenes* to stress conditions encountered in food and food processing Environments. *Front Microbiol.* 2018;9:2700.
4. Nyarko EB, Donnelly CW. *Listeria monocytogenes*: Strain heterogeneity, methods, and challenges of subtyping. *J Food Sci.* 2015;80:M2868-2878.
5. Kurpas M, Wieczorek K, Osek J. Ready-to-eat meat products as a source of *Listeria monocytogenes*. *J Vet Res.* 2018;62:49.
6. Buchanan RL, Gorris LGM, Hayman MM, Jackson TC, Whiting RC. A review of *Listeria monocytogenes*: An update on outbreaks, virulence, dose-response, ecology, and risk assessments. *Food Control.* 2017;75:1–13.
7. Chan YC, Wiedmann M. Physiology and genetics of *Listeria monocytogenes* survival and growth at cold temperatures. *Crit Rev Food Sci Nutr.* 2009;49:237–53.
8. Yamamoto K. Food processing by high hydrostatic pressure. *Biosci Biotechnol Biochem.* 2017;81:672–9.
9. Nikparvar B, Subires A, Capellas M, Hernandez-Herrero M, Crauwels P, Riedel CU, et al. A diffusion model to quantify membrane repair process in *Listeria monocytogenes* exposed to high pressure processing based on fluorescence microscopy data. *Front Microbiol.* 2021;12:598739.
10. Nikparvar B, Andreevskaya M, Duru IC, Bucur FI, Grigore-Gurgu L, Borda D, et al. Analysis of temporal gene regulation of *Listeria monocytogenes* revealed distinct regulatory response modes after exposure to high pressure processing. *BMC Genomics.* 2021;22:266.
11. Duru IC, Bucur FI, Andreevskaya M, Ylinen A, Crauwels P, Grigore-Gurgu L, et al. The complete genome sequence of *Listeria monocytogenes* strain S2542 and expression of selected genes under high-pressure processing. *BMC Res Notes.* 2021;14:137.
12. Duru IC, Andreevskaya M, Laine P, Rode TM, Ylinen A, Løvdal T, et al. Genomic characterization of the most barotolerant *Listeria monocytogenes* R015 strain compared to reference strains used to evaluate food high pressure processing. *BMC Genomics.* 2020;21:455.
13. Duru IC, Bucur FI, Andreevskaya M, Nikparvar B, Ylinen A, Grigore-Gurgu L, et al. High-pressure processing-induced transcriptome response during recovery of *Listeria monocytogenes*. *BMC Genomics.* 2021;22:117.

14. Liu Y, Orsi RH, Boor KJ, Wiedmann M, Guariglia-Oropeza V. Home Alone: Elimination of All but One Alternative Sigma Factor in *Listeria monocytogenes* Allows Prediction of New Roles for σ B. *Front Microbiol.* 2017;8.
15. Kazmierczak MJ, Mithoe SC, Boor KJ, Wiedmann M. *Listeria monocytogenes* sigma B regulates stress response and virulence functions. *J Bacteriol.* 2003;185:5722–34.
16. Sharma CM, Vogel J. Differential RNA-seq: the approach behind and the biological insight gained. *Curr Opin Microbiol.* 2014;19:97–105.
17. Ettwiller L, Buswell J, Yigit E, Schildkraut I. A novel enrichment strategy reveals unprecedented number of novel transcription start sites at single base resolution in a model prokaryote and the gut microbiome. *BMC Genomics.* 2016;17:199.
18. Ryan D, Jenniches L, Reichardt S, Barquist L, Westermann AJ. A high-resolution transcriptome map identifies small RNA regulation of metabolism in the gut microbe *Bacteroides thetaiotaomicron*. *Nat Commun.* 2020;11:3557.
19. Fuchs M, Lamm-Schmidt V, Sulzer J, Ponath F, Jenniches L, Kirk JA, et al. An RNA-centric global view of *Clostridioides difficile* reveals broad activity of Hfq in a clinically important gram-positive bacterium. *Proc Natl Acad Sci U S A.* 2021;118:e2103579118.
20. Bischler T, Tan HS, Nieselt K, Sharma CM. Differential RNA-seq (dRNA-seq) for annotation of transcriptional start sites and small RNAs in *Helicobacter pylori*. *Methods San Diego Calif.* 2015;86:89–101.
21. Wurtzel O, Sesto N, Mellin JR, Karunker I, Edelheit S, Bécavin C, et al. Comparative transcriptomics of pathogenic and non-pathogenic *Listeria* species. *Mol Syst Biol.* 2012;8:583.
22. Sass AM, Van Acker H, Förstner KU, Van Nieuwerburgh F, Deforce D, Vogel J, et al. Genome-wide transcription start site profiling in biofilm-grown *Burkholderia cenocepacia* J2315. *BMC Genomics.* 2015;16:775.
23. Knoop V. When you can't trust the DNA: RNA editing changes transcript sequences. *Cell Mol Life Sci CMLS.* 2011;68:567–86.
24. Bar-Yaacov D, Mordret E, Towers R, Biniashvili T, Soyris C, Schwartz S, et al. RNA editing in bacteria recodes multiple proteins and regulates an evolutionarily conserved toxin-antitoxin system. *Genome Res.* 2017;27:1696.
25. Nie W, Wang S, He R, Xu Q, Wang P, Wu Y, et al. A-to-I RNA editing in bacteria increases pathogenicity and tolerance to oxidative stress. *PLoS Pathog.* 2020;16.
26. Andreevskaya M, Johansson P, Jääskeläinen E, Rämö T, Ritari J, Paulin L, et al. *Lactobacillus oligofermentans* glucose, ribose and xylose transcriptomes show higher similarity between glucose and xylose catabolism-induced responses in the early exponential growth phase. *BMC Genomics.* 2016;17:539.
27. Duru IC, Ylinen A, Belanov S, Pulido AA, Paulin L, Auvinen P. Transcriptomic time-series analysis of cold- and heat-shock response in psychrotrophic lactic acid bacteria. *BMC Genomics.* 2021;22:28.
28. Leger A, Amaral PP, Pandolfini L, Capitanchik C, Capraro F, Miano V, et al. RNA modifications detection by comparative Nanopore direct RNA sequencing. *Nat Commun.* 2021;12:7198.
29. Frye M, Harada BT, Behm M, He C. RNA modifications modulate gene expression during development. *Science.* 2018;361:1346–9.

30. Zhao X, Yang Y, Sun B-F, Shi Y, Yang X, Xiao W, et al. FTO-dependent demethylation of N6-methyladenosine regulates mRNA splicing and is required for adipogenesis. *Cell Res.* 2014;24:1403–19.
31. Delaunay S, Frye M. RNA modifications regulating cell fate in cancer. *Nat Cell Biol.* 2019;21:552–9.
32. Roach NP, Sadowski N, Alessi AF, Timp W, Taylor J, Kim JK. The full-length transcriptome of *C. elegans* using direct RNA sequencing. *Genome Res.* 2020;30:299.
33. Parker MT, Knop K, Sherwood AV, Schurch NJ, Mackinnon K, Gould PD, et al. Nanopore direct RNA sequencing maps the complexity of *Arabidopsis* mRNA processing and m6A modification. *eLife.* 2020;9:e49658.
34. Price AM, Hayer KE, McIntyre ABR, Gokhale NS, Abebe JS, Della Fera AN, et al. Direct RNA sequencing reveals m6A modifications on adenovirus RNA are necessary for efficient splicing. *Nat Commun.* 2020;11:6016.
35. Kim D, Lee J-Y, Yang J-S, Kim JW, Kim VN, Chang H. The Architecture of SARS-CoV-2 Transcriptome. *Cell.* 2020;181:914–921.e10.
36. Sergeeva OV, Bogdanov AA, Sergiev PV. What do we know about ribosomal RNA methylation in *Escherichia coli*? *Biochimie.* 2015;117:110–8.
37. Yu S-H, Vogel J, Förstner KU. ANNOgesic: a Swiss army knife for the RNA-seq based annotation of bacterial/archaeal genomes. *GigaScience.* 2018;7.
38. Stoiber M, Quick J, Egan R, Eun Lee J, Celniker S, Neely RK, et al. De novo Identification of DNA Modifications Enabled by Genome-Guided Nanopore Signal Processing. *bioRxiv.* 2017;:094672.
39. Pawluk A, Davidson AR, Maxwell KL. Anti-CRISPR: discovery, mechanism and function. *Nat Rev Microbiol.* 2018;16:12–7.
40. Rauch BJ, Silvis MR, Hultquist JF, Waters CS, McGregor MJ, Krogan NJ, et al. Inhibition of CRISPR-Cas9 with Bacteriophage Proteins. *Cell.* 2017;168:150–158.e10.
41. Cerutti F, Mallet L, Painset A, Hoede C, Moisan A, Bécavin C, et al. Unraveling the evolution and coevolution of small regulatory RNAs and coding genes in *Listeria*. *BMC Genomics.* 2017;18:882.
42. Kim D, Hong JS-J, Qiu Y, Nagarajan H, Seo J-H, Cho B-K, et al. Comparative Analysis of Regulatory Elements between *Escherichia coli* and *Klebsiella pneumoniae* by Genome-Wide Transcription Start Site Profiling. *PLoS Genet.* 2012;8.
43. Voigt K, Sharma CM, Mitschke J, Lambrecht SJ, Voß B, Hess WR, et al. Comparative transcriptomics of two environmentally relevant cyanobacteria reveals unexpected transcriptome diversity. *ISME J.* 2014;8:2056–68.
44. Mitschke J, Georg J, Scholz I, Sharma CM, Dienst D, Bantscheff J, et al. An experimentally anchored map of transcriptional start sites in the model cyanobacterium *Synechocystis* sp. PCC6803. *Proc Natl Acad Sci U S A.* 2011;108:2124–9.
45. Perriman R, Ares M, Jr. Circular mRNA can direct translation of extremely long repeating-sequence proteins in vivo. *RNA.* 1998;4:1047.
46. Gourse RL, Ross W, Gaal T. UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition. *Mol Microbiol.* 2000;37:687–95.
47. Harley CB, Reynolds RP. Analysis of *E. coli* promoter sequences. *Nucleic Acids Res.* 1987;15:2343.

48. Djordjevic M. Redefining *Escherichia coli* σ 70 Promoter Elements: –15 Motif as a Complement of the –10 Motif. *J Bacteriol.* 2011;193:6305.
49. Vakulskas CA, Brutinel ED, Yahr TL. ExsA recruits RNA polymerase to an extended –10 promoter by contacting region 4.2 of sigma-70. *J Bacteriol.* 2010;192:3597–607.
50. Kline BC, McKay SL, Tang WW, Portnoy DA. The *Listeria monocytogenes* hibernation-promoting factor is required for the formation of 100s ribosomes, optimal fitness, and pathogenesis. *J Bacteriol.* 2015;197:581–91.
51. Mizuno CM, Guyomar C, Roux S, Lavigne R, Rodriguez-Valera F, Sullivan MB, et al. Numerous cultivated and uncultivated viruses encode ribosomal proteins. *Nat Commun.* 2019;10.
52. Antoine L, Bahena-Ceron R, Bunwaree HD, Gobry M, Loegler V, Romby P, et al. RNA Modifications in Pathogenic Bacteria: Impact on Host Adaptation and Virulence. *Genes.* 2021;12.
53. Förstner KU, Vogel J, Sharma CM. READemption—a tool for the computational analysis of deep-sequencing-based transcriptome data. *Bioinforma Oxf Engl.* 2014;30:3421–3.
54. Hoffmann S, Otto C, Kurtz S, Sharma CM, Khaitovich P, Vogel J, et al. fast mapping of short sequences with mismatches, insertions and deletions using index structures. *PLoS Comput Biol.* 2009;5.
55. Dugar G, Herbig A, Förstner KU, Heidrich N, Reinhardt R, Nieselt K, et al. High-resolution transcriptome maps reveal strain-specific regulatory features of multiple *Campylobacter jejuni* isolates. *PLoS Genet.* 2013;9:e1003495.
56. Harris RS. Improved pairwise alignment of genomic DNA. The Pennsylvania State University; 2007.
57. Kent WJ, Baertsch R, Hinrichs A, Miller W, Haussler D. Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. *Proc Natl Acad Sci U S A.* 2003;100:11484–9.
58. Zhao H, Sun Z, Wang J, Huang H, Kocher J-P, Wang L. CrossMap: a versatile tool for coordinate conversion between genome assemblies. *Bioinforma Oxf Engl.* 2014;30:1006–7.
59. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012;9:357–9.
60. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinforma Oxf Engl.* 2009;25:2078–9.
61. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics.* 2011;27:2987.
62. Wangsanuwat C, Heom KA, Liu E, O'Malley MA, Dey SS. Efficient and cost-effective bacterial mRNA sequencing from low input samples through ribosomal RNA depletion. *BMC Genomics.* 2020;21:717.
63. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinforma Oxf Engl.* 2018;34:3094–100.
64. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol.* 1994;2:28–36.

Figures

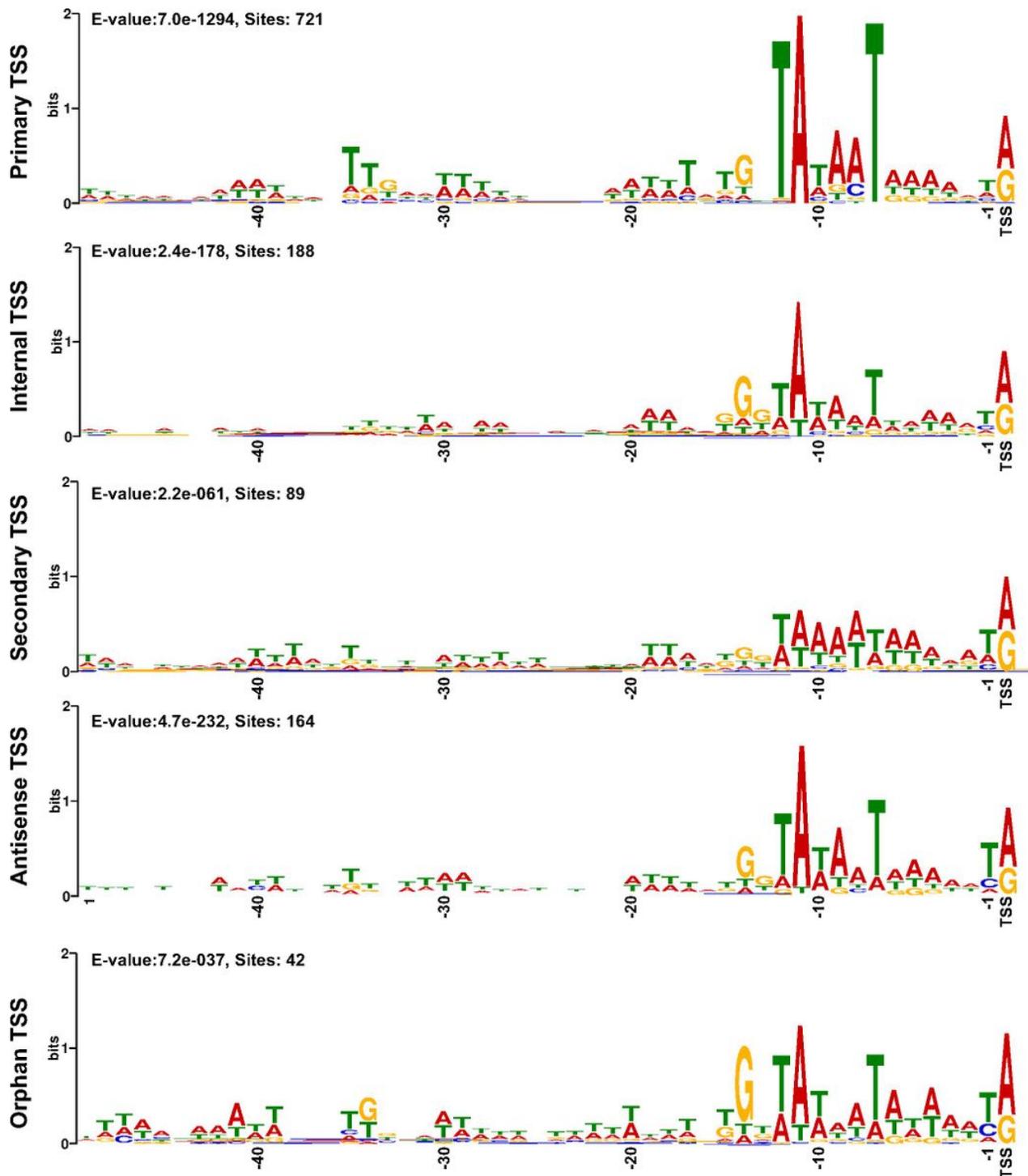


Figure 1

Predicted promoter motifs. Each row shows the lowest E-value sequence motif identified by a MEME suite tool on a window from -50 bp upstream to the TSS. The X axis represents the upstream position from TSS. In descending order: primary TSS, internal TSS, secondary TSS, antisense TSS, orphan TSS. E-value and the number of sites contributing to the construction of the motif can be found on the upper right corner of each motif logo.

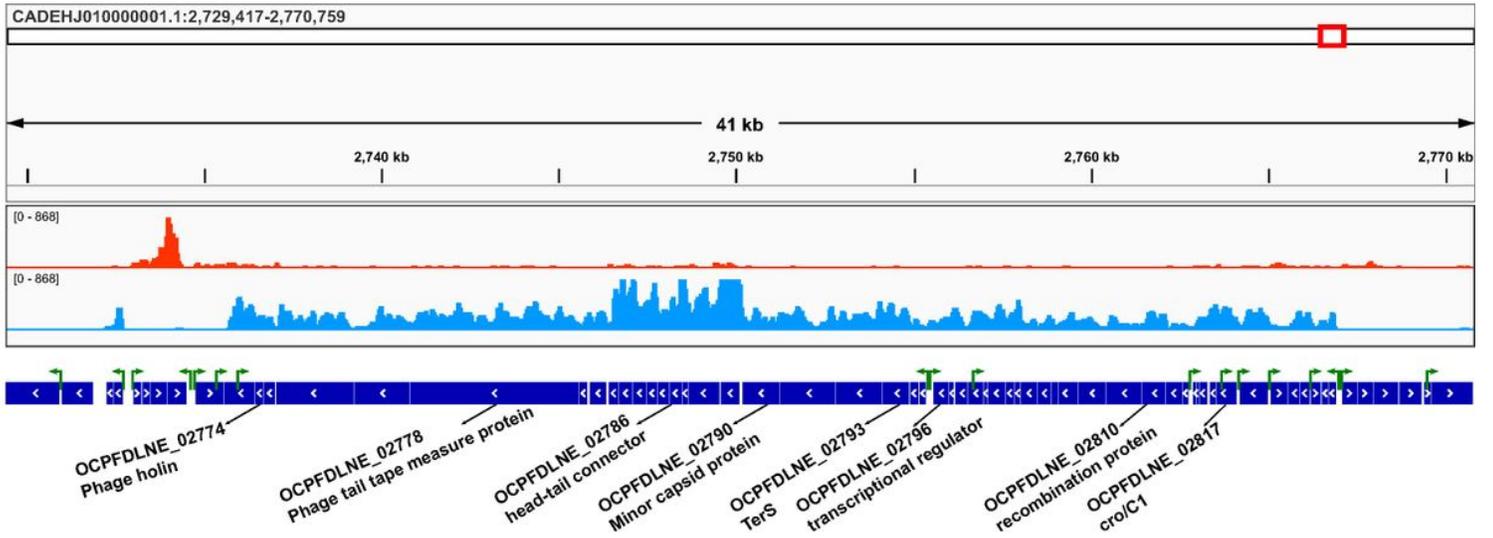


Figure 2

Visualization of Prophage 5 (2729417-2770759 bp) and TSSs predicted. Red bars represent forward RNA-seq reads coverage. Light blue bars represent reverse RNA-seq reads coverage. Dark blue boxes represent protein coding genes (direction of the gene is shown with a white arrow). Green arrows represent TSSs.

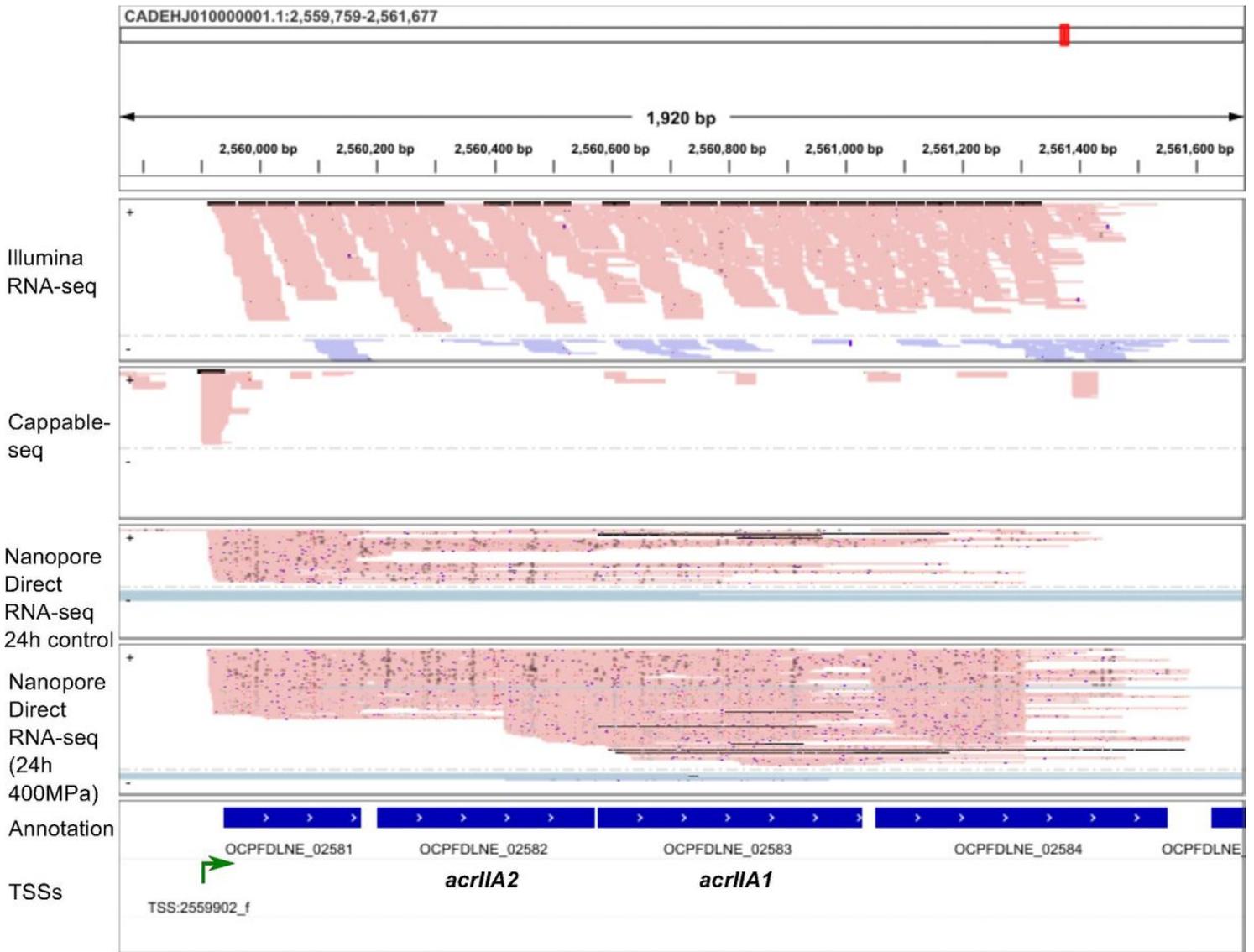
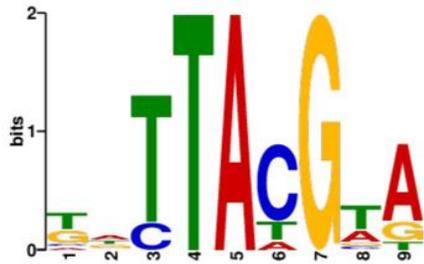


Figure 3

Visualization of anti-CRISPR genes in prophage5 on *L. monocytogenes* RO15 region 2.560 Mbp - 2.561 Mbp. Six rows of data are shown, in descending order: combined Illumina RNA-seq reads, combined cap-seq reads, Nanopore direct RNA-seq for control sample reads (R312), Nanopore direct RNA-seq for 400 MPa treated sample, (R173) reads, Genbank annotation, TSS prediction annotation. The region includes anti-CRISPR genes, marked in the figure as *acrIIA2* and *acrIIA1*. One forward TSS was predicted at the beginning of gene OCPFDLNE_02581. The reads were grouped by read strand in IGV, in each rectangle forward reads are located in the upper part and reverse reads in the lower part. Green arrow indicates TSS.

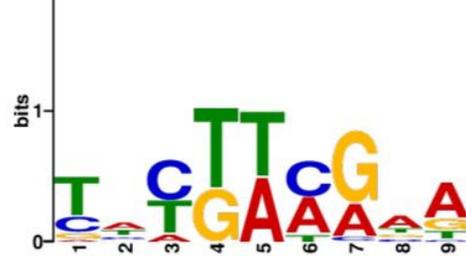
a) RO15 enriched motif identified by MEME

E-value: 2.8e-012, Site Count: 11

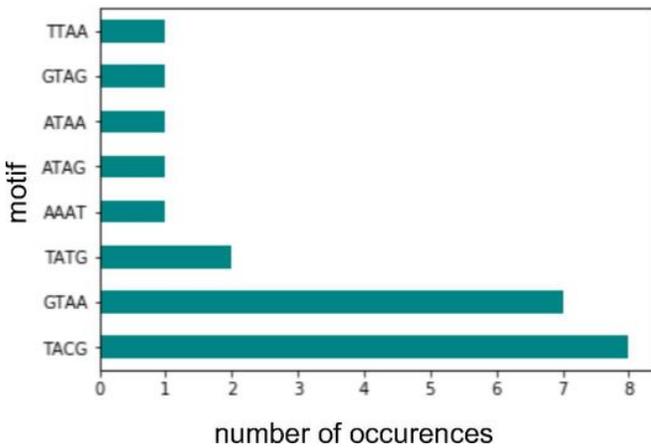


b) ScottA enriched motif identified by MEME

E-value: 1.8e-011, Site Count: 20



c) RO15 count of 4mer around all 22 editing site



d) ScottA count of 4mer around all 25 editing site

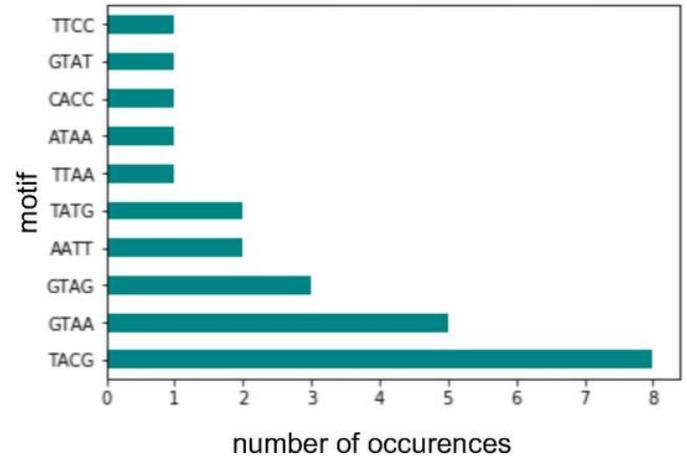
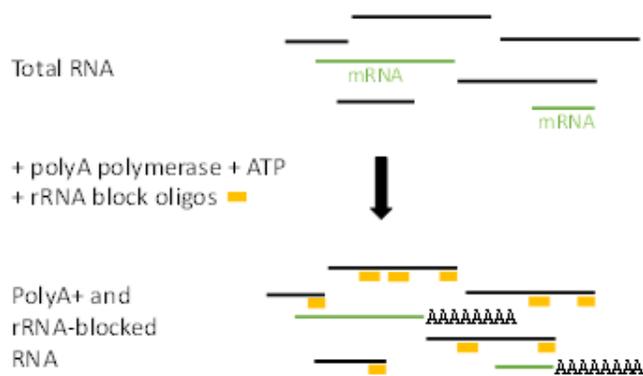


Figure 4

Motif around RNA editing sites based on short read data. Enriched motif around editing sites predicted by MEME was shown for both a) RO15 and b) ScottA. The count of four-mer around all editing sites was shown by bar plot for c) RO15 and d) ScottA.

Addition of RNA block oligos and polyA tail



Library preparation for Direct RNA sequencing

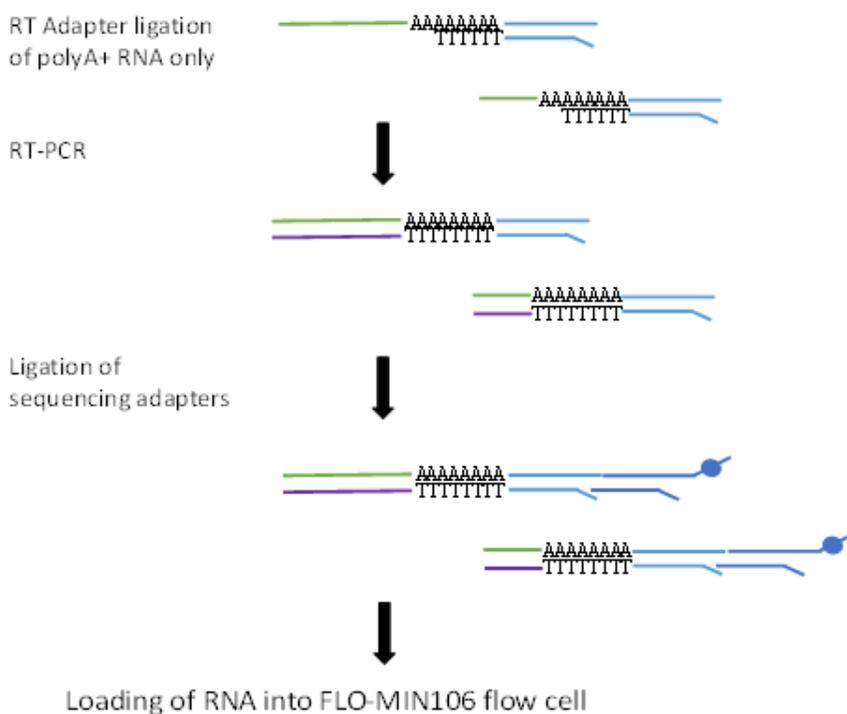


Figure 5

Customized direct RNA-seq protocol to sequence bacterial RNA reads by adding polyA tail.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigureS1.pdf](#)
- [FigureS2.pdf](#)

- [FigureS3.pdf](#)
- [FigureS4.pdf](#)
- [FigureS5.pdf](#)
- [FigureS6.pdf](#)
- [FigureS7.pdf](#)
- [FigureS8.pdf](#)
- [FigureS9.pdf](#)
- [FigureS10.pdf](#)
- [FigureS11.pdf](#)
- [FigureS12.pdf](#)
- [FigureS13.pdf](#)
- [FigureS14.pdf](#)
- [FigureS15.pdf](#)
- [FigureS16.pdf](#)
- [FigureS17.pdf](#)
- [TableS1.xlsx](#)
- [TableS2.xlsx](#)
- [TableS3.xlsx](#)
- [TableS4.xlsx](#)
- [TableS5.xlsx](#)
- [TableS6.xlsx](#)
- [TableS7.xlsx](#)
- [TableS8.xlsx](#)
- [TableS9.xlsx](#)
- [TableS10.xlsx](#)
- [TableS11.xlsx](#)
- [SupplementalFile1.txt](#)
- [SupplementalFile2.txt](#)