

Research article

Open Access

Deep RNA sequencing of *L. monocytogenes* reveals overlapping and extensive stationary phase and sigma B-dependent transcriptomes, including multiple highly transcribed noncoding RNAs

Haley F Oliver^{†1}, Renato H Orsi^{†1}, Lalit Ponnala², Uri Keich^{3,4}, Wei Wang⁵, Qi Sun², Samuel W Cartinhour^{6,7}, Melanie J Filiatrault^{6,7}, Martin Wiedmann¹ and Kathryn J Boor^{*1}

Address: ¹Department of Food Science, Cornell University, Ithaca, NY, USA, ²Computational Biology Service Unit, Cornell University, Ithaca, NY, USA, ³Department of Computer Science, Cornell University, Ithaca, NY, USA, ⁴School of Mathematics and Statistics, University of Sydney, NSW, Australia, ⁵Life Sciences Core Laboratories Center, Cornell University, Ithaca, NY, USA, ⁶United States Department of Agriculture-Agricultural Research Service, Robert W. Holley Center for Agriculture and Health, Ithaca, NY, USA and ⁷Department of Plant Pathology and Plant-Microbe Biology, Cornell University, Ithaca, NY, USA

Email: Haley F Oliver - haf9@cornell.edu; Renato H Orsi - rho2@cornell.edu; Lalit Ponnala - ponnala@cac.cornell.edu; Uri Keich - uri@maths.usyd.edu.au; Wei Wang - ww77@cornell.edu; Qi Sun - qisun@tc.cornell.edu; Samuel W Cartinhour - sam.cartinhour@ars.usda.gov; Melanie J Filiatrault - melanie.filiatrault@ars.usda.gov; Martin Wiedmann - mw16@cornell.edu; Kathryn J Boor* - kjb4@cornell.edu

* Corresponding author †Equal contributors

Published: 30 December 2009

Received: 1 June 2009

BMC Genomics 2009, 10:641 doi:10.1186/1471-2164-10-641

Accepted: 30 December 2009

This article is available from: <http://www.biomedcentral.com/1471-2164/10/641>

© 2009 Oliver et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Identification of specific genes and gene expression patterns important for bacterial survival, transmission and pathogenesis is critically needed to enable development of more effective pathogen control strategies. The stationary phase stress response transcriptome, including many σ^B -dependent genes, was defined for the human bacterial pathogen *Listeria monocytogenes* using RNA sequencing (RNA-Seq) with the Illumina Genome Analyzer. Specifically, bacterial transcriptomes were compared between stationary phase cells of *L. monocytogenes* I0403S and an otherwise isogenic $\Delta sigB$ mutant, which does not express the alternative σ factor σ^B , a major regulator of genes contributing to stress response, including stresses encountered upon entry into stationary phase.

Results: Overall, 83% of all *L. monocytogenes* genes were transcribed in stationary phase cells; 42% of currently annotated *L. monocytogenes* genes showed medium to high transcript levels under these conditions. A total of 96 genes had significantly higher transcript levels in I0403S than in $\Delta sigB$, indicating σ^B -dependent transcription of these genes. RNA-Seq analyses indicate that a total of 67 noncoding RNA molecules (ncRNAs) are transcribed in stationary phase *L. monocytogenes*, including 7 previously unrecognized putative ncRNAs. Application of a dynamically trained Hidden Markov Model, in combination with RNA-Seq data, identified 65 putative σ^B promoters upstream of 82 of the 96 σ^B -dependent genes and upstream of the one σ^B -dependent ncRNA. The RNA-Seq data also enabled annotation of putative operons as well as visualization of 5'- and 3'-UTR regions.

Conclusions: The results from these studies provide powerful evidence that RNA-Seq data combined with appropriate bioinformatics tools allow quantitative characterization of prokaryotic transcriptomes, thus providing exciting new strategies for exploring transcriptional regulatory networks in bacteria.

See minireview <http://jbiol.com/content/8/12/107>.

Background

The development of powerful new DNA sequencing technologies has yielded new tools with the potential for dramatically revolutionizing scientific approaches to biological questions [1]. These new technologies can be used for a variety of applications, including genome sequencing, identification of DNA-methylation sites, population studies, chromatin precipitation (CHIP-Seq), and transcriptome studies (RNA-Seq). For RNA-Seq, cDNA is generated from an mRNA-enriched total RNA preparation and sequenced using high-throughput technology. Here, we used the Illumina Genome Analyzer to characterize the transcriptome of stationary phase *Listeria monocytogenes* 10403S and its isogenic $\Delta sigB$ mutant, which lacks the general stress response sigma factor, σ^B .

L. monocytogenes, a Gram-positive foodborne pathogen of the Firmicutes family, is the etiological agent of the disease known as listeriosis. As 20% of listeriosis cases result in death in humans, with an estimated annual human death toll of ~ 500 in the US alone [2], this disease is a considerable public health concern. As a foodborne pathogen (with 99% of human illnesses caused by a foodborne route of infection [2]), this bacterium also presents challenging food safety concerns due to its ability to survive and grow under many conditions that are typically applied to control bacterial populations in foods, such as low pH, low temperature and high salt conditions [3-5]. The alternative general stress response sigma factor, σ^B , is an essential component of a regulatory mechanism that contributes to the ability of *L. monocytogenes* to respond to and survive exposure to harsh environmental conditions [6].

Sigma factors are dissociable subunits of prokaryotic RNA polymerase responsible for enzyme recognition of a conserved DNA sequence encoding a transcriptional promoter site. Promoter recognition specificities of bacterial RNA polymerase are determined by the transient association of an appropriate sigma factor with core polymerase in response to conditions affecting the cell [7]. The regulation of a single alternative sigma factor can include hundreds of transcriptional units, thus sigma factors provide an effective mechanism for simultaneously regulating large numbers of genes under appropriate conditions [7]. Critical phenotypic functions regulated by alternative sigma factors range from bacterial sporulation [8] to stress response systems [6,9].

Through microarray analyses, the σ^B regulon in *L. monocytogenes* has been reported to encompass more than 200 genes, including both virulence and stress response genes, many of them up-regulated upon entry into stationary phase [10-12]. However, interpretation of microarray analyses is dependent on the quality of existing genome annotations, which are rarely experimentally verified. Fur-

ther, transcripts that do not correspond to annotated features (e.g., noncoding RNA transcripts) cannot be identified. In addition, the utility of microarrays is limited by the genomic variation that exists among bacterial strains (i.e., ideally, a unique microarray should be constructed for each strain to be analyzed) and by technical biases such as cross-hybridization. Hence, microarray data can be difficult to analyze and occasionally, misleading [13,14]. Although interpretation of RNA-Seq data also relies on the availability of a genome sequence, it is probe- and annotation-independent and therefore, is free of cross-hybridization and low-hybridization biases, hence enabling genome-wide identification of all transcripts, including small noncoding RNAs (ncRNAs). Moreover, because RNA-Seq technology can generate multiple reads corresponding to each transcribed nucleotide on the genome, it is usually possible to identify 5' and 3' transcript ends with high resolution [15]. Therefore, in combination with bioinformatics tools, RNA-Seq data can be used to identify transcriptional promoters and terminators. We used *L. monocytogenes* as a model system to explore application of RNA-Seq for the dual purposes of genome-wide transcriptome characterization in a bacterial pathogen and comprehensive quantification of target gene expression for the alternative sigma factor, σ^B .

Results

RNA-Seq provided comprehensive coverage of the *L. monocytogenes* transcriptome

RNA-Seq analyses were performed on two independent replicate RNA samples collected from both the *L. monocytogenes* strain 10403S and an otherwise isogenic $\Delta sigB$ mutant (FSL A1-254) that had been grown to stationary phase. cDNA was generated from mRNA-enriched total RNA preparations from each strain and sequenced using the Illumina Genome Analyzer to yield a total number of reads for each sample ranging from 3,300,716 to 5,236,748 (Table 1). As the 10403S genome has not been completely closed, the sequence reads were aligned to a 10403S pseudochromosome that was created for this study using the completely closed genome of the *L. monocytogenes* strain EGD-e (accession no. AL591824) as a reference (see Material and Methods for details). The total number of reads matching regions other than rRNA and tRNA ranged from 451,548 to 683,746, yielding between $5 \times$ and $7.6 \times$ coverage of the pseudogenome. Between 87.3% and 92.1% of the reads in a given RNA-Seq run matched uniquely to the 10403S pseudochromosome and thus were used in subsequent analyses. Reads that did not match the 10403S pseudochromosome (i.e., reads that showed > 2 mismatches to the pseudochromosome) represented between 6.7% and 12.6% of the reads sequenced; another 0.1% to 0.7% of the reads matched to at least two different locations on the pseudochromosome and, therefore, were removed before further analyses. Reads identified as "matching two locations" did not

Table 1: Summary of RNA-Seq coverage data

Statistics	10403S replicate 1	10403S replicate 2	sigB replicate 1	sigB replicate 2
Reads that aligned uniquely with no mismatches (U_0)	2,290,717	3,111,726	2,320,447	3,866,492
Reads that aligned uniquely with 1 mismatch (U_1)	632,173	470,865	544,932	745,360
Reads that aligned uniquely with 2 mismatches (U_2)	234,886	110,882	173,903	181,684
$U_{SUM} = U_0 + U_1 + U_2$	3,157,776	3,693,473	3,039,282	4,793,536
Reads that aligned at more than one location (reads not used; R)	23,485	4,832	38,489	16,103
Reads that did not align to the pseudochromosome (NM)	299,034	533,462	222,945	427,109
Total number of reads in the sample (Total = $U_{SUM} + R + NM$)	3,480,295	4,231,767	3,300,716	5,236,748
Percentage of unique alignments, i.e. $100*(U_{SUM})/Total$	90.73	87.28	92.08	91.54
Reads that aligned to the 16S rRNA gene (16S)	490,381	482,845	434,263	760,863
Reads that aligned to the 23S rRNA gene (23S)	2,160,538	1,860,817	2,436,325	3,138,329
Reads that aligned to the 16S and 23S rRNA genes (16S + 23S)	2,650,919	2,919,170	2,295,080	3,899,192
Percentage of all reads that aligned to 16S and 23S rRNA genes	83.9	79	75.5	81.3
$U_{TOTAL} = U_{SUM} - (16S + 23S)$	506,857	774,303	744,202	894,344
Normalization factor ($f_{norm} = 894,344/U_{TOTAL}$) ^a	1.765	1.155	1.202	1

^aThis indicates the factor that was used for normalization of replicates

include those matching rRNA genes as the 10403S pseudochromosome created for this study was designed with only one unique rRNA gene sequence.

To allow for quantitative comparisons among genes and runs, the coverage for each run was normalized for the total number of reads in each run and for gene size. The normalized data are presented as the Gene Expression Index (GEI), which is expressed as the number of reads per 100 bases [16]. Although *in silico* analyses suggested that the sequencibility (i.e., the portion of the pseudochromosome that could yield unique 32 nt reads) of the 10403S pseudochromosome was 99.6% (Additional file 1: Sequencibility text file), approximately 77.5% of the genome was covered by reads from at least one of the four runs, suggesting that more than 20% of the genome is not transcribed or is transcribed at low levels.

RNA-Seq coverage correlated with qRT-PCR transcript levels indicating that RNA-Seq data are quantitative

We evaluated whether average GEI for specific genes correlated with transcript levels that had been measured using TaqMan qRT-PCR, the current gold standard for quantification of mRNA [17]. Based on transcript levels for 9 and 5 genes in 10403S and $\Delta sigB$, respectively, log transformed average GEI and log transformed TaqMan qRT-PCR absolute copy numbers were correlated (p -value < 0.001; adj. $R^2 = 0.83$; Figure 1; Additional file 2: RNA-Seq average GEI and TaqMan qRT-PCR absolute copy number of select genes), supporting that RNA-Seq provides reliable quantitative estimates of transcript levels in *L. monocytogenes*. RNA-Seq was previously reported to provide quantitative data on transcript levels in yeast [15], and more recently, in *Burkholderia cenocepacia* [16], thus,

our findings extend this important correlation to a new prokaryotic system.

Stationary phase *L. monocytogenes* transcribed at least 83% of annotated genes

Among the 2888 annotated coding sequences (CDS) in the 10403S pseudochromosome, 2417 (83.7%) showed

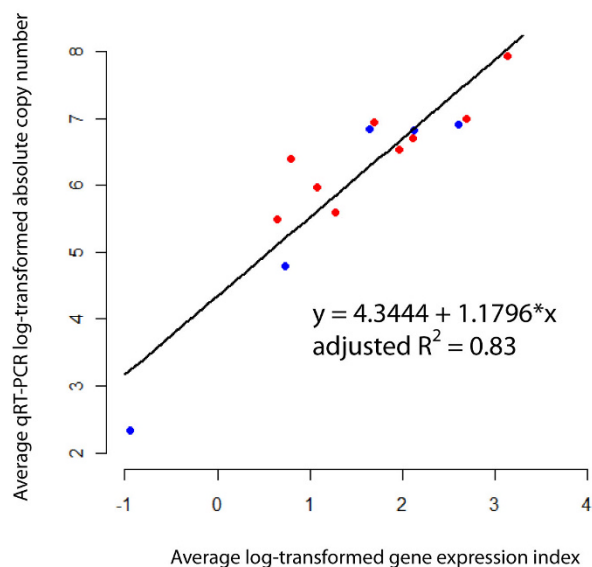


Figure 1
Correlation between qRT-PCR and RNA-Seq. Correlation between qRT-PCR and RNA-Seq data for selected genes in *L. monocytogenes* 10403S (red) and the $\Delta sigB$ strain (blue). The selected genes are: *ctc*, *gadA*, *gap*, *opuCA*, *rpoB* (qRT-PCR data from both strains were available for these 5 genes), *flaA*, *inlA*, *plcA* and *sigB* (only qRT-PCR data from 10403S were available for these 4 genes).

an average GEI ≥ 0.7 in 10403S (average of two biological replicates) suggesting that at least 83% of the annotated *L. monocytogenes* genes are transcribed in stationary phase (Additional file 3: Cumulative frequency of average GEI in *L. monocytogenes* 10403S; see Materials and Methods for calculation of coverage, rationale for defining transcribed genes, and criteria for classifying transcript levels as low, medium or high). Of these 2417 genes, 654 (22%) had high transcript levels, 586 (20.0%) had medium transcript levels, and 1177 (41.0%) had low transcript levels. A total of 471 genes (17%) had GEI < 0.7 and were considered "not transcribed". RNA-Seq data allowed visual examination of transcript units, aiding in identification of genes that are transcribed monocistronically or as part of an operon (Figure 2). A total of 355 transcription units appeared to represent operons; these units were identified and annotated (Additional file 4: Access database). A total of 1107 (38.3%) of the annotated 10403S CDS were located in these putative operons. Further experimental data are necessary to validate our predictions of transcription unit structure as some genes may have rho-dependent terminators that were not identified in this study and, therefore, they may be transcribed monocistronically despite the observation of GEI similar to those of their neighboring genes.

The three genes with the highest average GEI in 10403S all encoded predicted ncRNAs, including tmRNA, 6S and LhrA (Table 2). The annotated CDS (as annotated in EGD-e [18]) with the highest average GEI were lmo2257, fri, and lmo1847, which encode a hypothetical CDS, iron-binding ferritin, and an ABC transporter, respectively.

Other genes with well defined functions and high average GEI include flaA, which encodes a flagellin protein, sod, which encodes a superoxide dismutase involved in detoxification, and cspB and cspL, which encode cold-shock proteins involved in adaptation to atypical conditions (Table 2).

Both positive and negative associations were observed between GEI and the TIGR classification of sets of genes to physiological role categories <http://cmr.jcvi.org/cgi-bin/CMR/RoleIds.cgi> (Table 3). For example, genes involved in protein synthesis and protein fate showed higher average GEI in stationary phase 10403S as compared to genes involved in other functions, while genes involved in viral functions and amino acid biosynthesis were significantly associated with low average GEI in 10403S. Moreover, a positive significant association was observed between codon bias and the average GEI in 10403S (p -value < 0.001 ; linear regression analysis).

Identification and annotation of noncoding RNAs (ncRNAs)

Overall, we identified 67 ncRNAs (Additional file 5: ncRNAs identified by RNA-Seq) that showed average GEI ≥ 0.7 in 10403S, indicating that these ncRNAs are transcribed in stationary phase *L. monocytogenes* (see Materials and Methods for more details on ncRNA annotation). Among the 67 ncRNAs identified as transcribed in the present study, 60 matched ncRNAs previously described in *L. monocytogenes* (Additional file 5: ncRNAs identified by RNA-Seq) [19-22]. These 60 ncRNAs included 6S RNA, tmRNA, several S-box RNA and T-box leader RNA mole-

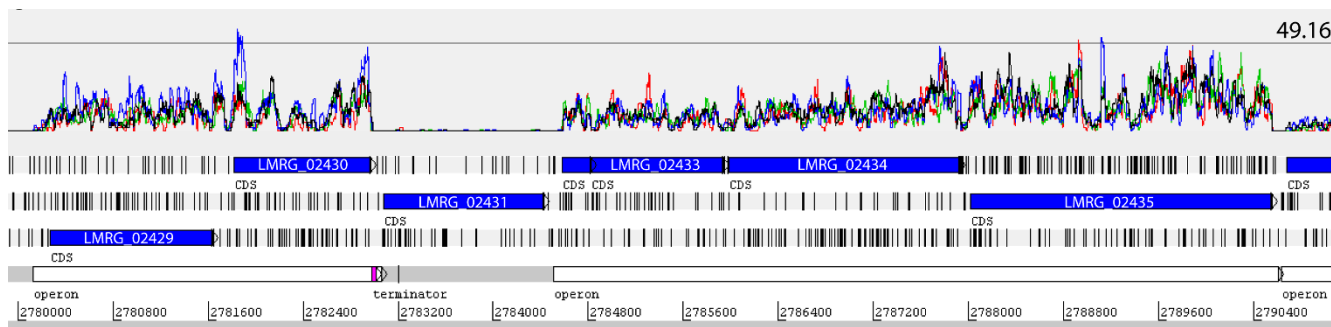


Figure 2

View of RNA-Seq data using the Artemis genome browser. This region of the 10403S chromosome includes six coding genes, i.e. LMRG_02429 to LMRG_02435, and the 5' end of LMRG_02436; genes are represented as blue arrows. The top part of the figure shows normalized RNA-Seq coverage (i.e. the number of reads that match an annotated gene after normalization across runs) with red and blue lines representing the two 10403S replicates and the green and black lines representing the $\Delta sigB$ strain. The horizontal line indicates a normalized RNA-Seq coverage of 49.16 reads. The middle part of the figure shows the three positive frames of translation with the coding regions and vertical black bars representing stop codons. The last line shows putative operons (white bars), a terminator (purple bar) downstream of LMRG_02430 and the chromosome coordinates. Notice the difference in coverage between LMRG_02431 (downstream of the terminator) and the other genes. All genes in the figure have sequencibility of 100% (See Additional file 1: Sequencibility text file for a complete sequencibility plot).

Table 2: Genes with highest GEI

Locus	Gene name ^a	EGD-e locus ^b	Description	I0403S Average GEI ^c
LMRG_04519	<i>ssrA</i>	NL	transfer-messenger RNA (tmRNA)	8566.2
LMRG_04503	<i>ssrS</i>	NL	6S RNA	7921.4
Noncoding	<i>lhrA</i>	NL	Hfq-binding RNA	4532.3
Noncoding	<i>sbrE</i>	NL	putative ncRNA	2359.9
LMRG_01574 ^d	<i>lmo2257</i>	<i>lmo2257</i>	hypothetical CDS	2066.3
LMRG_02041	<i>fri</i>	<i>lmo0943</i>	non-heme iron-binding ferritin	1572.6
LMRG_04515	NGN	NL	bacterial signal recognition particle RNA	1462.2
LMRG_02926 ^e	NGN	NL	-	1407
LMRG_00994	<i>lmo1847</i>	<i>lmo1847</i>	similar to adhesion binding proteins and lipoproteins with multiple specificity for metal cations (ABC transporter)	1378.9
LMRG_00378	<i>flaA</i>	<i>lmo0690</i>	flagellin protein	1366.9
LMRG_04523	<i>rnpB</i>	NL	bacterial RNAse P class B	1243.8
LMRG_01165	<i>cspB</i>	<i>lmo2016</i>	similar to major cold-shock protein	1109.5
Noncoding	NGN	NL	T-box leader	1086.7
LMRG_00891	<i>sod</i>	<i>lmo1439</i>	superoxide dismutase	845.4
LMRG_00996	<i>lmo1849</i>	<i>lmo1849</i>	similar to metal cations ABC transporter, ATP-binding proteins	827.4
LMRG_01986	<i>lmo2711</i>	<i>lmo2711</i>	similar to hypothetical proteins	802.1
LMRG_00921	<i>lmo1468</i>	<i>lmo1468</i>	similar to unknown proteins	738.5
LMRG_02618	<i>lmo0196</i>	<i>lmo0196</i>	similar to <i>B. subtilis</i> SpoVG protein	702.9
LMRG_00814	<i>cspL</i>	<i>lmo1364</i>	similar to cold shock protein	679.4
LMRG_01081	<i>hup</i>	<i>lmo1934</i>	similar to non-specific DNA-binding protein HU	631.8
LMRG_00995	<i>lmo1848</i>	<i>lmo1848</i>	similar metal cations ABC transporter (permease protein)	621.2
LMRG_00922	<i>rpsU</i>	<i>lmo1469</i>	30S ribosomal protein S21	609
LMRG_02619	<i>lmo0197</i>	<i>lmo0197</i>	similar to <i>B. subtilis</i> SpoVG protein	577.3
Noncoding	NGN	NL	putative ncRNA	561.9
LMRG_00679	<i>trxA</i>	<i>lmo1233</i>	thioredoxin	516.5
LMRG_01674	<i>lmo2158</i>	<i>lmo2158</i>	similar to <i>B. subtilis</i> YwmG protein	509.2
LMRG_02633	<i>ctc</i>	<i>lmo0211</i>	similar to <i>B. subtilis</i> general stress protein	496.4
LMRG_01479	<i>lmo2363</i>	<i>lmo2363</i>	similar to glutamate decarboxylase	491
LMRG_00517	<i>pdhD</i>	<i>lmo1055</i>	highly similar to dihydrolipoamide dehydrogenase, E3 subunit of pyruvate dehydrogenase complex	483.5
LMRG_00703	<i>lmo1254</i>	<i>lmo1254</i>	similar to alpha, alpha-phosphotrehalase	395.9
LMRG_02718	<i>lmo2373</i>	<i>lmo2373</i>	similar to phosphotransferase system beta-glucoside-specific enzyme IIB component	378.5
LMRG_01737	<i>lmo2511</i>	<i>lmo2511</i>	similar to <i>B. subtilis</i> YvyD protein	377.1
LMRG_00515	<i>pdhB</i>	<i>lmo1053</i>	highly similar to pyruvate dehydrogenase (E1 beta subunit)	356.4
LMRG_00704	<i>lmo1255</i>	<i>lmo1255</i>	similar to PTS system trehalose-specific enzyme IIBC	353.6
LMRG_00516	<i>pdhC</i>	<i>lmo1054</i>	highly similar to pyruvate dehydrogenase (dihydrolipoamide acetyltransferase E2 subunit)	351.3
LMRG_01480	<i>lmo2362</i>	<i>lmo2362</i>	similar to amino acid antiporter (acid resistance)	351.1
LMRG_02239	<i>lmo2692</i>	<i>lmo2692</i>	unknown	344.1
LMRG_00875	<i>lmo1423</i>	<i>lmo1423</i>	unknown	341.2
LMRG_01835	<i>lmo2413</i>	<i>lmo2413</i>	similar to aminotransferase	333.1
LMRG_01429	<i>lmo1541</i>	<i>lmo1541</i>	similar to unknown protein	318.8

^aNGN = No gene name given;

^bNL = No EGD-e locus;

^cAverage normalized number of reads matching each of the σ^B -dependent genes in the two I0403S datasets divided by the length of the genes times 100 bp;

^dThe high coverage of LMRG_01574 is restricted to the portion that overlaps with *lhrA*. LMRG_01574 may not be a valid coding gene;

^eLMRG_02926 completely overlaps with the bacterial RNAse P class B noncoding gene. LMRG_02926 may not be a valid coding gene as no Pfam matches were found for the putative protein coded by this gene.

cules. A total of 7 putative ncRNAs identified here were not previously identified in *L. monocytogenes* and did not match ncRNA entries in Rfam (Table 4). The regions representing these putative ncRNAs showed contiguous coverage by RNA-Seq reads (i.e., at least 100 bp completely covered by RNA-Seq reads), but did not fully match annotated genes. Overall, 36 of the ncRNAs recently identified by tiling microarray analyses in *L. monocytogenes* strain EGD-e [20] were not identified in this study (see Additional file 6: ncRNAs previously described in *L. monocytogenes* strain EGD-e but not identified in this study for a

list of these EGD-e ncRNAs). The most likely explanations for the absence of these EGD-e ncRNAs in I0403S are one or more of the following: (i) low (<0.7 GEI) or no RNA-Seq coverage in I0403S (indicating no transcription in stationary phase I0403S or loss of small RNAs during RNA isolation); (ii) the homolog may be absent in the *L. monocytogenes* I0403S genome (e.g., for EGD-e RliC; Table S3); (iii) ncRNAs determined to be antisense RNA in EGD-e [20] were not identified in I0403S, as the RNA-Seq protocol did not provide for directional reads; (iv) the corresponding I0403S genome region has not been com-

Table 3: Associations between GEI and role categories

	Role categories	Significance ^a
Low average GEI in I0403S	Signal transduction	0.006
	Amino acid biosynthesis	< 0.001
	Transport and binding	0.003
	Viral function	< 0.001
High average GEI in I0403S	Cellular processes	0.011
	DNA metabolism	0.011
	Protein fate	< 0.001
	Protein synthesis	< 0.001
	Purines, pyrimidines, nucleosides, and nucleotides	0.043
	Transcription	< 0.001
	Unknown functions	0.043

^aBased on one-sided Wilcoxon rank sum test and FDR correction.

pletely sequenced and closed (e.g., for EGD-e LhrC, which falls in a repetitive region in the EGD-e chromosome [19]), and (v) the EGD-e ncRNA did not meet our criterion of 100 bases of contiguous coverage.

Three putative ncRNAs with high GEI covered either part or all of each of three annotated CDS, suggesting that ncRNAs overlap with these CDS or that some putative CDS actually encode ncRNAs rather than proteins. Specifically, LMRG_01574 (lmo2257), LMRG_02926 (no homolog in EGD-e), and LMRG_1986 (lmo2711) overlapped with lhrA (partial overlap), with the bacterial RNase P class B ncRNA (full overlap), and with the bacterial signal recognition particle RNA (partial overlap), respectively. In concert with our findings, lmo2257 was previously hypothesized not to be a CDS [19,21].

RNA-Seq identified 96 annotated CDS and one ncRNA as σ^B -dependent and provided comprehensive data on transcript levels for genes in the σ^B regulon

Our RNA-Seq data analyses identified a total of 96 genes as up-regulated by σ^B (Additional file 7: Genes up-regulated by σ^B). No annotated genes were identified as significantly down-regulated by σ^B in this study. Although various genes have been identified previously as down-regulated by σ^B [10,12,20], we have observed that genes

with significantly higher transcript levels in the $\Delta sigB$ strain (i.e., genes identified as down-regulated by σ^B): (i) are likely to be indirectly regulated by σ^B , as σ^B is a transcriptional activator, (ii) generally show a lower fold-difference in transcript levels between the parent strain and the $\Delta sigB$ strain as compared to genes identified as up-regulated by σ^B [10], and (iii) have not been consistently identified as down-regulated by σ^B between different studies, even in microarray studies using the same strain and condition (see Figure 3, which indicates that only 7 genes were identified as down-regulated by σ^B in both of two separate studies with strain 10403S). Down-regulation of genes by σ^B thus appears stochastic as compared to up-regulation by σ^B . Overall, our findings suggest that RNA-Seq combined with stringent criteria for detection of statistically significant differences in transcript levels (i.e., the requirement for statistical significance for all four binomial comparisons) may generate fewer false positives as compared to some microarray-based approaches.

As illustrated in Figure 4A, RNA-Seq data are useful for predicting multi-gene operons controlled by a given regulator such as σ^B . Thirty-eight of the 96 genes up-regulated by σ^B are organized into a total of 20 operons, including (i) opuCABCD, which encodes the subunits of a glycine betaine/carnitine/choline ABC transporter, (ii) lmo0781-

Table 4: New *L. monocytogenes* ncRNAs^a identified in this study

Description	Coordinates in I0403S	Length	I0403S Average GEI ^b	sigB Average GEI ^c
rli64	222952..223741	790	1.99	2.17
rli65	409956..410100	145	43.80	82.82
rli66	938236..938563	328	14.47	29.94
rli67	1393256..1393496	241	52.11	65.68
rli68	2020305..2020575	271	189.49	224.23
rli69	2305436..2305610	175	20.62	49.18
rli70	2370319..2370547	229	45.73	17.84

^aNone of the ncRNAs in this table had matches in the Rfam database;

^bAverage normalized number of reads matching each of the σ^B -dependent genes in the two I0403S datasets divided by the length of the genes times 100 bp;

^cAverage normalized number of reads matching each of the σ^B -dependent genes in the two $\Delta sigB$ datasets divided by the length of the genes times 100 bp.

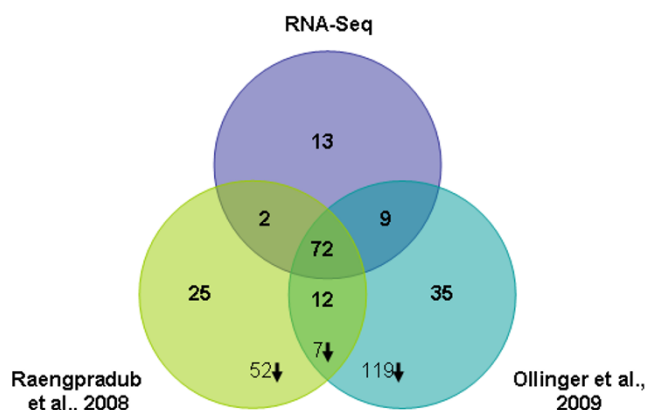


Figure 3
 σ^B -dependent genes identified by RNA-Seq and microarray analyses. Venn diagram of σ^B -dependent genes identified in stationary phase cells in this study and in previous microarray studies of stationary phase *L. monocytogenes* [10,12]. Numbers in bold are the number of up-regulated annotated CDS identified as σ^B -dependent in each study; numbers followed by down arrows are down-regulated σ^B -dependent genes. No down-regulated σ^B -dependent genes were identified by RNA-Seq. The 13 genes identified as σ^B -dependent in stationary phase only by RNA-Seq, but not by previous microarray studies of *L. monocytogenes* 10403S, include 5 genes that had been found to be σ^B -dependent, by microarray studies [10] in salt stressed cells (see Table 5). In a number of instances, (e.g. *opuCB*, *rsbX*; See Additional file 8: Comparison of genes found to be σ^B -dependent by microarray analysis and not by RNA-Seq) genes with significantly different transcript levels in both microarrays [10,12] had significant binomial probabilities ($q < 0.05$) and a fold change ≥ 2.0 for most of the possible combinations (i.e. 10403S replicate 1 vs $\Delta sigB$ replicate 1; 10403S replicate 1 vs $\Delta sigB$ replicate 2; 10403S replicate 2 vs $\Delta sigB$ replicate 1; 10403S replicate 2 vs $\Delta sigB$ replicate 2), but not for all four comparisons and these genes were, therefore, not identified as showing significant differences in normalized RNA-Seq coverage (based on our conservative definition of genes with significant differences in normalized RNA-Seq coverage); see Additional file 8: Comparison of genes found to be σ^B -dependent by microarray analysis and not by RNA-Seq for detailed RNA-Seq data for genes identified as σ^B -dependent by microarrays, but not by RNA-Seq.

lmo0784, which encode the four subunits of a putative mannose-specific phosphotransferase system, (iii) lmo2484-lmo2485, which encode a putative membrane-associated protein and a putative transcriptional regulator similar to PspC, respectively, and (iv) lmo0133 and lmo0134 (Figure 4A), which encode proteins similar to *E. coli* YjdI and YjdJ, respectively.

One-sided Fisher's exact tests were used to determine if σ^B -dependent genes are over-represented within specific TIGR role categories. Genes identified as σ^B -dependent

were over-represented among genes involved in cellular functions (q -value = 0.045). σ^B -dependent genes in this category include genes involved in pathogenesis (*inlA*, *inlB*, *inlH*), adaptation to atypical conditions (lmo0515, lmo0669, lmo2673, *lrcC*), detoxification (lmo1433, lmo2230), cell division (lmo1624) and an unknown protein that may be involved in toxin production and resistance (lmo0321).

We evaluated RNA-Seq transcript levels for the 96 σ^B -dependent genes identified here (Additional file 7: Genes up-regulated by σ^B). The average fold change (10403S GEI/ $\Delta sigB$ GEI) for the 96 σ^B -dependent genes ranged from 2.6 to 479.4. The σ^B -dependent genes with the highest average GEI in 10403S were lmo2158, lmo1602, and lmo0539, which encode a protein similar to *B. subtilis* YwmG, an unknown protein, and a tagatose-1,6-diphosphate aldolase, respectively (Table 5).

An ~ 500 nt σ^B -dependent ncRNA was identified between lmo2141 and lmo2142 (Figure 4B); this ncRNA was recently designated *rli47* [20]. To be consistent with the nomenclature for other σ^B -dependent ncRNA [21], we propose that *rli47* be named *sbrE* (sigma B-dependent RNA). Although BLASTX searches (using 6 possible reading frames) and searches against the Pfam database did not yield significant matches, a σ^B -dependent promoter was identified upstream of the transcript and a Rho-independent terminator was found by TransTermHP (Figure 4B). The sequence for this putative ncRNA was also present in 17 other *L. monocytogenes* genomes, including EGD-e (GenBank accession no. [NC_003210](#)), F2365 (GenBank accession no. [NC_002973](#)), and 15 unfinished genome sequences by the Broad Institute http://www.broad.mit.edu/annotation/genome/listeria_group/MultiHome.html as well as in one *L. innocua* (GenBank accession no. [NC_003212](#)) and one *L. welshimeri* (GenBank accession no. [NC_008555](#)) genome. The 514 nt *sbrE* (*rli47*) sequence was 96.6% conserved among the 18 *L. monocytogenes* genomes.

HMM showed that 84% of σ^B -dependent genes and operons identified by RNA-Seq are preceded by σ^B promoters and therefore, appear to be directly regulated by σ^B

An HMM representing *L. monocytogenes* σ^B -dependent promoters was dynamically created by using an initial training set of experimentally verified *L. monocytogenes* σ^B -dependent promoters to search the RNA-Seq data. The final model yielded a total of 5,387 motifs with scores > 5.00 bits throughout the pseudochromosome sequence. Among these motifs, we identified 65 possible σ^B -dependent promoter sequences upstream of genes and operons identified as σ^B -dependent based on RNA-Seq data (see Figure 5 for the *L. monocytogenes* σ^B promoter sequence

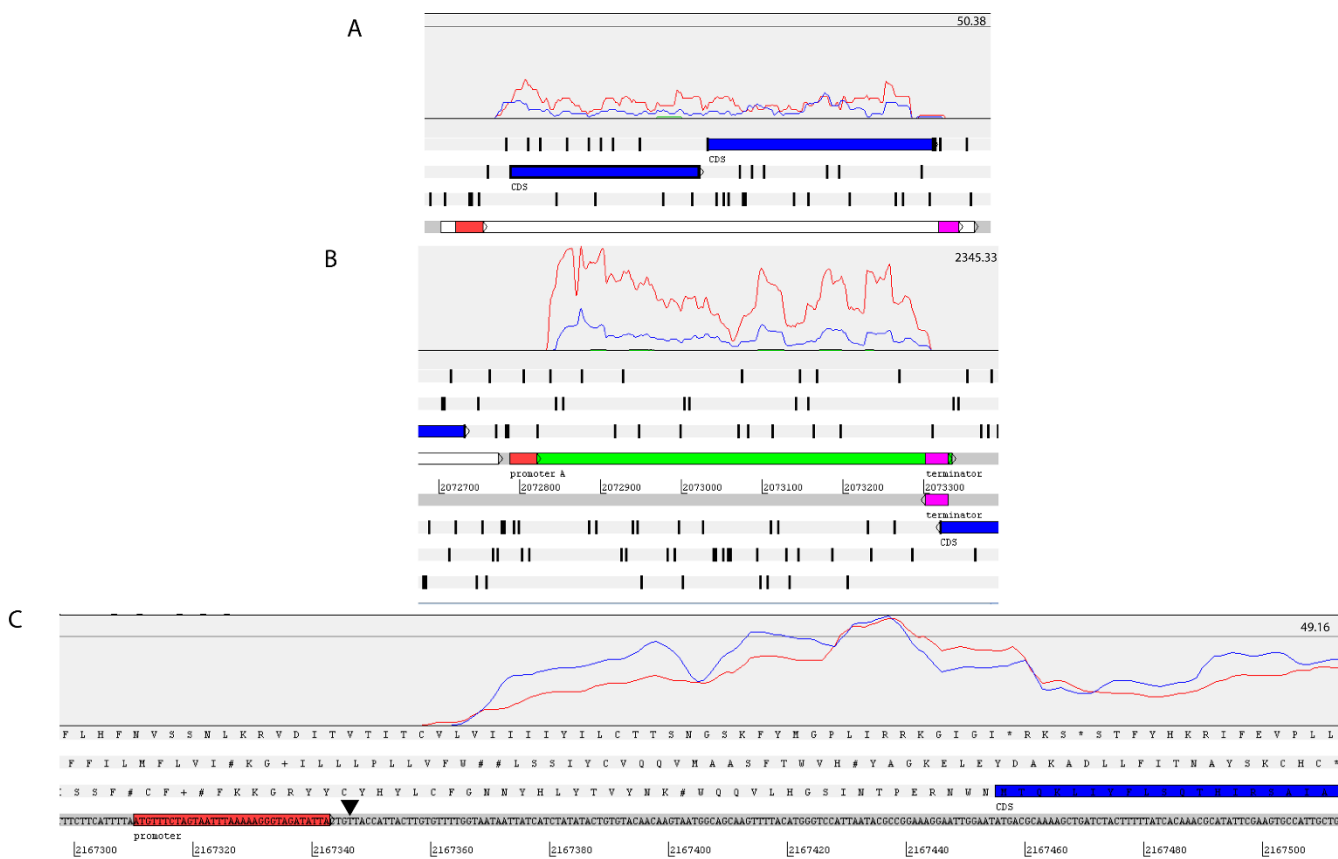


Figure 4
Examples of σ^B -dependent transcripts identified by RNA-Seq. In each panel (A, B, and C), red and blue lines representing normalized RNA-Seq coverage (i.e. the number of reads that match an annotated gene after normalization across runs) in the two 10403S replicates and green and black lines represent normalized RNA-Seq coverage in the $\Delta sigB$ strain replicates; the numbers at the top right in each panel indicates the normalized RNA-Seq coverage represented by the horizontal line shown. Panel (A) depicts LMRG_02382 and LMRG_02383 (shown as blue bars), which form an operon (indicated by a long white bar) with a defined Rho-independent terminator (purple bar) downstream of LMRG_02383; the three positive frames of translation with the coding regions in blue and stop codons shown as vertical black bars are also shown. A σ^B -dependent promoter (red bar) was identified upstream of the operon and the RNA-Seq coverage data clearly shows that the transcription of this operon is positively regulated by σ^B (i.e. almost no coverage was obtained from the $\Delta sigB$ strain). Panel (B) depicts SbrE (Rli47), a σ^B -dependent noncoding RNA (ncRNA) with Rho-independent terminator and a σ^B -dependent promoter identified; annotated features as well as positive and negative frames of translation are shown at the bottom with stop codons shown as vertical black bars. Panel (C) shows the 5' end of LMRG_01602 illustrating the position of a σ^B -dependent promoter in relation to the start codon of the gene and the transcriptional start site determined by RNA-Seq. The black triangle indicates the transcriptional start site determined by RACE-PCR as previously described by Kazmierczak et al. [23].

logo). Because some of the genes with experimentally validated σ^B promoters were not found to be significantly up-regulated by σ^B in our study (e.g. *prfA* and the *rsbV* operon) and because the *ltrC* promoter, which was in the initial training set, had a score below our threshold of 5.00 bits in the final search, our annotation does not include all promoters present in the training set (i.e., only promoters identified upstream of genes that were significantly up-regulated by σ^B in the present study were annotated). Specifically, σ^B -dependent promoter sequences were found upstream of 15 of the 20 putative σ^B -depend-

ent operons, 49 of the 58 monocistronic σ^B -dependent genes, and the one σ^B -dependent ncRNA identified here (Figure 4B). We compared RNA-Seq defined transcriptional start sites for 8 genes with σ^B promoters to transcriptional start sites determined by Rapid Amplification of cDNA Ends PCR (RACE-PCR) in a previous study [23]. Transcriptional start sites identified with RNA-Seq were located between 0 to 29 bases downstream (and therefore sometimes 3') of start sites determined by RACE-PCR (see Figure 4C for LMRG_01602 transcriptional start site mapped by RACE-PCR and RNA-Seq), indicating that

Table 5: Summary of genes up-regulated by σ^B

Locus	EGD-e locus	Description	Avg. fold change (WT/ Δ sigB) ^a	I0403S Average GEI ^b	Δ sigB Average GEI ^c
σ^B-dependent genes found by RNA-Seq and not previously identified by microarray analyses of stationary phase cells					
LMRG_02371	lmo0122	similar to phage proteins	3.9	2.37	0.6
LMRG_02611	lmo0265	similar to succinyldiaminopimelate desuccinylase	204.5	17.95	0
LMRG_02602	lmo0274	unknown	3.17	2.89	0.91
LMRG_00064	lmo0372	similar to beta-glucosidase	4.26	2.4	0.66
LMRG_00126 ^d	lmo0433 (<i>inlA</i>)	Internalin A	5.86	6.19	1.06
LMRG_00127 ^d	lmo0434 (<i>inlB</i>)	Internalin B	6	2.71	0.47
LMRG_02244	lmo0819	unknown	3.01	18.35	6.09
LMRG_00873 ^d	lmo1421	similar to glycine betaine/carnitine/choline ABC transporter (ATP-binding protein)	28.44	5.27	0.67
LMRG_00877 ^d	lmo1425 (<i>opuCD</i>)	similar to betaine/carnitine/choline ABC transporter (membrane p)	3.56	22.59	6.51
LMRG_00878 ^d	lmo1426 (<i>opuCC</i>)	similar to glycine betaine/carnitine/choline ABC transporter (osmoprotectant-binding protein)	3.77	19.78	5.41
LMRG_01013	lmo1866	similar to conserved hypothetical proteins	2.63	4.87	1.79
LMRG_01151	lmo2003	similar to transcription regulator GntR family	14.67	3.15	0.32
LMRG_01963	lmo2733	similar to PTS system, fructose-specific IIABC component	7.95	1.35	0.32
Noncoding	ND	putative ncRNA, <i>sbrE</i>	186.09	2359.89	20.95
σ^B-dependent genes with Average GEI 25 in I0403S					
Noncoding	ND	<i>rliA</i> (<i>sbrE</i>)	186.09	2359.89	20.95
LMRG_01674	lmo2158	similar to <i>B. subtilis</i> YwmG protein	479.39	509.23	22.8
LMRG_01365	lmo1602	similar to unknown proteins	5.47	157.02	30.08
LMRG_00221	lmo0539	similar to tagatose-1,6-diphosphate aldolase	14.54	132.74	9.3
LMRG_01602	lmo2230	similar to arsenate reductase	411	96.43	0
LMRG_02052	lmo0953	unknown	167	73.18	0.48
LMRG_00357	lmo0669	similar to oxidoreductase	75.93	64.6	0.89
LMRG_00358	lmo0670	unknown	105.5	59.6	0.58
LMRG_00341	lmo0654	unknown	7.1	56.61	7.94
LMRG_02219	lmo2674	similar to ribose 5-phosphate epimerase	5.42	52.93	9.94
LMRG_01794	lmo2454	unknown	84.5	50.24	0.76
LMRG_01850	lmo2398 (<i>ltrC</i>)	low temperature requirement C protein, also similar to <i>B. subtilis</i> YutG protein	2.8	50.03	18.94
LMRG_00745	lmo1295 (<i>hfq</i>)	similar to host factor-I protein	4.83	49.77	11.19
LMRG_01948	lmo2748	similar to <i>B. subtilis</i> stress protein YdaG	207.5	49.37	0
LMRG_00583	lmo1140	unknown	11.93	47.84	4.28
LMRG_02036	lmo0937	unknown	54.38	44.68	0.91
LMRG_00484	lmo0796	conserved hypothetical protein	4.21	43.88	10.61
LMRG_02772	lmo1698	similar to ribosomal-protein-alanine N-acetyltransferase	4.1	42.94	10.92
LMRG_02736	lmo2391	conserved hypothetical protein similar to <i>B. subtilis</i> YhfK protein	11.76	39.48	4.54

Table 5: Summary of genes up-regulated by σ^B (Continued)

LMRG_02011	lmo0911	unknown	4.04	33.9	8.58
LMRG_01763	lmo2485	similar to <i>B. subtilis</i> yvIC protein	3.93	32.87	8.47
LMRG_00482	lmo0794	similar to <i>B. subtilis</i> YwnB protein	67.02	32.5	0.72
LMRG_00278	lmo0596	similar to unknown proteins	170.5	32.33	0.09
LMRG_02218	lmo2673	conserved hypothetical protein	150.5	31.92	0.11
LMRG_02013	lmo0913	similar to succinate semialdehyde dehydrogenase	330.38	30.05	0.11
LMRG_00469	lmo0781	similar to mannose-specific phosphotransferase system (PTS) component IID	59.58	29.59	0.65
LMRG_00470	lmo0782	similar to mannose-specific phosphotransferase system (PTS) component IIC	18.99	29.59	1.58
LMRG_01360	lmo1606	similar to DNA translocase	7.88	29.5	3.97
LMRG_02696	lmo2572	similar to Chain A, Dihydrofolate Reductase	8.05	29.05	3.59
LMRG_02768	lmo1694	similar to CDP-abequose synthase	155.31	27.51	0.2
LMRG_02216	lmo2671	unknown	3.13	27.29	8.82
LMRG_02695	lmo2573	similar to zinc-binding dehydrogenase	7.52	25.91	3.83
LMRG_00472	lmo0784	similar to mannose-specific phosphotransferase system (PTS) component IIA	88.5	25.25	0.21
LMRG_02215	lmo2670	conserved hypothetical protein	3	25.23	8.58
LMRG_02697	lmo2571	similar to nicotinamidase	9.84	25.15	2.99

^aAverage fold changes from the I0403S and $\Delta sigB$. Genes with no matching reads in $\Delta sigB$ had their GEI set to 1 to allow for calculation of the fold change;

^bAverage normalized number of reads matching each of the σ^B -dependent genes in the two I0403S datasets relative to the length of the genes times 100 bp;

^cAverage normalized number of reads matching each of the σ^B -dependent genes in the two $\Delta sigB$ datasets relative to the length of the genes times 100 bp;

^dGenes previously identified as σ^B -dependent under salt stress in *L. monocytogenes* I0403S by Raengpradub et al., 2008.

RNA-Seq successfully approximates transcriptional start sites, but sometimes does not provide full sequence coverage to the 5' end of a transcript. Some transcriptional start sites could not be specifically mapped to a σ^B promoter site using RNA-Seq as some genes (e.g. *opuCA*) have multiple promoters. A dendrogram of the putative σ^B promoter sequences showed no apparent clustering of these

promoter sequences by either average GEI in I0403S or by σ^B -dependence (average fold change). These results suggest that additional regulatory elements or mechanisms other than promoter sequence *per se* (e.g., RNA stability) also influence transcript levels and/or σ^B -dependence for these genes (data not shown).

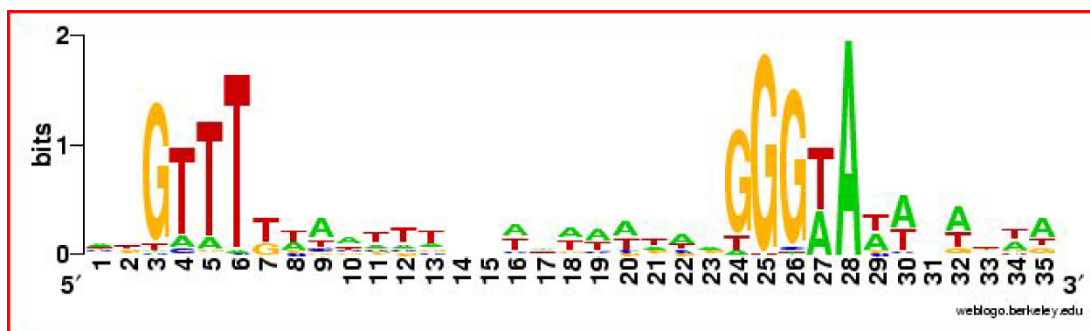


Figure 5
Logo of the σ^B promoter. This logo was created from the alignment of 65 σ^B promoters identified in this study.

RNA-Seq successfully identifies a number of previously identified as well as novel σ^B -dependent genes

To evaluate the ability of RNA-Seq to identify *L. monocytogenes* σ^B -dependent genes, we compared the σ^B -dependent genes identified here with those identified in two independent microarray studies by our research group. Specifically, we compared our results with microarray data reported by (i) Raengpradub et al. [10], who identified σ^B -dependent genes using *L. monocytogenes* strains and growth conditions identical to those in this study, and by (ii) Ollinger et al. [12], who identified σ^B -dependent genes by comparing transcripts from *L. monocytogenes* 10403S with a PrfA* (G155S) allele [24], which constitutively expresses the PrfA-regulated virulence genes [24-26], with those from an isogenic $\Delta sigB$ mutant grown to stationary phase under the same conditions used here. Further, we compared our results with those from a microarray study using another *L. monocytogenes* strain (EGD-e) and its isogenic $\Delta sigB$ mutant, grown under similar conditions (i.e., growth to early stationary phase [11]). Among the 96 σ^B -dependent annotated CDS identified in the present study, 72 were also identified as σ^B -dependent in previous microarray studies of stationary phase *L. monocytogenes* 10403S cells [10,12] (Figure 3). In addition, 64 (66.7%) of the 96 σ^B -dependent genes identified here were identified as positively regulated by σ^B in *L. monocytogenes* strain EGD-e cells grown to early stationary phase (8 h growth in BHI) [11]. Overall, 12 genes identified as σ^B -dependent in stationary phase cells in both previous microarray studies by our group [10,12], were not identified as σ^B -dependent by the RNA-Seq experiments reported here (Figure 3); 9 of these genes showed a σ^B -dependent promoter based on the HMM analyses in this study and are likely to be directly regulated by σ^B (see Additional file 8: Comparison of genes found to be σ^B -dependent by microarray analysis and not by RNA-Seq for further details on these genes).

Finally, a total of 13 annotated CDS identified as σ^B -dependent by RNA-Seq (including 9 genes that also showed a σ^B -dependent promoter in our HMM analysis) had not been identified as σ^B -dependent in either of the previous microarray studies with strain 10403S grown to stationary phase [10,12] (see Table 3). Among these 13 genes not previously identified as σ^B -dependent in stationary phase *L. monocytogenes* 10403S, five had previously been identified as σ^B -dependent in salt-stressed cells [10], including the well-characterized virulence genes *inlA* and *inlB*, which have also been shown by qRT-PCR and promoter mapping to be directly regulated by σ^B [27]. In addition, two of these 13 genes had been identified as positively regulated by σ^B in *L. monocytogenes* strain EGD-e [11], even though they had not been identified as σ^B -dependent in previous microarray studies of strain 10403S [10,12]. For one of these genes (i.e. lmo0265), the microarray probe (designed based on the genome of *L.*

monocytogenes strain EGD-e) showed a low hybridization index (HI; % match between strain-specific sequence and oligonucleotide probe) to 10403S (< 80%). Interestingly, lmo2003, which encodes a transcription regulator similar to the GntR family, was identified as σ^B -dependent by RNA-Seq, but had not been previously identified as σ^B -dependent in either 10403S or EGD-e.

Discussion

In this study, we used deep RNA sequencing to define and characterize the transcriptomes of *L. monocytogenes* strain 10403S and an otherwise isogenic $\Delta sigB$ mutant, which does not express the general stress-response sigma factor, σ^B . The data generated using this approach showed that (i) at least 83% of annotated *L. monocytogenes* genes are transcribed in stationary phase cells; and (ii) stationary phase *L. monocytogenes* transcribes 67 ncRNAs, including one σ^B -dependent ncRNA and seven ncRNAs that, to our knowledge, have not previously been identified in *L. monocytogenes*. Additionally, RNA-Seq data provided for quantitation of transcript levels and approximate identification of transcriptional start sites on a genome scale. Use of a novel, iterative, dynamic HMM, in combination with RNA-Seq data, identified putative σ^B -dependent promoters and further defined the *L. monocytogenes* σ^B regulon.

The majority of annotated *L. monocytogenes* genes are transcribed in stationary phase cells

While genome sequencing and microarray approaches have provided important insight into the biology of prokaryotic organisms, including a number of human bacterial pathogens, identification of all genes and their transcriptional patterns remains a major challenge in all areas of biology. Our results demonstrate that global probe-independent approaches for transcriptome characterization are valuable tools for analyzing bacterial transcriptomes [16,28,29]. A major challenge that currently hinders analysis of transcriptomic data generated by approaches such as RNA-Seq is the ability to differentiate between genes with low levels of transcription and background levels of coverage. Several approaches have been used to define cut-off values between background GEI and GEI indicative of low transcript levels (e.g., [15,30,31]). We chose a comparative analysis of *L. monocytogenes* 10403S transcript levels with those of a mutant strain that does not express a transcription factor (i.e., the alternative sigma factor σ^B) as a novel approach for robustly defining background RNA-Seq coverage. Our results show that a number of σ^B -dependent genes were solely σ^B -dependent (at least under the conditions used here), as supported by the lack of detectable RNA-Seq coverage in the $\Delta sigB$ strain, despite considerable RNA-Seq coverage of the same genes in the isogenic parent strain 10403S. This is an important observation as a number of σ^B -dependent *L. monocytogenes* genes are also activated by other sigma factors (e.g., σ^A [32,33]). Using the average GEI for *L. mono-*

cytogenes genes that were solely σ^B -dependent in the $\Delta sigB$ strain as a conservative cut-off value for transcribed genes, we found that approximately 83% of *L. monocytogenes* 10403S annotated CDS were transcribed in stationary phase cells. These transcribed genes include 355 putative operons, which cover a total of 1,107 genes, indicating that a considerable proportion of *L. monocytogenes* genes appear to be transcribed polycistronically. In comparison, a recent study using a tiling microarray identified 517 polycistronic operons that encompass 1,719 genes in *L. monocytogenes* EGD-e [20]. Taken together, these data indicate that the majority of annotated *L. monocytogenes* genes are transcribed. This conclusion is consistent with results from a whole-genome tiled microarray transcriptome study of *E. coli* MG1655 [34], which reported transcription of 4052 *E. coli* MG1655 genes in bacteria grown under different conditions, suggesting that about 98% of the *E. coli* MG1655 genes are transcribed.

Our results also demonstrate that RNA-Seq coverage levels (generated with the Illumina Genome Analyzer System) correlate well with quantitative RT-PCR-based mRNA transcript level data. Therefore, in combination with results from previous studies (e.g., in yeast [15,31], human cell lines [35], human tissue [36], murine tissue [30]), our findings indicate that RNA-Seq tools can be broadly applied in biological studies to enable quantitative analysis of transcript levels. We also found a positive correlation between RNA-Seq-based transcript levels and codon bias, consistent with the well-documented observation that genes with high codon bias are often highly expressed [37-39]. Genes in four role categories, including (i) signal transduction, (ii) viral functions, (iii) amino acid biosynthesis, and (iv) transport and binding, were significantly associated with lower transcript levels. These categories include a number of genes that encode proteins predominantly required for growth and survival under specialized environmental conditions (e.g., viral replication genes) or under conditions other than stationary phase (e.g., amino acid biosynthesis may be less important in stationary phase than during exponential growth as sufficient amino acids from dead bacteria are likely to be available for scavenging), and/or proteins that may only be required in small amounts. On the other hand, we found that genes in seven role categories, including (i) cellular processes, (ii) DNA metabolism, (iii) protein fate, (iv) protein synthesis, (v) purines, pyrimidines, nucleosides, and nucleotides, (vi) transcription, and (vii) genes encoding proteins with unknown functions, showed, on average, higher transcript levels in stationary phase *L. monocytogenes*. These findings suggest that genes in these particular categories are important for bacterial cells transitioning from exponential growth to stationary phase.

Overall, the *L. monocytogenes* genes with the highest transcript levels were ncRNAs, specifically the transfer-mes-

senger RNA (tmRNA) and 6S RNA, consistent with the observation that tmRNAs are involved with bacterial recovery from a variety of stresses including entry into stationary phase, amino acid starvation, and heat shock [40]. 6S RNA accumulates in cells during stationary phase; cells lacking 6S RNA have reduced fitness relative to wildtype stationary phase cells [41]. In addition to down-regulating some housekeeping genes, 6S RNA has been shown to up-regulate expression of some σ^S -dependent genes in Gram-negative bacteria [41]. σ^S is the stationary phase stress response alternative sigma factor in *E. coli* [42]. Taken together, we hypothesize that 6S RNA plays a critical role in the ability of *L. monocytogenes* to survive stationary phase associated stress conditions.

Specific protein-encoding genes with very high transcript levels in stationary phase *L. monocytogenes* include *fri*, *sod*, *cspB*, and *cspL*, all genes with some previous evidence for contributions to *L. monocytogenes* stationary phase and stress survival [43-49]. *flaA*, which encodes a flagellin protein, was also highly transcribed in stationary phase cells at 37°C. Although *L. monocytogenes* has been reported to show flagellar motility only when grown at $\leq 30^\circ\text{C}$ [50,51], our results are consistent with the observation that strain 10403S, which was used in this study, has been shown to express flagellin at 37°C [51]. Interestingly, we also found some annotated CDS without known function to be highly transcribed, including lmo1847 and lmo1849, which encode putative ABC transporters based on BLAST and Pfam [52] searches, respectively, and lmo1468, which encodes an unknown protein.

RNA-Seq identifies ncRNA molecules in *L. monocytogenes*, including a σ^B -dependent ncRNA, in 10403S

Using RNA-Seq, we found 67 previously identified or putative ncRNAs that were transcribed in stationary phase *L. monocytogenes*. Of these, 7 represent ncRNAs that have not been identified previously as transcribed in *L. monocytogenes*. Sixty of the ncRNAs identified here have previously been reported by Toledo-Arana et al. [20], Nielsen et al. [53], Mandin et al. [22] and Christiansen et al. [19]. Interestingly, 16 *L. monocytogenes* ncRNAs with similarities to ncRNAs identified in other bacterial organisms are putative riboswitches. We also found that *sbrE* (*rli47*), which has no homologies to ncRNA entries in Rfam, appears to be directly regulated by σ^B , based on the considerably higher transcript levels (186 fold) present in the parent strain as compared to the *sigB*-null mutant, consistent with results from a recent tiling microarray study [20]. As the RNA isolation procedure used here selected against small RNA molecules (see Materials and Methods for details), it is likely that additional small ncRNAs not detected here (e.g., some small ncRNAs identified by Toledo-Arana et al. [20]), are also transcribed in stationary phase *L. monocytogenes* 10403S.

Prior to this study, *L. monocytogenes* ncRNAs, including potential σ^B -dependent ncRNAs [53], had been identified using *in silico* modeling [22,53], co-precipitation with the RNA-binding protein Hfq [19], and, most recently, tiling microarrays [20]. While, among these approaches, tiling microarrays [20] provided the most comprehensive characterization of *L. monocytogenes* ncRNAs, deep RNA sequencing also identified a large number of transcribed *L. monocytogenes* ncRNAs, including ncRNAs with no similarities to previously identified ncRNAs. Our results, taken together with previous studies that have identified numerous novel transcripts with RNA-Seq in bacteria (*S. meliloti* [28], *B. cenocepacia* [16], *V. cholerae* [29]), yeast [15,31], mouse [30], Arabidopsis [54], human cell lines [35,55], and human tissue [36], clearly show the power of this technique for characterizing bacterial transcriptomes and ncRNAs.

The *L. monocytogenes* σ^B regulon is composed of at least 96 genes, including 82 genes and 1 ncRNA that are preceded by putative σ^B promoters

As alternative sigma factors, such as σ^B , are known to play critical roles in gene regulation across bacterial genera [33], we used *L. monocytogenes* 10403S and an isogenic $\Delta sigB$ null mutant as a model system for exploring the use of RNA-Seq, in combination with *in silico* analyses, for characterization of transcriptional blueprints associated with bacterial regulatory elements. In our study, RNA-Seq identified 96 annotated CDS and one ncRNA SbrE (Rli47) that are up-regulated by σ^B . Quantitative RT-PCR experiments also confirmed σ^B -dependent transcript levels of SbrE (Rli47) (Mujahid et al., unpublished). Among the 96 σ^B -dependent annotated CDS identified in this study, 74 (77.1%) [10] and 81 (84.4%) [12] were also identified as σ^B -dependent in stationary phase cells in two previous microarray studies using the same strain background. Also, 63 of the 96 σ^B -dependent genes identified here were reported as positively regulated by σ^B in another *L. monocytogenes* strain (EGD-e) grown to early stationary phase [11]. Twelve genes were identified as σ^B -dependent in both previous microarray studies performed with the same *L. monocytogenes* strain background and the same conditions used here, but were not identified as σ^B -dependent by RNA-Seq in this study. This disparity is likely due to the fact that the thresholds and statistical cut-offs used to define σ^B -dependent genes were very stringent in the present study (e.g., a *q*-value < 0.05 in all four comparisons).

Overall, in addition to confirming a previously identified σ^B -dependent ncRNA [20], RNA-Seq identified 13 genes that had not been defined as σ^B -dependent in previous microarray studies of stationary phase *L. monocytogenes* 10403S cells [10,12], including 5 genes that had been identified as σ^B -dependent in salt stressed cells, but not in

stationary phase cells. One gene not previously identified as σ^B -dependent was lmo2003, which encodes a transcription regulator similar to the GntR family. The GntR family of regulators has been characterized as global regulators of primary metabolism in a number of bacteria [56-58]. This finding further supports that *L. monocytogenes* σ^B appears to be involved in a number of transcriptional regulatory networks [6]. Increasing evidence indicates that regulatory RNAs also contribute to regulatory networks that involve *L. monocytogenes* σ^B . For example, in addition to the σ^B -dependent SbrE ncRNA described here, tiling array analyses also identified additional σ^B -dependent ncRNAs. While previous *in silico* studies in *L. monocytogenes* strain EGD-e [53] identified four putative σ^B -dependent ncRNAs (i.e., SbrA, SbrB, SbrC, SbrD), only SbrA was confirmed *in vivo* as σ^B -dependent in EGD-e [20,53]. Even though our RNA-Seq analyses in 10403S identified SbrA transcripts, transcript levels for this ncRNA were not σ^B -dependent under the conditions used in our study. The fact that SbrA was not found to be σ^B -dependent in 10403S may be due to differences in strains or growth conditions used (e.g., Nielsen et al. [53] and Toledo-Arana et al. [20] used strain EGD-e, while we used strain 10403S). Further studies in different *L. monocytogenes* strains will thus be needed to understand the full complexity of regulatory networks in this pathogen, including those involving σ^B and ncRNAs.

The quantitative nature of RNA-Seq allowed us to also identify highly transcribed σ^B -dependent genes, including lmo2158 (which encodes a protein similar to the *B. subtilis* YwmG), lmo1602 (which encodes an unknown protein), and lmo0539 (which encodes a tagatose-1,6-diphosphate aldolase). Interestingly, none of these genes encode proteins that appear to contribute to any of the presently recognized σ^B -dependent phenotypes in *L. monocytogenes*, such as acid resistance [9,59], oxidative stress resistance [59,60], or virulence [27,33,61,62]. As there are no published reports of construction and characterization of null mutations in these highly transcribed σ^B -dependent genes, our data clearly suggest that σ^B and the σ^B regulon make additional important contributions to *L. monocytogenes* physiology that remain to be characterized.

In conjunction with appropriate bioinformatics tools, such as the iterative, dynamic HMM developed in this study to identify putative σ^B promoters, RNA-Seq data also allowed mapping of approximate transcriptional start and termination sites. Specifically, putative σ^B -dependent promoters were identified upstream of (i) 49 monocistronic σ^B -dependent genes, (ii) 15 σ^B -dependent operons (covering a total of 40 genes), and (iii) 1 σ^B -dependent ncRNA. By comparison, in the absence of genome wide transcriptional start site data, a previous study that solely relied on HMM and genome sequence

data identified putative σ^B -dependent promoters upstream of only 40 genes that had been identified as σ^B -dependent by microarray analyses [10]. Our data reported here show that the majority of σ^B -dependent genes are directly regulated by σ^B and illustrate the power of combining RNA-Seq data and bioinformatics approaches for characterizing transcriptional regulatory systems. Specifically, combining transcriptional start site information with an HMM that identifies promoter motifs (e.g., the motif for σ^B -dependent promoters) provides a powerful approach for identifying genes directly regulated by a given transcription factor. This approach facilitates rapid genome-wide identification of putative transcriptional start sites, which currently represents a critical bottleneck in genome-wide characterization of transcriptional regulation and regulatory networks, as many current strategies for promoter mapping (e.g., primer extension, rapid amplification of cDNA ends (RACE-PCR), RNase protection assays) are time- and labor-intensive.

Conclusions

Using the human foodborne pathogen *L. monocytogenes* as a model system, we have shown that RNA-Seq provides a powerful approach to (i) rapidly, comprehensively, and quantitatively characterize prokaryotic genome-wide transcription profiles without hybridization bias, and (ii) characterize putative transcriptional start sites and operon structures. We also show that RNA-Seq transcriptomic evaluation of a bacterial strain bearing a deletion in a transcriptional regulator in comparison with its parent strain can provide rapid, comprehensive insights into the blueprints of prokaryotic transcriptional regulation. Such tools and approaches will revolutionize our ability to characterize genome-wide transcriptional regulatory networks, with wide ranging applications from medicine to ecology, e.g., by providing a means to quickly characterize transcriptional networks contributing to pathogen transmission and virulence as well as environmental growth and gene expression in bacteria used for specific purposes, such as bio-remediation. When applied to both genome and transcriptome sequencing, novel high throughput sequencing approaches can also provide rapid and comprehensive characterization of bacterial genomes, representing an important tool for initial rapid characterization of novel and emerging bacterial pathogens.

Methods

Strains and growth conditions

RNA-Seq was performed on the *L. monocytogenes* parent strain 10403S and a previously described [9] isogenic mutant ($\Delta sigB$, FSL A1-254) with an internal non-polar deletion of *sigB*, which encodes the stress response alternative sigma factor σ^B .

Prior to RNA isolation, bacteria were grown in 5 ml Brain Heart Infusion (BHI) broth (BD Difco, Franklin Lakes,

NJ) at 37°C with shaking (230 rpm) for 15 h, followed by transfer of a 1% inoculum to 5 ml pre-warmed BHI. After growth to $OD_{600} \sim 0.4$, a 1% inoculum was transferred to a 300 ml nephelo flask (Bellco, Vineland, NJ) containing 50 ml pre-warmed BHI. This culture was incubated at 37°C with shaking until cells reached stationary phase (defined as growth to $OD_{600} = 1.0$, followed by incubation for an additional 3 h). Two independent growth replicates and RNA isolations were performed for each strain.

RNA isolation, integrity and quality assessment

RNA isolation was performed as previously described [10]. Briefly, RNAProtect bacterial reagent (Qiagen, Valencia, CA) was added according to the manufacturer's instructions to the cultures grown to stationary phase; treated cells were stored at -80°C (for no longer than 24 h) until RNA isolation was performed. Bacterial cells were treated with lysozyme followed by 6 sonication cycles at 18W on ice for 30 s. Total RNA was isolated and purified using the RNeasy Midi kit (Qiagen) according to the manufacturer's protocol; RNA molecules <200 nt in length are not recovered well with this procedure, according to the manufacturer. RNA was eluted from the column using RNase-free water. Total RNA was incubated with RQ1 DNase (Promega, Madison, WI) in the presence of RNasin (Promega) to remove remaining DNA. Subsequently, RNA was purified using two phenol-chloroform extractions and one chloroform extraction, followed by RNA precipitation and resuspension of the RNA in RNase free TE (10 mM Tris, 1 mM EDTA; pH 8.0; Ambion, Austin, TX). UV spectrophotometry (Nanodrop, Wilmington, DE) was used to quantify and assess purity of the RNA.

Efficacy of the DNase treatment was assessed by TaqMan qPCR analysis of DNA levels for two housekeeping genes, *rpoB* [63] and *gap* [33]. qPCR was performed using TaqMan One-Step RT-PCR Master Mix Reagent and the ABI Prism 7000 Sequence Detection System (all from Applied Biosystems, Foster City, CA). Each RNA sample was run in duplicate and standard curves for each target gene were included for each assay to allow for absolute quantification of residual DNA. Data were analyzed using the ABI Prism 7000 Sequence Detection System software as previously described [64]. Normalization and log transformation were performed as described by Kazmierczak et al. [23]. All samples showed log copy numbers ≤ 1.5 and C_t values > 35 for both *rpoB* and *gap*, indicating negligible levels of DNA contamination. As a final step, RNA integrity was assessed using the 2100 Bioanalyzer (Agilent, Foster City, CA).

mRNA enrichment

Removal of 16S and 23S rRNA from total RNA was performed using MicroExpress™ Bacterial mRNA Purification Kit (Ambion) according to the manufacturer's protocol with the exception that no more than 5 μ g total

RNA was treated per enrichment reaction. Each RNA sample was divided into multiple aliquots of $\leq 5 \mu\text{g}$ RNA and separate enrichment reactions were performed for each sample. Enriched mRNA samples were pooled and run on the 2100 Bioanalyzer (Agilent) to confirm reduction of 16S and 23S rRNA prior to preparation of cDNA fragment libraries.

Preparation of cDNA fragment libraries

Ambion RNA fragmentation reagents were used to generate 60-200 nucleotide RNA fragments with an input of 100 ng of mRNA. Following precipitation of fragmented RNA, first strand cDNA synthesis was performed using random N₆ primers and Superscript II Reverse Transcriptase, followed by second strand cDNA synthesis using RNaseH and DNA pol I (Invitrogen, CA). Double-stranded cDNA was purified using Qiaquick PCR spin columns according to the manufacturer's protocol (Qiagen).

RNA-Seq using the Illumina Genome Analyzer

The Illumina Genomic DNA Sample Prep kit (Illumina, Inc., San Diego, CA) was used according to the manufacturer's protocol to process double-stranded cDNA for RNA-Seq, including end repair, A-tailing, adapter ligation, size selection, and pre-amplification. Amplified material was loaded onto independent flow cells; sequencing was carried out by running 36 cycles on the Illumina Genome Analyzer.

The quality of the RNA-Seq reads was analyzed by assessing the relationship between the quality score and error probability; these analyses were performed on Illumina RNA-Seq quality scores that were converted to phred format <http://www.phrap.com/phred/>. Quality scores are reported in Additional file 9: Distribution of quality scores for all RNA-Seq runs.

RNA-Seq data will be available in the NCBI GEO Short Read Archives: <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE15651>.

RNA-Seq alignment and coverage

The program nucmer, which is part of the MUMmer package <http://mummer.sourceforge.net/>, was used to align the 10403S unfinished genome sequences (available at http://www.broad.mit.edu/annotation/genome/listeria_group/MultiHome.html as supercontigs 5.1 to 5.21) against the finished genome sequence of the *L. monocytogenes* reference strain EGD-e [18] to create a pseudochromosome for 10403S. Creation of the 10403S pseudochromosome was performed using the order and orientation of the 10403S supercontigs provided by the alignment with EGD-e; the assembled pseudochromosome was 2.87 Mb long. The annotation of the genes in the individual 10403S supercontigs, as provided by the

Broad Institute http://www.broad.mit.edu/annotation/genome/listeria_group/MultiHome.html was then mapped to the 10403S pseudochromosome (Additional file 10: Genbank (gbk) file with ncRNAs identified here). The 5S, 16S and 23S rRNA genes as well as the various tRNA genes in 10403S were identified using blastn and the EGD-e annotated rRNA and tRNA genes as a reference (Genbank ID: [AL591824](http://www.ncbi.nlm.nih.gov/nuccore/AL591824)).

Based on quantitative analyses of RNA-Seq data, throughout this manuscript, transcript levels of a given gene are reported as the Gene Expression Index (GEI), which is expressed as number of reads per 100 bases. To obtain the GEI, the 10403S pseudochromosome was used to align Illumina RNA-Seq reads. These alignments were performed using the whole genome alignment software Eland (Illumina), which reports unique alignments of the first 32 bases of each read, allowing up to 2 mismatches. Coverage at each base position along the pseudochromosome was calculated by enumerating the number of reads that align to a given base. The coverage for each base from the first to last nt in an annotated CDS was summed then divided by 32 (i.e., the length of each aligned read) to obtain the RNA-Seq coverage for that gene before normalization. The following data were discarded prior to further analyses: (i) reads with more than 2 mismatches, (ii) reads that matched to multiple locations, (iii) reads that did not map to the chromosome, and (iv) reads that mapped to the 16S or 23S genes (Table 1). Reads identified as "matching two locations" did not include those matching rRNA genes as the 10403S pseudochromosome created for this study was designed with only one unique rRNA gene sequence. Reads matching the 16S and 23S genes were removed prior to normalizing the total number of aligned reads across the four samples because of the technical bias introduced by our deliberate partial removal of 16S and 23S transcripts from the samples. Despite removal of 16S and 23S rRNA, in a given run, between 1,860,817 and 3,138,329 reads aligned to the 23S gene and between 434,263 and 760,863 reads aligned to the 16S gene. In a given run, between 101,419 and 242,246 reads matched the 5S rRNA gene and between 7,778 and 62,699 reads matched the various tRNA genes present in the pseudochromosome.

Because of the inherent differences in the total number of reads among the four runs, the total number of reads for each run was normalized to the run with the highest coverage (i.e. $\Delta sigB$ replicate 2, Table 1). The ratio of total number of reads for $\Delta sigB$ replicate 2 to the total number of reads for 10403S replicate 1, 10403S replicate 2, or $\Delta sigB$ replicate 2 was used as a multiplier to normalize the approximate number of reads matching a given gene (Table 1). The GEI was then obtained by dividing the normalized number of reads matching each gene by the gene

length. The average GEI was the number of reads that match each nt in a given gene after normalization; this value represented the average of the 2 biological replicates for a given strain and is presented as reads per 100 bases (as opposed to reads per 1 base) to simplify identification of differences. The distribution of the coefficient of variation for each gene between replicates is depicted in Additional file 11: Coefficient of variation among RNA-Seq replicates by strain.

Identification of transcribed annotated CDS

Sequence reads matching annotated CDS in the 10403S genome were used to identify those annotated CDS that were transcribed under the experimental conditions used. As our RNA-Seq analyses included both a wildtype strain and an isogenic mutant with a deletion in a transcriptional regulator (i.e., the alternative sigma factor σ^B), our data also provide a novel approach for characterizing background RNA-Seq coverage for genes that are not transcribed, similar to a previous approach that used background RNA-Seq coverage of so-called "gene deserts" in human chromosomes to characterize background average GEI [65]. The observations that (i) eight genes that showed average GEI between 8.64 reads and 96.43 reads per 100 bases in the parent strain showed 0 reads per 100 bases in the $\Delta sigB$ strain; (ii) 42 genes with average GEI of 1.21 to 73.81 reads per 100 bases in the parent strain showed between 0.01 and 0.7 reads per 100 bases in the $\Delta sigB$ strain; and (iii) 0.7 reads per 100 bases is the approximate median of the average GEI in σ^B -dependent genes in the $\Delta sigB$ strain, clearly indicate that extremely low background RNA-Seq coverage is expected for genes that are not transcribed. Overall, 50/96 σ^B -dependent genes show an average GEI < 0.7 in the $\Delta sigB$ strain (Additional file 7: Genes up-regulated by σ^B); genes with GEI < 0.7 reads are overrepresented in the $\Delta sigB$ strain (Figure 6). It is not unexpected that some σ^B -dependent genes showed average GEI ≥ 0.7 as a number of genes are not solely dependent on σ^B and will still be transcribed in the absence of σ^B (e.g., *opuCABCD* operon [32,66,67]). Based on these observations, we set an average GEI ≥ 0.7 as a conservative cut-off to identify genes that are transcribed (i.e., we define genes with average GEI ≥ 0.7 as being transcribed as the RNA-Seq data indicate that non-specific reads [e.g., from DNA] are highly unlikely to provide average GEI ≥ 0.7).

Depending on RNA-Seq coverage, genes were classified into four categories, including (i) not transcribed (average GEI < 0.7), (ii) low transcript levels (average GEI ≥ 0.7 and < 10), (iii) medium transcript levels (average GEI ≥ 10 and < 25), and (iv) high transcript levels (average GEI ≥ 25). While cut-offs between low, medium, and high transcript level categories were somewhat arbitrary, they were chosen to yield a relative distribution of genes into these categories similar to the distribution of yeast genes into low,

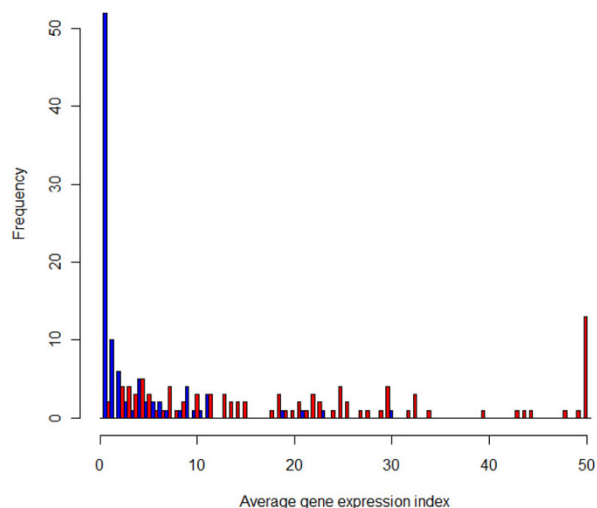


Figure 6
Average gene expression indices for σ^B -dependent genes. The histogram shows the average GEI of σ^B -dependent genes in 10403S (red) and the $\Delta sigB$ (blue) strains. GEIs were grouped in intervals of 0.7, i.e., the first bar represents genes with GEIs between 0 and 0.7; the second bar represents GEIs between > 0.7 and ≤ 1.4 , etc. Genes with average GEI ≥ 50 were grouped together.

medium, and high expression categories reported previously by Nagalakshmi et al. [15].

Annotation of Rho-independent terminators and putative operons

Potential operons were manually annotated based on the continuity of a similar level of RNA-Seq coverage across consecutive genes and the (i) absence of putative Rho-independent terminators between genes, and/or (ii) presence of a putative Rho-independent terminator at the end of a putative operon. Putative Rho-independent terminators in the 10403S pseudochromosome were identified using the program TransTermHP v2.04 [68].

Discovery and annotation of regions transcribing ncRNAs

To aid in identification of transcribed ncRNAs, ncRNAs previously identified in *L. monocytogenes* EGD-e [19-22] were mapped onto the 10403S pseudochromosome and were identified as transcribed in 10403S in this study.

New putative ncRNAs (i.e., ncRNAs not previously reported or previously identified by Rfam) were manually identified using the genome browser Artemis [69]. Specifically, regions not matching annotated genes, but showing contiguous coverage by RNA-Seq reads (i.e., regions that contain at least 100 bp completely covered by RNA-Seq reads) were designated putative ncRNAs. Further, RNA-Seq reads that did not cover an entire annotated CDS, but showed partial contiguous coverage within a

CDS, were also designated as putative ncRNAs. All ncRNAs, including those reported in previous publications [19,20,22,53], those identified by Rfam, and those with no matches to the Rfam database were annotated into a Genbank (gbk) file that is available as Additional file 10: Genbank (gbk) file with ncRNAs identified here. ncRNAs identified by RNA-Seq, but with no matches to the Rfam database were designated "putative ncRNA" and received designations from *rli64* to *rli70*. The presence of rho-independent transcriptional terminators was used to assign the strand of putative ncRNAs. For two instances where terminators were not observed, the ncRNAs were annotated on both strands.

Differential expression analysis

To identify genes that showed significantly different transcript levels in the parent strain (10403S) and the $\Delta sigB$ strain, statistical analyses were performed using the normalized RNA-Seq coverage of each coding gene (as annotated by the Broad Institute). Normalized RNA-Seq coverage (i.e. the number of reads that match an annotated CDS after normalization across runs) was used in lieu of the GEI (in which the normalized RNA-Seq coverage number is divided by the gene length) for statistical analyses. Corresponding analyses were also performed for each region encoding a putative ncRNA transcript identified as described above. A coverage file of normalized RNA-Seq coverage is available in Additional file 12: Coverage file with the normalized RNA-Seq coverage for the 4 RNA-Seq runs.

For each gene, a binomial probability was calculated for the normalized RNA-Seq coverage, using each of the four possible comparisons between the 10403S and $\Delta sigB$ transcripts (i.e. 10403S replicate 1 vs $\Delta sigB$ replicate 1; 10403S replicate 1 vs $\Delta sigB$ replicate 2; 10403S replicate 2 vs $\Delta sigB$ replicate 1; 10403S replicate 2 vs $\Delta sigB$ replicate 2). The binomial probability was calculated under the hypothesis that genes that are not regulated by σ^B will show the same normalized number of reads in the two strains ($p = 0.5$ and $q = 0.5$). For a gene to be considered up-regulated by σ^B , the binomial probability of observing as many reads in the $\Delta sigB$ strain as those observed for 10403S had to be < 0.05 for each of the four possible combinations. Conversely, for a gene to be considered down-regulated by σ^B , the binomial probability of observing as many reads as those observed for $\Delta sigB$ had to have q -values < 0.05 for each of the four possible combinations. To control for multiple comparisons, a False Discovery Rate (FDR) approach was used. q -values (representing the FDR) were calculated using the program Q-Value [70] for R. Only genes with q -values < 0.05 and fold change ≥ 2 or ≤ 0.5 among all four possible comparisons between 10403S and $\Delta sigB$ were considered significantly up-regulated or down-regulated by σ^B .

Iterative HMM-based promoter identification

An initial training set containing 17 experimentally validated σ^B -dependent promoter motifs was used to build a Hidden Markov Model (HMM) of these motifs (Additional file 13: σ^B -dependent promoters used for HMM search). HMM construction and searches were performed using the program hmmer version 1.8.5. The HMM was constructed from unaligned sequences (using hmmt) and then used to search the 10403S pseudochromosome (using the hmmls tool). The null frequencies of each nucleotide used were those observed in the *L. monocytogenes* genome (i.e., A/T = 0.31 and G/C = 0.19).

To identify new promoter motifs that could be added to the training set, we used an iterative HMM approach. In each given HMM iteration, the only hits added to the training set were those that met four conservative criteria, including (i) location within 100 bp upstream of the start codon of an annotated CDS (or 100 bp upstream the first nt for the manually annotated noncoding genes), (ii) q -values < 0.05 (from the binomial probabilities) for σ^B dependence of a given gene (based on RNA-Seq data), and (iii) fold change ≥ 2 among all possible comparisons between 10403S and $\Delta sigB$, and (iv) a score higher than the lowest score for which 50% of the motifs fall in noncoding regions (i.e. for each iteration, we adaptively chose a threshold score such that 50% of the motifs that score higher than this threshold lie in noncoding regions). After adding all hits that met these criteria (in a given iteration) to the training set, a new model was built and used to search the 10403S pseudochromosome. This process was repeated until no new motifs could be added to the training set; the final training set can be found in Additional file 13: σ^B -dependent promoters used for HMM search. When no new motifs that matched our criteria were discovered, the model was considered complete and the results from the last search were used for promoter identification. The final model was used to search the 10403S pseudochromosome for potential σ^B promoters. Potential σ^B promoters identified by this HMM upstream of σ^B -dependent genes and the σ^B -dependent putative ncRNA were visually evaluated. Potential σ^B promoters identified by HMM were considered probable σ^B promoters if the promoter was within 50 bp upstream of the transcriptional start site (as identified by RNA-Seq). In some instances, the transcriptional start site was not discernable due to an upstream gene transcript that overlapped with a σ^B -dependent gene transcript or because the gene had a low average relative normalized RNA-Seq coverage. For these instances, putative promoters were considered if they were located within 200 bp from the start codon of the σ^B -dependent gene. σ^B -dependent genes with probable σ^B promoters are described in Figure 7; the σ^B promoter sequence logo is presented in Figure 5 [http://weblogo.berkeley.edu/\[71\]](http://weblogo.berkeley.edu/[71]).

	10	20	30	Score (bits)
<i>Imo0133/Imo0134</i>	ACGTTT	TC TTTTGGTTGATg	AGTGG AATAGATGG	13.64
<i>Imo0169/Imo0170</i>	TCATTT	TAA - CAGATAATT	ACGGGAATAGGAAT	8.56
<i>inIH</i>	TTGTTA	ATTTGGTCTAAAA	AAGGGTATCTATTA	16.55
<i>Imo0265</i>	GAGTTT	GCCTTTATAGAGA	ACGGGAAAACATAG	8.24
<i>Imo0274</i>	GCGGTT	ACATTGGCTAAAA	AAGGGTATTTTCAT	11.83
<i>Imo0321</i>	TGGTTT	GCGAAGGGAATAA	GAGGGAATAAGTA	8.60
<i>Imo0372</i>	TTGTTT	TTT - AAATAAATG	TATGCTATATTTAT	6.35
<i>Imo0405</i>	CTTTTT	ATATTTGTATAAA	AGGGGTATAGACAA	20.35
<i>inIA</i>	ATGTGT	TATTTTGAACATA	AAGGGTAGAGGATA	22.73
<i>inIB</i>	TAGCTA	TTATTTT AGTTT	ATGGATAATTATTG	5.51
<i>Imo0439</i>	TTGTTT	CACCGC ACTGCTT	t CAGGGAAC TATTA	12.26
<i>Imo0515</i>	AGGCC T	AAAATC ATTTT TA	TAGGGTATGGAATA	8.12
<i>Imo0539</i>	ATGTTT	TAAAAAATTATT	c AGTGGTATAAT AAG	8.04
<i>Imo0554/Imo0555</i>	AGGTTT	AAATTTCTAAAAa	AAGTGTATTATTAA	16.32
<i>Imo0593</i>	GTGTTT	T AAGAGTTTGAAAa	CGGGGAATTAACA	11.50
<i>Imo0596</i>	AGGTTT	TAAATTCGTTTTT	TAGGCTATTATAGT	11.53
<i>Imo0602</i>	CGCATT	CTTTTGGTTAAAA	AGGGGTAAAGTCAG	7.04
<i>Imo0610</i>	TTGTTT	AACATATTACTAAa	AGAGGAATTGTCTT	11.89
<i>Imo0628/Imo0629</i>	TC TTTT	GAATAAAGTTAAAa	TCGGGTATACAGAT	10.63
<i>Imo0654/Imo0655</i>	AGGATT	ACATTTCTATTTA	t TGGGGAAGTAGA	13.52
<i>Imo0669/Imo0670</i>	ACGTTT	TAGCGTAAAAC TG	GAGGGAAGACATAA	17.07
<i>Imo0722</i>	ATGAAT	ACCTTCTAAAAA	CAGGGTAAACGAAA	12.34
<i>Imo0781-Imo0784</i>	GCGTTT	t CTGACTAATCTTT	TAGGGTAAATGTTG	8.85
<i>Imo0794</i>	ATGTTT	CCCAGTCCCTCT	t TCGGGAATAATTCT	6.61
<i>Imo0796</i>	AGGTTT	AATTTCTT AAGAT	t TAGGCTAGATTATA	15.98
<i>Imo0880</i>	CGGTTT	t TAAC AAGC AAGTT	GTGGGAAC TATAAA	9.87
<i>Imo0911</i>	TTGTTT	TAACTTGCCCTC	Ag GCGGGTATTTATTA	18.74
<i>Imo0913</i>	CTGATT	AAATTTTTCGATT	TGTGGAAAACACTA	16.63
<i>Imo0937</i>	ATGTTT	AAAGACTGATCTC	ACGGGAATATATAA	14.14
<i>Imo0953</i>	TTGTTT	TACTTCTACTTTT	t TAGGGAATAAAAATA	18.32
<i>Imo0994</i>	CGGTAT	TATTATCGAGTCG	ATTGGTAAAAATAT	6.90
<i>Imo1241</i>	GCGATT	GAGCATCCAAAAc	AGGGGTATCATATA	5.23
<i>Imo1295</i>	CTGTTT	GGTAAGAAGAAAT	a AAGGGTATTTGAAA	12.68
<i>Imo1421</i>	AGGAAT	ATTTAGGGATGAT	t TAGGGTAAATGGAT	6.21
<i>opuCA-opuCD</i>	AAGTTT	AAATCTACTACT	AGT TAGGGAATAGTT	14.51
<i>Imo1433</i>	TCGTTT	GAAAGTGAAATC	Ag ACGGGAAC AAGC	12.89
<i>Imo1526</i>	TCGTTT	TTAATAGGACAG	Aa ACGGGTACAGATA	19.27
<i>Imo1601/Imo0602</i>	AAGTTT	TAGAGGGGAATAC	t CAGGGTATAGAAAA	16.36
<i>Imo1694</i>	TGGTTT	TAATACTACTAAA	AAGGGAATAAACTA	20.16
<i>Imo1698</i>	AAGTTT	ATTTTTTAAATAAA	ATGGGTATATAGAA	20.02
<i>Imo1830</i>	CCGTTT	TTTCTTTCTAATT	t TAGGGTAGATGTGT	19.11
<i>Imo1883</i>	TAGTTT	TATTTTCTACTATG	TTGGGTATTTTCTA	19.91
<i>Imo2067</i>	ATGTTT	TACTCCAAACTCC	GAGGGTACTGGTAT	11.91
<i>Imo2085</i>	CTGTTT	TC TTTTGTGTTTT	t ATGGGTATTTAATG	16.67
<i>Imo2132</i>	AAGTTT	TATGCGCTTATAT	t GCGGGAAC TAAATG	10.12
<i>sepA</i>	AGGTTT	TGAATAATTTTAT	GGAGGTATAAAAAG	8.00
<i>Imo2158</i>	ATGTTT	TAGCTTCTATAT	TGTGGAAAACACTA	19.87
<i>Imo2213</i>	CTGTTT	CAATTATGAAAAA	CGTGGAAAATAAAG	13.26
<i>Imo2230</i>	ATGTTT	CTAGTAATTTAAA	AAGGGTAGATATTA	19.51
<i>Imo2269</i>	ATTTTT	TGTCATTATAGCA	CATGGTAATTTGCA	7.60
<i>Imo2387</i>	TAGTTT	ACAGCTATATGTT	a AAGGGAATCATG	11.16
<i>Imo2397</i>	TGGTTT	TATTTTTTACTCA	CCGGGAAGTTCT	18.02
<i>Imo2434</i>	CGGTTT	GTCTCTGTGTTTT	a ATGGGTATTGGTGA	15.30
<i>Imo2454</i>	CTGTTT	TAAAAATAACGAG	AGGGGTAAATGATTT	16.55
<i>Imo2463</i>	ATGTTT	g GCATATGTAAAAA	AGAGGTATAAATTA	12.36
<i>Imo2570-Imo2573</i>	GTGCAT	TATTTTAAAGAAAT	TCGGGAAAAGGAAA	9.41
<i>Imo2602/Imo2603</i>	TTGTTT	TGGTTTAAATGCC	AAGGGAATATATTA	22.31
<i>Imo2670-Imo2672</i>	ATGATT	AAAGAGAAAATTT	TGTGGTACTATGAG	5.03
<i>Imo2673/Imo2674</i>	ATGCCT	CTTTC TTTTATTT	ATGGGTATTAAGTA	12.52
<i>Imo2724</i>	TAGTTT	AAGGTAAAACGAA	TTGGGTATTTTCTA	15.79
<i>Imo2733</i>	TGGTAT	AACTAAAGTAACC	TAGGGAAGACTG	8.66
<i>Imo2748</i>	ATGTTT	AAAGCCGGGAGCC	g AGTGGAAAGGTACA	5.16
<i>Imo0019</i>	TCTTTT	TATTTTTCCAAAA	TAGGGTATACATAA	20.78
<i>Imo0043</i>	CTGTGC	TTTTTTTTATTTT	CAGGGAATCTCTC	6.07
<i>ncRNA</i>	GCGTTT	ACATTTATTTAGA	ACGGTTATATATAT	9.54

Figure 7

Alignment of the 65 putative B-dependent promoters identified in this study. EGD-e homologs of genes or operons downstream of a given promoters are indicated on the left. Positions 3 to 6 in the alignment represent the -35 region while positions 24 to 29 represent the -10 region. Darker nucleotides are more conserved than lighter nucleotides in the alignment. Gene names that are boxed indicated promoters that have been experimentally validated (e.g., by RACE-PCR).

Correlation of RNA-Seq relative coverage (GEI) with TaqMan absolute transcript copy number

Average GEI was correlated with absolute transcript copy numbers quantified by TaqMan qRT-PCR. qRT-PCR-based transcript level data obtained for selected genes in *L. monocytogenes* grown under the same conditions used here (i.e., stationary phase) were obtained from previous studies and unpublished work (see Additional file 2: RNA-Seq average GEI and TaqMan qRT-PCR absolute copy number); qRT-PCR methods are detailed in Raengpradub et al. [10]. qRT-PCR data from these studies were used to calculate absolute transcript copy numbers (using a standard curve as described by Sue et al. [64]); values were log transformed.

Statistical Analyses

One-sided Wilcoxon rank sum tests were used to assess whether genes in certain role categories showed lower or higher average GEI in 10403S than genes in other role categories. One-sided Fisher's exact tests were used to assess whether σ^B -dependent genes were overrepresented in certain TIGR role categories <http://cmr.jcvi.org/cgi-bin/CMR/RoleIds.cgi>. Linear regression analysis was used to assess correlations between average GEI and qRT-PCR data as well as between codon bias and average GEI in 10403S. The effective number of codons used in a gene (N_c), a measure of the codon bias, was assessed using the program "chips" implemented in the EMBOSS package [72]. All tests were carried out in R (version 2.7.0; <http://www.r-project.org/>). Correction for multiple testing was performed using the procedure reported by Benjamini & Hochberg [73], as implemented in the program Q-Value [70]. Significance was set at 5%.

Data access

RNA-Seq data will be available in the NCBI GEO Short Read Archives. All RNA-Seq data are provided in an Access database file (Additional file 4: Access database). This database contains information on the annotated CDS and ncRNAs with their 10403S locus name, 10403S start and end coordinates, lengths, strand, EGD-e locus, EGD-e gene name, EGD-e common name, EGD-e role category, codon bias, GEI, average GEI in 10403S and $\Delta sigB$ strains, fold change for the four possible comparisons involving the two replicates with 10403S and the $\Delta sigB$ strains, q -values of the binomial tests, operon annotation, promoter annotation, list of σ^B -dependent genes identified in this study, and data from 3 other studies of the σ^B regulon in *L. monocytogenes* using microarrays including Ollinger et al. [12], Hain et al. [11], and Raengpradub et al. [10].

Abbreviations

GEI: Gene Expression Index; RNA-Seq: RNA Sequencing; ncRNA: noncoding RNA; RACE-PCR: Rapid Amplification of cDNA Ends PCR; FDR: False Discovery Rate; HMM: Hidden Markov Model

Authors' contributions

HFO and RHO participated in the design of the study, prepared RNA for sequencing, completed all data analysis and method comparisons and drafted the manuscript. LP and QS assembled the *L. monocytogenes* 10403S pseudochromosome, aligned RNA-Seq reads, and implemented the iterative HMM searches. UK guided HMM design. WW processed enriched RNA for sequencing. SWC and MJF participated in study design. MW and KJB conceived of the study, and participated in its design and coordination and helped to draft the manuscript. All authors read and approved the final manuscript.

Additional material

Additional file 1

Sequencibility text file. The resulting plot, when used in conjunction with the Artemis genome browser, shows the regions that can (0) and cannot (1) be sequenced in the 10403S pseudochromosome with the Illumina Genome Analyzer. Regions that cannot be sequenced appear as high peaks.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-641-S1.TXT>]

Additional file 2

RNA-Seq average GEI and TaqMan qRT-PCR absolute copy number of select genes.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-641-S2.XLS>]

Additional file 3

Cumulative frequency of average GEI in L. monocytogenes 10403S. The vertical line indicates an average GEI of 0.7 reads, which is the cut-off used to identify transcription. The graph shows that about 83% of the genes fall at the right of the average GEI cut-off of 0.7 reads and were therefore considered transcribed.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-641-S3.TIFF>]

Additional file 4

Access database. All RNA-Seq data are provided in an Access database file. This database contains information on the annotated CDS and ncRNAs with their 10403S locus name, 10403S start and end coordinates, lengths, strand, EGD-e locus, EGD-e gene name, EGD-e common name, EGD-e role category, codon bias, GEI, average GEI in 10403S and $\Delta sigB$ strains, fold change for the four possible comparisons involving the two replicates with 10403S and the $\Delta sigB$ strains, q -values of the binomial tests, operon annotation, promoter annotation, list of σ^B -dependent genes identified in this study, and data from the other 3 studies of the σ^B regulon in *L. monocytogenes* using microarrays including Ollinger et al. [12], Hain et al. [11], and Raengpradub et al. [10].

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2164-10-641-S4.MDB>]

Additional file 5*ncRNAs identified by RNA-Seq.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2164-10-641-S5.XLS]

Additional file 6*ncRNAs previously described in L. monocytogenes strain EGD-e but not identified in this study.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2164-10-641-S6.XLS]

Additional file 7*Genes up-regulated by σ^B .*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2164-10-641-S7.XLS]

Additional file 8*Comparison of genes found to be σ^B -dependent by microarray analysis and not by RNA-Seq.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2164-10-641-S8.XLS]

Additional file 9*Distribution of quality scores for all RNA-Seq runs. The quality of the RNA-Seq reads was analyzed using the correspondence between the quality score and error probability; these analyses were performed on Illumina RNA-Seq quality scores that were converted to phred format <http://www.phrap.com/phred/>.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2164-10-641-S9.TIFF]

Additional file 10*Genbank (gbk) file with ncRNAs identified here.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2164-10-641-S10.GBK]

Additional file 11*Coefficient of variation among RNA-Seq replicates by strain. (A) Histogram of the coefficient of variation (standard deviation/mean) for genes with GEI > 0 in both replicates for 10403S and Δ sigB strain. There is less variation between Δ sigB replicates compared to the 10403S replicates, but very few genes have a coefficient > 0.6. (B) Histogram depicting the GEI of one replicate for genes where the other replicate GEI = 0. The replicate GEI of the gene for which the other replicate is 0 (zero) is typically very low (GEI < 0.7).*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2164-10-641-S11.TIFF]

Additional file 12*Coverage file with the normalized RNA-Seq coverage for the 4 RNA-Seq runs.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2164-10-641-S12.ZIP]

Additional file 13 *σ^B -dependent promoters used for HMM search.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2164-10-641-S13.XLS]

Acknowledgements

This work was funded by NIH-NIAID (R01 AI052151 to K.J.B.). U.K. was supported by NSF (award no. 0644136). We thank P. Schweitzer and the staff at the Cornell DNA Sequencing and Genotyping Core Facility for sample preparation and sequencing and A. G. Clark and T. B. Sackton for helpful discussion.

References

1. Wang Z, Gerstein M, Snyder M: **RNA-Seq: a revolutionary tool for transcriptomics.** *Nat Rev Genet* 2009, **10(1)**:57-63.
2. Mead PS, Slutsker L, Dietz V, McCaig LF, Bresee JS, Shapiro C, Griffin PM, Tauxe RV: **Food-related illness and death in the United States.** *Emerg Infect Dis* 1999, **5(5)**:607-625.
3. Begley M, Gahan CG, Hill C: **Bile stress response in *Listeria monocytogenes* LO28: adaptation, cross-protection, and identification of genetic loci involved in bile resistance.** *Appl Environ Microbiol* 2002, **68(12)**:6005-6012.
4. Phan-Thanh L, Gormon T: **Analysis of heat and cold shock proteins in *Listeria* by two-dimensional electrophoresis.** *Electrophoresis* 1995, **16(3)**:444-450.
5. Watkins J, Sleath KP: **Isolation and enumeration of *Listeria monocytogenes* from sewage, sewage sludge and river water.** *J Appl Bacteriol* 1981, **50(1)**:1-9.
6. Chaturongakul S, Raengpradub S, Wiedmann M, Boor KJ: **Modulation of stress and virulence in *Listeria monocytogenes*.** *Trends Microbiol* 2008, **16(8)**:388-396.
7. Kazmierczak MJ, Wiedmann M, Boor KJ: **Alternative sigma factors and their roles in bacterial virulence.** *Microbiol Mol Biol Rev* 2005, **69(4)**:527-543.
8. Piggott PJ, Hilbert DW: **Sporulation of *Bacillus subtilis*.** *Curr Opin Microbiol* 2004, **7(6)**:579-586.
9. Wiedmann M, Arvik TJ, Hurley RJ, Boor KJ: **General stress transcription factor σ^B and its role in acid tolerance and virulence of *Listeria monocytogenes*.** *J Bacteriol* 1998, **180(14)**:3650-3656.
10. Raengpradub S, Wiedmann M, Boor KJ: **Comparative analysis of the σ^B -dependent stress responses in *Listeria monocytogenes* and *Listeria innocua* strains exposed to selected stress conditions.** *Appl Environ Microbiol* 2008, **74(1)**:158-171.
11. Hain T, Hossain H, Chatterjee SS, Machata S, Volk U, Wagner S, Brors B, Haas S, Kuenne CT, Billion A, et al.: **Temporal transcriptomic analysis of the *Listeria monocytogenes* EGD-e σ^B regulon.** *BMC Microbiol* 2008, **8**:20.
12. Ollinger J, Bowen B, Wiedmann M, Boor KJ, Bergholtz TM: ***Listeria monocytogenes* σ^B modulates PrfA-mediated virulence factor expression.** *Infect Immun* 2009, **77(5)**:2113-2124.
13. Asmann YW, Wallace MB, Thompson EA: **Transcriptome profiling using next-generation sequencing.** *Gastroenterology* 2008, **135(5)**:1466-1468.
14. Mockler TC, Ecker JR: **Applications of DNA tiling arrays for whole-genome analysis.** *Genomics* 2005, **85(1)**:1-15.
15. Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, Snyder M: **The transcriptional landscape of the yeast genome defined by RNA sequencing.** *Science* 2008, **320(5881)**:1344-1349.
16. Yoder-Himes DR, Chain PSG, Zhu Y, Wurtzel O, Rubin EM, Tiedje JM, Sorek R: **Mapping the *Burkholderia cenocepacia* niche response via high-throughput sequencing.** *Proc Natl Acad Sci USA* 2009, **106(10)**:3976-3981.
17. Schmittgen TD, Lee EJ, Jiang J, Sarkar A, Yang L, Elton TS, Chen C: **Real-time PCR quantification of precursor and mature microRNA.** *Methods* 2008, **44(1)**:31-38.

18. Glaser P, Frangeul L, Buchrieser C, Rusniok C, Amend A, Baquero F, Berche P, Bloecker H, Brandt P, Chakraborty T, et al.: **Comparative genomics of *Listeria* species.** *Science* 2001, **294(5543)**:849-852.
19. Christiansen JK, Nielsen JS, Ebersbach T, Valentin-Hansen P, Sogaard-Andersen L, Kallipolitis BH: **Identification of small Hfq-binding RNAs in *Listeria monocytogenes*.** *RNA (NY)* 2006, **12(7)**:1383-1396.
20. Toledo-Arana A, Dussurget O, Nikitas G, Sesto N, Guet-Revillet H, Balestrino D, Loh E, Gripenland J, Tiensuu T, Vaitkevicius K, et al.: **The *Listeria* transcriptional landscape from saprophytism to virulence.** *Nature* 2009, **459**:950-956.
21. Nielsen JS, Olsen AS, Bonde M, Valentin-Hansen P, Kallipolitis BH: **Identification of a sigma B-dependent small noncoding RNA in *Listeria monocytogenes*.** *J Bacteriol* 2008, **190(18)**:6264-6270.
22. Mandin P, Repoila F, Vergassola M, Geissmann T, Cossart P: **Identification of new noncoding RNAs in *Listeria monocytogenes* and prediction of mRNA targets.** *Nucleic Acids Res* 2007, **35(3)**:962-974.
23. Kazmierczak MJ, Mithoe SC, Boor KJ, Wiedmann M: ***Listeria monocytogenes* σ^B regulates stress response and virulence functions.** *J Bacteriol* 2003, **185(19)**:5722-5734.
24. Shetron-Rama LM, Mueller K, Bravo JM, Bower HG, Way SS, Freitag NE: **Isolation of *Listeria monocytogenes* mutants with high-level in vitro expression of host cytosol-induced gene products.** *Mol Microbiol* 2003, **48(6)**:1537-1551.
25. McGann P, Raengpradub S, Ivanek R, Wiedmann M, Boor KJ: **Differential regulation of *Listeria monocytogenes* internalin and internalin-like genes by σ^B and PrfA as revealed by subgenomic microarray analyses.** *Foodborne Pathog Dis* 2008, **5(4)**:417-435.
26. Mueller KJ, Freitag NE: **Pleiotropic enhancement of bacterial pathogenesis resulting from the constitutive activation of the *Listeria monocytogenes* regulatory factor PrfA.** *Infect Immun* 2005, **73(4)**:1917-1926.
27. Kim H, Marquis H, Boor KJ: **σ^B contributes to *Listeria monocytogenes* invasion by controlling expression of *inlA* and *inlB*.** *Microbiology* 2005, **151(Pt 10)**:3215-3222.
28. Mao C, Evans C, Jensen RV, Sobral BV: **Identification of new genes in *Sinorhizobium meliloti* using the Genome Sequencer FLX system.** *BMC Microbiol* 2008, **8**:72.
29. Liu JM, Livny J, Lawrence MS, Kimball MD, Waldor MK, Camilli A: **Experimental discovery of sRNAs in *Vibrio cholerae* by direct cloning, 5S/tRNA depletion and parallel sequencing.** *Nucleic Acids Res* 2009, **37(6)**:e46-.
30. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B: **Mapping and quantifying mammalian transcriptomes by RNA-Seq.** *Nat Meth* 2008, **5(7)**:621-628.
31. Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I, Penkett CJ, Rogers J, Bahler J: **Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution.** *Nature* 2008, **453(7199)**:1239-1243.
32. Chan YC, Boor KJ, Wiedmann M: **σ^B -dependent and σ^B -independent mechanisms contribute to transcription of *Listeria monocytogenes* cold stress genes during cold shock and cold growth.** *Appl Environ Microbiol* 2007, **73(19)**:6019-6029.
33. Kazmierczak MJ, Wiedmann M, Boor KJ: **Contributions of *Listeria monocytogenes* σ^B and PrfA to expression of virulence and stress response genes during extra- and intracellular growth.** *Microbiology* 2006, **152(6)**:1827-1838.
34. Tjaden B, Saxena RM, Stolyar S, Haynor DR, Kolker E, Rosenow C: **Transcriptome analysis of *Escherichia coli* using high-density oligonucleotide probe arrays.** *Nucleic Acids Res* 2002, **30(17)**:3732-3738.
35. Sultan M, Schulz MH, Richard H, Magen A, Klingenhoff A, Scherf M, Seifert M, Borodina T, Soldatov A, Parkhomchuk D, et al.: **A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome.** *Science* 2008, **321(5891)**:956-960.
36. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y: **RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays.** *Genome Res* 2008, **18(9)**:1509-1517.
37. Gouy M, Gautier C: **Codon usage in bacteria: correlation with gene expressivity.** *Nucl Acids Res* 1982, **10(22)**:7055-7074.
38. Ikemura T: **Codon usage and tRNA content in unicellular and multicellular organisms.** *Mol Biol Evol* 1985, **2(1)**:13-34.
39. Kanaya S, Yamada Y, Kudo Y, Ikemura T: **Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of *Bacillus subtilis* tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis.** *Gene* 1999, **238(1)**:143-155.
40. Keiler KC: **Biology of trans-translation.** *Ann Rev Microbiol* 2008, **62**:133-151.
41. Trotochaud AE, Wassarman KM: **6S RNA function enhances long-term cell survival.** *J Bacteriol* 2004, **186(15)**:4978-4985.
42. Loewen PC, Hengge-Aronis R: **The role of the sigma factor sigma S (KatF) in bacterial global regulation.** *Annu Rev Microbiol* 1994, **48**:53-80.
43. Archambaud C, Nahori MA, Pizarro-Cerda J, Cossart P, Dussurget O: **Control of *Listeria* superoxide dismutase by phosphorylation.** *J Biol Chem* 2006, **281(42)**:31812-31822.
44. Chan YC, Raengpradub S, Boor KJ, Wiedmann M: **Microarray-based characterization of the *Listeria monocytogenes* cold regulon in log- and stationary-phase cells.** *Appl Environ Microbiol* 2007, **73(20)**:6484-6498.
45. Graumann PL, Marahiel MA: **Cold shock proteins CspB and CspC are major stationary-phase-induced proteins in *Bacillus subtilis*.** *Arch Microbiol* 1999, **171(2)**:135-138.
46. Jin B, Newton SM, Shao Y, Jiang X, Charbit A, Klebba PE: **Iron acquisition systems for ferric hydroxamates, haem and haemoglobin in *Listeria monocytogenes*.** *Mol Microbiol* 2006, **59(4)**:1185-1198.
47. Olsen KN, Larsen MH, Gahan CGM, Kallipolitis B, Wolf XA, Rea R, Hill C, Ingmer H: **The Dps-like protein Fri of *Listeria monocytogenes* promotes stress tolerance and intracellular multiplication in macrophage-like cells.** *Microbiology* 2005, **151(3)**:925-933.
48. Schmid B, Klumpp J, Raimann E, Loessner MJ, Stephan R, Tasara T: **Role of cold shock proteins (Csp) for growth of *Listeria monocytogenes* under cold and osmotic stress conditions.** *Appl Environ Microbiol* 2009, **75(6)**:1621-1627.
49. Vasconcelos JA, Deneer HG: **Expression of superoxide dismutase in *Listeria monocytogenes*.** *Appl Environ Microbiol* 1994, **60(7)**:2360-2366.
50. Bigot A, Pagniez H, Botton E, Frehel C, Dubail I, Jacquet C, Charbit A, Raynaud C: **Role of FliF and FliI of *Listeria monocytogenes* in flagellar assembly and pathogenicity.** *Infect Immun* 2005, **73(9)**:5530-5539.
51. Way SS, Thompson LJ, Lopes JE, Hajjar AM, Kollmann TR, Freitag NE, Wilson CB: **Characterization of flagellin expression and its role in *Listeria monocytogenes* infection and immunity.** *Cell Microbiol* 2004, **6(3)**:235-242.
52. Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, Hotz H-R, Ceric G, Forslund K, Eddy SR, Sonnhammer ELL, et al.: **The Pfam protein families database.** *Nucl Acids Res* 2008, **36(suppl_1)**:D281-288.
53. Nielsen JS, Olsen AS, Bonde M, Valentin-Hansen P, Kallipolitis BH: **Identification of a σ^B -dependent small noncoding RNA in *Listeria monocytogenes*.** *J Bacteriol* 2008, **190(18)**:6264-6270.
54. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR: **Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*.** *Cell* 2008, **133(3)**:523-536.
55. Morin R, Bainbridge M, Fejes A, Hirst M, Krzywinski M, Pugh T, McDonald H, Varhol R, Jones S, Marra M: **Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing.** *BioTechniques* 2008, **45(1)**:81-94.
56. Chai Y, Kolter R, Losick R: **A widely conserved gene cluster required for lactate utilization in *Bacillus subtilis* and its involvement in biofilm formation.** *J Bacteriol* 2009, **191(8)**:2423-2430.
57. Hillerich B, Westpheling J: **A new GntR family transcriptional regulator in *Streptomyces coelicolor* is required for morphogenesis and antibiotic production and controls transcription of an ABC transporter in response to carbon source.** *J Bacteriol* 2006, **188(21)**:7477-7487.
58. Ogasawara H, Ishida Y, Yamada K, Yamamoto K, Ishihama A: **PdhR (pyruvate dehydrogenase complex regulator) controls the respiratory electron transport system in *Escherichia coli*.** *J Bacteriol* 2007, **189(15)**:5534-5541.
59. Ferreira A, O'Byrne CP, Boor KJ: **Role of σ^B in heat, ethanol, acid, and oxidative stress resistance and during carbon starvation in *Listeria monocytogenes*.** *Appl Environ Microbiol* 2001, **67(10)**:4454-4457.

60. Moorhead SM, Dykes GA: **The role of the *sigB* gene in the general stress response of *Listeria monocytogenes* varies between a strain of serotype 1/2a and a strain of serotype 4c.** *Curr Microbiol* 2003, **46(6)**:461-466.
61. Garner MR, Njaa BL, Wiedmann M, Boor KJ: **Sigma B contributes to *Listeria monocytogenes* gastrointestinal infection but not to systemic spread in the guinea pig infection model.** *Infect Immun* 2006, **74(2)**:876-886.
62. Sleator RD, Clifford T, Hill C: **Gut osmolarity: A key environmental cue initiating the gastrointestinal phase of *Listeria monocytogenes* infection?** *Med Hypotheses* 2007, **69(5)**:1090-1092.
63. Sue D, Boor KJ, Wiedmann M: **σ^B -dependent expression patterns of compatible solute transporter genes *opuCA* and *lmo1421* and the conjugated bile salt hydrolase gene *bsh* in *Listeria monocytogenes*.** *Microbiology* 2003, **149(Pt 11)**:3247-3256.
64. Sue D, Fink D, Wiedmann M, Boor KJ: **σ^B -dependent gene induction and expression in *Listeria monocytogenes* during osmotic and acid stress conditions simulating the intestinal environment.** *Microbiology* 2004, **150(Pt 11)**:3843-3855.
65. Core LJ, Waterfall JJ, Lis JT: **Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters.** *Science* 2008, **322(5909)**:1845-1848.
66. Cetin MS, Zhang C, Hutkins RW, Benson AK: **Regulation of transcription of compatible solute transporters by the general stress sigma factor, σ^B , in *Listeria monocytogenes*.** *J Bacteriol* 2004, **186(3)**:794-802.
67. Fraser KR, Sue D, Wiedmann M, Boor K, O'Byrne CP: **Role of σ^B in regulating the compatible solute uptake systems of *Listeria monocytogenes*: osmotic induction of *opuC* is σ^B dependent.** *Appl Environ Microbiol* 2003, **69(4)**:2015-2022.
68. Kingsford C, Ayanbule K, Salzberg S: **Rapid, accurate, computational discovery of Rho-independent transcription terminators illuminates their relationship to DNA uptake.** *Genome Biol* 2007, **8(2)**:R22.
69. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B: **Artemis: sequence visualization and annotation.** *Bioinformatics* 2000, **16(10)**:944-945.
70. Storey JD, Tibshirani R: **Statistical significance for genomewide studies.** *Proc Natl Acad Sci USA* 2003, **100(16)**:9440-9445.
71. Crooks GE, Hon G, Chandonia JM, Brenner SE: **WebLogo: a sequence logo generator.** *Genome Res* 2004, **14(6)**:1188-1190.
72. Rice P, Longden I, Bleasby A: **EMBOSS: the european molecular biology open software suite.** *Trends Genet* 2000, **16(6)**:276-277.
73. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *J Royal Stat Soc B* 1995, **57(1)**:289-300.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

