molecular systems biology

Comparative transcriptomics of pathogenic and non-pathogenic Listeria species

Omri Wurtzel, Nina Sesto, Jeff R. Mellin, Iris Karunker, Sarit Edelheit, Christophe Becavin, Cristel Archambaud, Pascale Cossart, Rotem Sorek

Corresponding author: Rotem Sorek, Weizmann Institute of Science

Review timeline:

Submission date:	04 January 2012
Editorial Decision:	17 February 2012
Revision received:	07 March 2012
Accepted:	12 March 2012

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

1st Editorial Decision 17 February 2012

Dear Dr Sorek,

Thank you again for submitting your work to Molecular Systems Biology. We have now heard back from the two of the three referees who accepted to evaluate the study. Given that their recommendations are very similar, I prefer to make a decision now rather than delaying further the process. As you will see, the referees find the topic of your study of potential interest and are supportive. They raise however a series of concerns and make suggestions for modifications, which should be convincingly addressed in a revision of the present work. The recommendations provided by the reviewers are very clear in this regard.

We would also kindly ask you to prepare supplementary information according to our guidelines (http://www.nature.com/msb/authors/index.html#a3.4.6). Essentially, supplementary text and figures should be merged in a single PDF file starting with a table of content. Each figure legend should be written clearly beneath its respective figure.

Please include accession numbers of the datasets described in this study in Materials & Methods, if possible in a dedicated sub-section labeled "Data availability". The suggested wording for referring to accession identifiers in a manuscript is the following: "The [protein interaction | microarray | mass spectrometry] data from this publication have been submitted to the [name of the database] database [URL] and assigned the identifier [accession | permalink | hashtag]." (http://www.nature.com/msb/authors/index.html#a3.5)

*** PLEASE NOTE *** As part of the EMBO Publications transparent editorial process initiative (see our Editorial at http://www.nature.com/msb/journal/v6/n1/full/msb201072.html), Molecular Systems Biology will publish online a Review Process File to accompany accepted manuscripts. When preparing your letter of response, please be aware that in the event of acceptance, your cover

© European Molecular Biology Organization

letter/point-by-point document will be included as part of this File, which will be available to the scientific community. More information about this initiative is available in our Instructions to Authors. If you have any questions about this initiative, please contact the editorial office (msb@embo.org).

Please resubmit your revised manuscript online, with a covering letter listing amendments and responses to each point raised by the referees.

Thank you for submitting this paper to Molecular Systems Biology.

Yours sincerely,

Editor

Molecular Systems Biology

http://www.nature.com/msb

Referee Reports:

Reviewer #1 (Remarks to the Author):

This manuscript from the Sorek laboratory reports the--to the best of my knowledge--first comparative analysis of transcriptome structure of two related bacteria, the pathogenic Listeria monocytogenes and its non-pathogenic sibling, Listeria innocua. The authors use a specialized strand-specific RNA-seq protocol to build a global of map transcription start sites (TSS) in both species, and subsequently analyze and compare the data sets with respect to 5' UTR length, potential consensus elements in promoter regions, as well as noncoding transcripts.

The comparative interrogation of the two Listeria species reveals dozens of new small RNAs, including their differential expression, as well as a numerous new long antisense RNAs, a class of transcripts that have generally been underrated in bacteria. Some of these new long antisense RNAs are likely to decide the fate of expression of physically adjacent and functional-linked Listeria genes, in a mutually exclusive fashion, which leads the authors to propose a concept of a bacterial excludon.

Overall, this is a carefully-designed study that provides much new insight into the structure of bacterial transcriptomes, many aspects of which will also be important for model bacteria beyond Listeria. The manuscript is well-written, and I shall gladly recommend it be published in MSB following some revision.

Major points

1. The authors use a protocol that involves differential treatment of RNA with tobacco acid pyrophosphatase to evaluate the 5' status of transcripts (5'PPP would be primary 5' end; 5'P or 5'OH mark the end of processed species), and infer from this transcription start sites. How accurate are the calls of primary 5' ends? How many of the newly discovered TSS have been validated by independent methods?

2. Given the massive number of newly determined TSS, the authors should look for consensus motifs in Listeria promoters.

3. In Figure 3, the Northern blot seems to indicate the Rli130 sRNA is not expressed in L. innocua. Is this because of ineffective termination, or does the Rli130 locus of L. innocua have a transcriptional organization that deviates from L. monocytogenes?

4. The discussion of candidate mRNA targets predicted using the TargetRNA algorithm (middle of page 18) makes the assumption that Rli87 is a regulatory small RNA that acts by base-pairing, for which there is presently no evidence. This should be clarified in the text.

© European Molecular Biology Organization

Minor points

5. Page 3, line 10 in Introduction: no need to use past perfect (... and it had been at first estimated ...).

6. Page 4, line 6: Consider adding the fairly comprehensive review by Papenfort & Vogel in Cell Host & Microbe (2010 Jul 22;8(1):116-27) to the list of recent reviews.

7. Figure 1. I understand the authors employed a strand-specific RNA sequencing protocol, but why doesn't the browser show the strand-specific information?

8. Figure 2, comparing panels A and B. There seem to be great differences in the 5' UTR distribution between L. monocytogenes and L. innoucia, such that the former species expresses more leaderless mRNAs. What are these leaderless, and can they be functionally classified somehow?

9. Page 10, first paragraph. The 5' UTR of lmo1654 is 73nt longer than that of lin1694. What causes the difference? Promoter shifting, sequence insertion/deletion? It would be great to show the sequence alignment, and the difference in predicted secondary structures (which is alluded to Discussion on Page 18).

10. Figures 3 & 4. It would be nice to add simple plots to summarize the number of sRNAs and asRNAs that are specific to lmo and/or lin. It might also help if Figure S2 were moved to the main text, provided that the expression patterns of more sRNAs were validated by independent methods.

11. In Figure 3, it would be more consistent to illustrate the predicted terminators with black color.

12. Figure 4 showing expression patterns of long antisense RNAs is little informative unless the associated mRNAs are also shown. In other words, what is the correlation of expression of the transcripts from the opposite strands?

13. Right panel of Fig 6A: Anti1845 should be anti1846.

14. Page 21, last paragraph: Sharma et al. 2010 detected in Helicobacter antisense TSS for ~40% of all genes, thus more than in most of the species listed.

15. Next time, please include page numbers.

Reviewer #3 (Remarks to the Author):

Wurtzel et al describe a comparative analysis of the transcriptome of two strains of Listeria (one pathogenic and the other non-pathogenic). Combining newly generated transcription start site analysis with prior tiling array data the authors have developed a web browser that will allow researchers to compare the transcriptomes of each strain. The authors also report the identification of several start sites that appear to produce long transcripts that are antisense to genes on the opposite strand and that contain one or more genes on the same strand (downstream of a long 5' UTR). The authors show that in at least one case, transcription of the long transcript reduces expressing of the gene on the opposite strand and increases the expression of genes on the same strand. The work presented here and the web browser will be of use to the field and I recommend publication.
Comments to address:
- The start site analysis was done only for a single sample. Biological replicates would have been preferable.
- Figure 1 has panels A and B but it is unclear why.
Figure 2D- In cases such as this, where genes appear to have distinct 5' UTRs in the different strains it would be useful to know if there are sequence differences in the respective promoter regions that can account for these differences.
- The abstract should make it clear how many of the 113 sRNAs and 70asRNAs are newly identified.

© European Molecular Biology Organization

- In methods section (last paragraph of "Determination of TSSs and comparison of 5' UTRs length") it states "...; the higher number of 5' ends of this kind in L. innocua is the result of less sequencing reads... the words: "for the L. monocytogenes samples" should be added.

© European Molecular Biology Organization

Reviewer #1 (Remarks to the Author):

Major points

1. The authors use a protocol that involves differential treatment of RNA with tobacco acid pyrophosphatase to evaluate the 5' status of transcripts (5'PPP would be primary 5' end; 5'P or 5'OH mark the end of processed species), and infer from this transcription start sites. How accurate are the calls of primary 5' ends? How many of the newly discovered TSS have been validated by independent methods?

Answer: We compared our results with a set of genes having long 5' UTRs for which the UTR length was determined by other methods (summarized in a review by Loh et al, Trends in Microbiology, 2006). In eight of 9 such genes that were expressed in our study, the 5' UTR size determined by our method deviated by 0-2 bp from the previously determined 5' UTR, with the 9th gene deviating by 6 nt. Considering that these UTR sizes are 100 bp on average, these results confirm the accuracy of our TSS determination. These data are now included in Supplementary Table S3, and we added the following sentence to the results text: "Comparison to a set of 9 TSSs previously determined by other studies (Loh et al, 2006) validated the accuracy of our TSS determination (Supplementary Table S3)".

2. Given the massive number of newly determined TSS, the authors should look for consensus motifs in Listeria promoters.

Answer: We thank the reviewer for this point. To determine the consensus motifs we performed a PWM promoter analysis for two sets of genes: highly expressed genes and sigB-regulated genes. Within the highly expressed genes, our analysis revealed a conserved -10 box motif, while the promoters of sigB regulated genes clearly showed a sigB consensus promoter motif (GTTT-N_{12-17}-GGGWWW). We added these results as Supplementary Figure S2. The results text now reads: "Further analysis of sequences upstream to the determined TSSs revealed the existence of sigma-like consensus promoter motifs in the expected upstream positions (Supplementary Figure S2)."

3. In Figure 3, the Northern blot seems to indicate the Rli130 sRNA is not expressed in L. innocua. Is this because of ineffective termination, or does the Rli130 locus of L. innocua have a transcriptional organization that deviates from L. monocytogenes?

Answer: We thank the reviewer noticing this point. We removed from Figure 3C the illustration that suggested the smaller form of Rli130 in *L. innocua*. Sequence analysis of that locus showed divergence at the 3' end of Rli130, such that bases 179 to 230 (52 nt)

contain 11 mismatches (as opposed to the high conservation in the 5' end of that sRNA). This divergence might have caused the observed difference in expression. To clarify this point, we added the following text to the legend of Figure 3: "In *L. innocua*, the shortest form of Rli130 (230 nt) is missing, possibly due to sequence divergence between the two organisms (11 differences between positions 179 and 230 of that gene)."

4. The discussion of candidate mRNA targets predicted using the TargetRNA algorithm (middle of page 18) makes the assumption that Rli87 is a regulatory small RNA that acts by base-pairing, for which there is presently no evidence. This should be clarified in the text.

Answer: To clarify this point, we added to the text the following reservation: “It is tempting to speculate that Rli87 functions by base-pairing to InlA and that the loss of its expression…”.

Minor points

5. Page 3, line 10 in Introduction: no need to use past perfect (... and it had been at first estimated ...).

Answer: Text was changed from “had been” to “was”.

6. Page 4, line 6: Consider adding the fairly comprehensive review by Papenfort & Vogel in Cell Host & Microbe (2010 Jul 22;8(1):116-27) to the list of recent reviews.

Answer: The citation was added to the text.

7. Figure 1. I understand the authors employed a strand-specific RNA sequencing protocol, but why doesn't the browser show the strand-specific information?

Answer: The strand-specific TSS-mapping and tiling arrays appear as arrows pointing to the directionality of the transcript (right and left, forward and reverse strands, respectively). This is indicated in the browser legend that can be presented by selecting the checkbox marked "legend" within the browser, and also in Figure 1. The whole-gene sequencing appearing in the browser is taken from Oliver et al, *BMC Genomics*, 2009, which did not use strand-specific RNA-seq.

8. Figure 2, comparing panels A and B. There seem to be great differences in the 5' UTR distribution between L. monocytogenes and L. innoucia, such that the former species expresses more leaderless mRNAs. What are these leaderless, and can they be functionally classified somehow?

Answer: We searched for gene functions that were enriched in genes lacking 5’ UTR in *L. monocytogenes*, but no significant results were found. We used Kolmogorov-Smirnov test on genes with annotated TSS in both organisms to assess whether the lengths of the

UTRs represented different distributions, a test that resulted in insignificant result (p=0.9997).

9. Page 10, first paragraph. The 5' UTR of lmo1654 is 73nt longer than that of lin1694. What causes the difference? Promoter shifting, sequence insertion/deletion? It would be great to show the sequence alignment, and the difference in predicted secondary structures (which is alluded to Discussion on Page 18).

Answer: The sequence in the intergenic region between *lmo1654* and *lmo1655* is very different from the corresponding intergenic region in *L. innocua*. The 3' ends of the two UTRs show similarity, while the 5' ends and their promoter sequences show significant differences. This indeed causes changes to the predicted secondary structures of the 5' UTRs. As suggested by the reviewer, we added the sequence alignment and the predicted structures as Supplementary Figure S3.

10. Figures 3 & 4. It would be nice to add simple plots to summarize the number of sRNAs and asRNAs that are specific to lmo and/or lin. It might also help if Figure S2 were moved to the main text, provided that the expression patterns of more sRNAs were validated by independent methods.

Answer: Table 1 of the manuscript contains the numbers of conserved and sequence-specific sRNAs and asRNAs as asked for by the reviewer. Since the primary text already contains 6 figures, we think that Figure S2 should be kept in the Supplementary material.

11. In Figure 3, it would be more consistent to illustrate the predicted terminators with black color.

Answer: For homogeneity with other examples, we removed the representation of predicted terminators from the Figure 3.

12. Figure 4 showing expression patterns of long antisense RNAs is little informative unless the associated mRNAs are also shown. In other words, what is the correlation of expression of the transcripts from the opposite strands?

Answer: The intention of Figure 4 was to validate the expression of the new asRNAs, rather than determine their functionality, and therefore mRNAs expression was not co-analyzed. The functions of selected asRNAs was indeed further validated by detailed analysis of the cognate sense mRNA as shown in Figure 5.

13. Right panel of Fig 6A: Anti1845 should be anti1846.

Answer: Fixed.

14. Page 21, last paragraph: Sharma et al. 2010 detected in Helicobacter antisense TSS for ~40% of all genes, thus more than in most of the species listed.

Answer: We modified the text and it now reads: "…in several organisms such as *Helicobacter pylori* (Sharma et al, 2010),…"

Reviewer #3 (Remarks to the Author):

Comments to address:

1. **The start site analysis was done only for a single sample. Biological replicates would have been preferable.**

Answer: To assess the reproducibility of our TSS detection we compared the TSSs determined in WT *L. monocytogenes*, ΔsigB and ΔprfA strains, all of which were grown to mid-log phase at 37°C. Since the genes regulated by SigB and PrfA are largely known, we expect that apart from these genes the TSSs should largely reproduce between the 3 strains as they were grown in the same conditions. We found that over 98% of the TSSs supported by 10 reads or more (our threshold for TSS detection, Methods) in the WT strain reproduced to exactly the same nucleotide in the ΔprfA strain as well as 97% reproducibility in ΔsigB. These results demonstrate high reproducibility between biological samples. We further performed a Spearman correlation test of 5' UTRs lengths determined through TSS detection. In all cases, the correlation between each pair of samples was 1, indicating again the high reproducibility of our TSS detection.

To clarify this point the following text was added to the Materials and Methods: "To assess the reproducibility of the TSS-mapping the positions of TSSs in WT and ΔsigB *L. monocytogenes* grown to mid-log phase at 37°C were compared. The TSS-mapping of genes known to be independent of SigB (Oliver, 2009) reproduced to a single-nucleotide resolution in 97% of the TSSs supported by over 10 reads in the WT bacteria. Additional comparison of WT and the ΔprfA TSSs showed a reproducibility of over 98%."

We also compared our results with a set of genes having long 5' UTRs for which the UTR length was previously determined by other methods (summarized in a review by Loh et al, *Trends in Microbiology*, 2006). In eight of 9 such genes that were expressed in our study, the 5' UTR size determined by our method deviated by 0-2 nt from the previously determined 5' UTR, with the 9th gene deviating by 6 nt. Considering that these UTR sizes are 100 bp on average, these results confirm the accuracy of our TSS determination. These data are now included in Supplementary Table S3, and we added the following sentence to the results text: "Comparison to a set of 9 TSSs previously determined by other studies (Loh et al 2006) validated the accuracy of our TSS determination (Supplementary Table S3).

2. **Figure 1 has panels A and B but it is unclear why.**

Answer: Our intention is to show the major genomic features presented in the viewer, including TSSs of ORFs, ncRNAs and asRNAs. Since these features are found in different genome locations (e.g. panel A does not contain an asRNA) it was necessary to present these data in two panels. To clarify this point, we added the following sentence to the Figure legend: "Panels A, B, and C represent different windows of the genome in the unified browser." In addition, we now added panel C, to highlight the representation of operons in the browser, which was missing from the previous figure.

3. **Figure 2D- In cases such as this, where genes appear to have distinct 5' UTRs in the different strains it would be useful to know if there are sequence differences in the respective promoter regions that can account for these differences.**

Answer: As suggested by both reviewers, we added the sequence alignment and the predicted secondary structures of the two 5' UTRs, as well as the promoter region, as Supplementary Figure S3.

4. **The abstract should make it clear how many of the 113 sRNAs and 70asRNAs are newly identified.**

Answer: The numbers were added to the abstract.

5. **In methods section (last paragraph of "Determination of TSSs and comparison of 5' UTRs length") it states "...; the higher number of 5' ends of this kind in L. innocua is the result of less sequencing reads... the words: "for the L. monocytogenes samples" should be added.**

Answer: We added “than in the *L. monocytogenes* samples” to that sentence.