

Listeriomics: an Interactive Web Platform for Systems Biology of Listeria

Christophe Bécavin, Mikael Koutero, Nicolas Tchitchek, Franck Cerutti, Pierre Lechat, Nicolas Maillet, Claire Hoede, Hélène Chiapello, Christine Gaspin, Pascale Cossart

▶ **To cite this version:**

Christophe Bécavin, Mikael Koutero, Nicolas Tchitchek, Franck Cerutti, Pierre Lechat, et al.. Listeriomics: an Interactive Web Platform for Systems Biology of Listeria. mSystems, 2017, 2 (2), pp.e00186-16. 10.1128/mSystems.00186-16 . pasteur-01574970

HAL Id: pasteur-01574970

https://pasteur.hal.science/pasteur-01574970

Submitted on 17 Aug 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution| 4.0 International License

AMERICAN SOCIETY FOR MICROBIOLOGY | mSystems®

RESOURCE REPORT
Host-Microbe Biology

Listeriomics: an Interactive Web Platform for Systems Biology of *Listeria*

Christophe Bécavin,[a,b,c,d] Mikael Koutero,[a,b,c] Nicolas Tchitchek,[e] Franck Cerutti,[f] Pierre Lechat,[d] Nicolas Maillet,[d] Claire Hoede,[f] Hélène Chiapello,[f] Christine Gaspin,[f] Pascale Cossart[a,b,c]

Département de Biologie Cellulaire et Infection, Institut Pasteur, Unité des Interactions Bactéries-Cellules, Paris, France[a]; INSERM, U604, Paris, France[b]; INRA, USC2020, Paris, France[c]; Institut Pasteur–Bioinformatics and Biostatistics Hub, C3BI, USR 3756 IP CNRS, Paris, France[d]; CEA-Division of Immuno-Virology, Institute of Emerging Diseases and Innovative Therapies, IDMIT Center, Fontenay-aux-Roses, France[e]; INRA-UR MIAT, Université de Toulouse, Castanet-Tolosan, France[f]

ABSTRACT As for many model organisms, the amount of *Listeria* omics data produced has recently increased exponentially. There are now >80 published complete *Listeria* genomes, around 350 different transcriptomic data sets, and 25 proteomic data sets available. The analysis of these data sets through a systems biology approach and the generation of tools for biologists to browse these various data are a challenge for bioinformaticians. We have developed a web-based platform, named Listeriomics, that integrates different tools for omics data analyses, i.e., (i) an interactive genome viewer to display gene expression arrays, tiling arrays, and sequencing data sets along with proteomics and genomics data sets; (ii) an expression and protein atlas that connects every gene, small RNA, antisense RNA, or protein with the most relevant omics data; (iii) a specific tool for exploring protein conservation through the *Listeria* phylogenomic tree; and (iv) a coexpression network tool for the discovery of potential new regulations. Our platform integrates all the complete *Listeria* species genomes, transcriptomes, and proteomes published to date. This website allows navigation among all these data sets with enriched metadata in a user-friendly format and can be used as a central database for systems biology analysis.

IMPORTANCE In the last decades, *Listeria* has become a key model organism for the study of host-pathogen interactions, noncoding RNA regulation, and bacterial adaptation to stress. To study these mechanisms, several genomics, transcriptomics, and proteomics data sets have been produced. We have developed Listeriomics, an interactive web platform to browse and correlate these heterogeneous sources of information. Our website will allow listeriologists and microbiologists to decipher key regulation mechanism by using a systems biology approach.

KEYWORDS *Listeria*, transcriptomics, database, genomics, proteomics, systems biology

Received 1 December 2016 **Accepted** 2 February 2017 **Published** 14 March 2017

Citation Bécavin C, Koutero M, Tchitchek N, Cerutti F, Lechat P, Maillet N, Hoede C, Chiapello H, Gaspin C, Cossart P. 2017. Listeriomics: an interactive web platform for systems biology of *Listeria*. mSystems 2: e00186-16. https://doi.org/10.1128/mSystems.00186-16.

Editor David Fenyo, NYU School of Medicine

Copyright © 2017 Bécavin et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Christophe Bécavin, christophe.becavin@pasteur.fr, or Pascale Cossart, pascale.cossart@pasteur.fr.

#Listeria systems biology study with Listeriomics web platform - #genomics #transcriptomics #proteomics

Listeria monocytogenes is a foodborne pathogen responsible for foodborne infections with a mortality rate of 25%. This pathogen is responsible for gastroenteritis, sepsis, and meningitis and can cross three host barriers, the intestinal, placental, and blood-brain barriers. It is a major concern for pregnant women, as it induces abortions (1). *L. monocytogenes* can enter, replicate in, and survive in a wide range of human cell types, such as macrophages, epithelial cells, and endothelial cells. Moreover, *Listeria* has emerged as a model organism for the study of host-pathogen interactions (1–3).

Listeria belongs to the *Firmicutes* phylum. The *Listeria* genus is made up of the widely studied pathogenic species *L. monocytogenes*; another pathogenic species, *Listeria ivanovii*, that mostly affects ruminants; and 15 nonpathogenic species (4–10). In

Downloaded from http://msystems.asm.org/ on April 13, 2017 by guest

2001, the genomes of *L. monocytogenes* strain EGD-e and one *Listeria innocua* strain were sequenced (11). Since then, many other *Listeria* genomes, covering all the lineages, have been sequenced (12–17). Currently, the NCBI refSeq database contains 83 complete *Listeria* genomes, including 70 *L. monocytogenes* genomes. The number of *Listeria* strains sequenced will probably grow exponentially in the coming years. Efforts have been made to summarize all these genomes on specific databases like ListiList (11), GenoList (18), GECO-LisDB server (16), and ListeriaBase (19) to find common gene features and to develop pangenome studies of *Listeria* species.

The first *Listeria* transcriptomic data set was published in 2007 (20). Since that report, 64 ArrayExpress studies, corresponding to 362 different biological conditions, have been produced (21). Only seven of them are transcriptome sequencing (RNA-Seq) studies, and all the others correspond to transcription profiling by microarrays, with the EGD-e strain being the most frequently used strain. *Listeria* is also a key organism in the study of bacterial regulatory small noncoding RNAs (sRNAs). Despite the high number of studies on *Listeria* noncoding RNAs, only two websites with *Listeria*-related data sets have been published. The first one is a genome viewer published along with a transcription start site (TSS) study of *Listeria* (22). The second is the sRNAdb database (23), which provides tools to visualize the conservation of gene loci surrounding noncoding RNAs in different Gram-positive bacteria.

The ability of *L. monocytogenes* to enter into various types of cells is due to the variety of proteins it secretes or anchors to its cell wall and external membrane. Consequently, many proteomic studies have been performed to analyze the exoproteome of *Listeria* (24–35). Other studies have focused on cytoplasmic proteins (27, 33, 35–44). To our knowledge, 74 proteome studies have been conducted to decipher the production and localization of *Listeria* proteins. Nevertheless, no database exists that combines all these proteomics data sets into a single, user-friendly resource.

The number of omics data sets produced has increased exponentially. The number of tools to analyze these data, as well as the diversity of databases to store them, has also burgeoned. In parallel with this increase, many efforts have been made to develop accurate web-based tools to integrate diverse omics data for each model organism. One of the most complete resources is certainly the University of California Santa Cruz Encyclopedia of DNA Elements (ENCODE at UCSC) Genome Browser (45), which allows the visualization of a large variety of human and mouse omics data sets. For prokaryotic organisms, the BioCyc (46, 47) and Pathosystems Resource Integration Center (48) websites have been created. These websites connect all the published genomic and transcriptomic data sets for prokaryotic organisms to metabolic pathways. Such wide-ranging web resources are useful for microbiologists, but for in-depth analyses, the development of individual web resources with curated metadata per model organism is also required. In the case of bacteria, few heterogeneous omics data sets are available (49) and few model organisms have dedicated web resources, including *Escherichia coli*, with RegulonDB (50) and PortEco (51), and *Bacillus subtilis*, with SubtiWiki (52). As yet, resources for *Listeria* species are limited.

Here, we present Listeriomics (http://listeriomics.pasteur.fr/), a highly interactive web resource summarizing many omics data sets related to the genus *Listeria*. We have curated and integrated all the available *Listeria* transcriptomic, proteomic, and genomic data sets to date. The Listeriomics platform was developed not only to integrate these diverse data sets but also to display them in a single viewer. To interactively explore these data sets, our website also provides different tools, i.e., (i) a genome viewer for displaying gene expression arrays, tiling arrays, and sequencing data, along with proteomic and genomic data sets; (ii) an expression atlas and protein atlas, inspired by the EBI Expression Atlas, that connects genomic elements (genes, small RNAs, antisense RNAs [asRNAs]) to the most relevant omics data; (iii) a specific tool for exploring protein conservation through the *Listeria* phylogenomic tree; and (iv) a coexpression network analysis tool for the discovery of potential new regulations.

Downloaded from http://msystems.asm.org/ on April 13, 2017 by guest

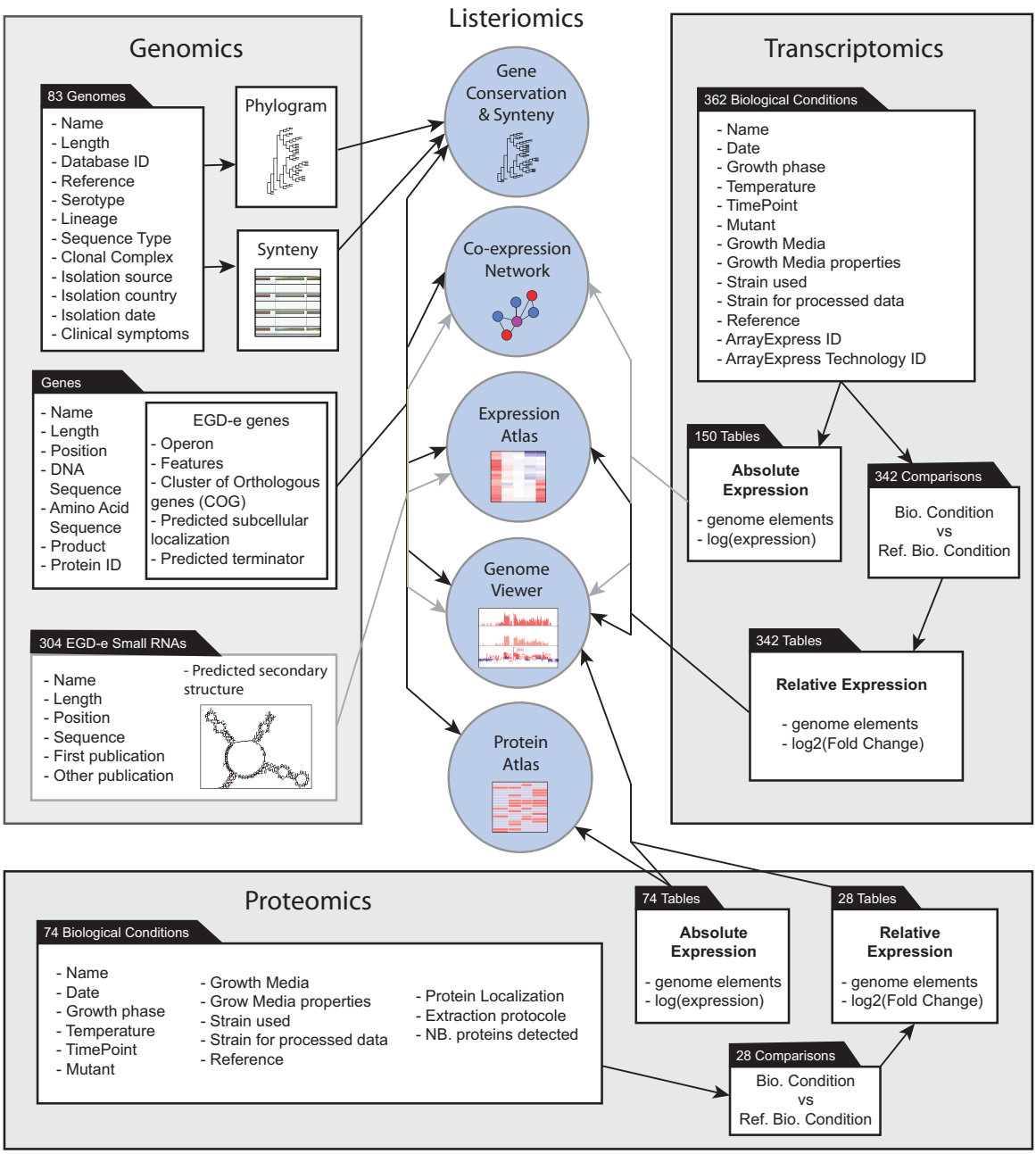


FIG 1 Overview of the Listeriomics platform. (Center) The five major tools of Listeriomics, i.e., gene conservation and synteny, coexpression network, genome viewer, expression, and protein atlas. (Left) Summary of all the available genomic information available on the website. (Right) List of all the transcriptomic information available in Listeriomics. (Bottom) View of all the proteomic information that can be accessed.

RESULTS

The Listeriomics web interface. Genomic, transcriptomic, or proteomic data can be browsed by using the Listeriomics website (http://listeriomics.pasteur.fr/) main page (Fig. 1; Table 1; see Fig. S1 in the supplemental material). For each type of data, we designed a summary panel to navigate through the different data sets. The top banner of the website gives direct access to them. As summarized in Table 1, users can search 83 complete *Listeria* genomes and browse 492 transcriptome and 74 proteome data sets. Listeriomics integrates four tools for omics data management, i.e., (i) a genome viewer for displaying gene expression array, tiling array, and sequencing data along with proteomics and genomics data; (ii) an expression atlas and protein atlas that

Downloaded from http://msystems.asm.org/ on April 13, 2017 by guest

TABLE 1 Summary of omics data sets included in the Listeriomics database

Category	Data sizes	Data type(s)	Tools available
Genomics	83 complete genomes (NCBI), all protein coding genes and noncoding RNAs, 304 small RNAs	Genome, phylogeny, genome elements, homologs	Genome summary, gene panel, small RNA panel, genome viewer
Transcriptomics	362 biological conditions, 8 *Listeria* strains, 342 comparisons	Gene expression array, tiling array, TSS, RNA-Seq	Transcriptome summary, expression atlas, heat map, genome viewer
Proteomics	74 biological conditions, 4 *Listeria* strains, 28 comparisons	Mass spectrometry	Proteome summary, protein atlas, heat map, genome viewer

connect every genomic element (genes, small RNAs, asRNAs) to the most relevant omics data; (iii) a protein conservation tool for the direct visualization of the presence or absence of a protein in a specific *Listeria* strain; and (iv) a coexpression network analysis tool for the visualization of genome features with the same expression profile.

The genomic interface is designed to browse every complete genome of the Listeriomics resource. Users can access strain name, serotype, lineage, and isolation information, along with a complete phylogenomic tree of *Listeria* strains (Fig. 1). From this table, scientists can access all the annotated genes of a specific strain. For each *Listeria* gene, five different information panels are available. The first panel shows all the general information about the position of the gene, its predicted annotated function. DNA and amino acid sequences can be accessed and saved as FASTA files or sent directly for a BLASTn or BLASTp search (53). The predicted subcellular localization (cytoplasm, cytoplasmic membrane, cell wall, cell surface, and extracellular milieu [27]) of each protein is also displayed along with information about the secretion pathway possibly used by the protein. The second panel provides an instant view of the conservation of a specific protein in other *Listeria* strains. This panel dynamically displays homologs on the *Listeria* reference tree in each existing *Listeria* strain. It also displays a summary table of all the homologous proteins with their similarity percentages and amino acid sequences. Users can also create a multialignment file of the homologous proteins. With the third panel, the user can visualize the protein locus synteny in all *Listeria* strains. We built an external synteny website by using the SynTView architecture (54). A fourth panel uses the expression atlas to show in which transcriptomics data sets the selected gene is differently expressed. The fifth panel displays every proteomics data set in which the protein encoded by the selected gene has been detected. Finally, from the home webpage, a summary panel with all the small RNAs in *L. monocytogenes* EGD-e can be accessed (Fig. 1). For each noncoding RNA element, one can display its position, its nucleotide sequence, its predicted secondary structure at 37°C, and a table displaying all supplementary information provided in source references (22, 55–58).

In the transcriptomic interface of the Listeriomics website, researchers can access all the *Listeria* transcriptomic data sets published so far. A searchable table shows every data set available with a precise description of the biological conditions studied. In total, four different transcriptome technologies (gene expression array, tiling array, RNA-Seq, and TSS) are included for seven different *L. monocytogenes* strains grown in four different broth media under seven intracellular conditions (Fig. 2A). Once transcriptomic data sets have been selected, it is possible to obtain the number of genes and small RNAs that are differently expressed between selected data sets and their corresponding reference biological conditions. Users can then directly visualize the relative gene expression values, shown as log fold changes, in a heat map representation. Also, specific lists of elements extracted from key publications can be chosen, such as a list of internalin genes (17) or surface proteins with the LPXTG motif (59), to display their expressions under specific biological conditions.

The coexpression network interface is designed to display possible correlations between genes and small RNAs in accordance with the "guilty by association" paradigm (60). This paradigm states that two genomic features that share the same expression profile might be involved in the same functional process. Pearson correlation coeffi-

Downloaded from http://msystems.asm.org/ on April 13, 2017 by guest

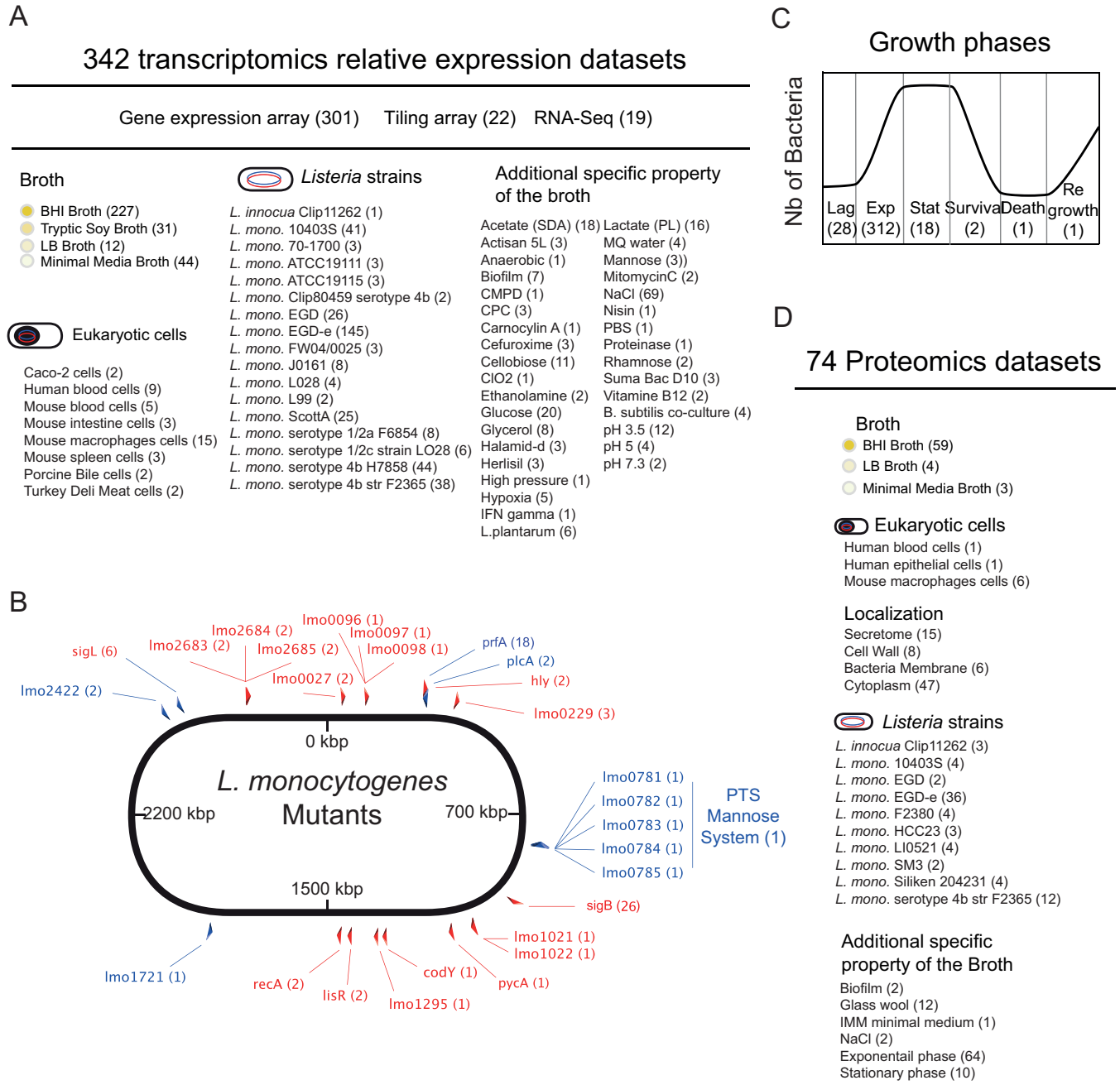


FIG 2 Transcriptomic and proteomic data sets available in the Listeriomics database. (A) Summary of all the transcriptomic data sets available at the Listeriomics website. In parentheses is the number of transcriptomics data sets available in the Listeriomics database for a specific biological condition. (B) Schematic representation of all the *L. monocytogenes* mutants for which transcriptomic data sets are available in the Listeriomics database. (C) Schematic representation of the number of transcriptomics data sets available for each *L. monocytogenes* growth phase. (D) Summary of all the proteomics data sets available at the Listeriomics website. In parentheses is the number of proteomics data sets available in the Listeriomics database for a specific biological condition.

cients have been calculated for 42 tiling array and RNA-Seq data sets. Once a specific Pearson correlation coefficient cutoff has been selected, this interface filters the coexpression network to show only the selected genome elements and the genome elements having a correlation coefficient above the selected cutoff. Two types of displays are available for the coexpression network, including a standard force-directed graph visualization (Fig. 3B) and a circular graph visualization (Fig. 3C). The latter viewer shows the Pearson correlation coefficient between genome elements by using a representation of the circular bacterial genome. This visualization highlights possible coexpression between distant genomic loci.

Downloaded from http://msystems.asm.org/ on April 13, 2017 by guest

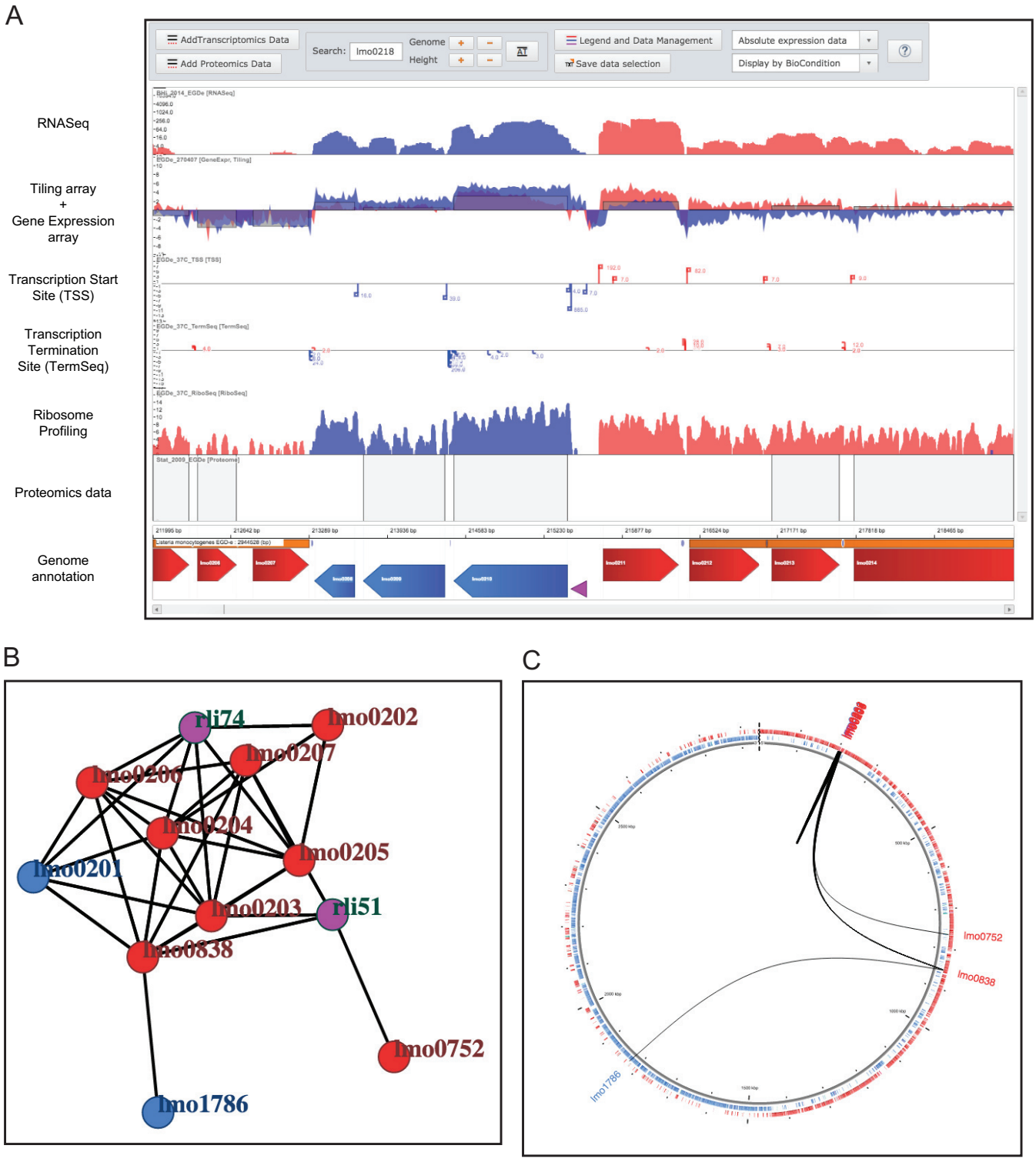


FIG 3 Multi-omics genome viewer and coexpression network tool. (A) Genome viewer of representative omics data sets for *L. monocytogenes* EGD-e grown in BHI at 37°C to the exponential and stationary growth phases as indicated in the text. The genome viewer shows positive genome strand genes (in red), negative genome strand genes (blue), tRNAs and rRNAs (in yellow), small RNAs (in purple), riboswitches (in green), asRNAs (in light green), predicted operons (in orange) from reference 56, and predicted transcription terminators (22) (in blue circles). Exp, exponential. (B) Coexpression network of the virulence locus genes (*lmo0200* to *lmo0207*) of *L. monocytogenes* EGD-e. Network nodes are genome elements (genes and noncoding RNAs) with the same color code as in the genome viewer tool. (C) Circular graph visualization of the coexpression network of the virulence locus genes (*lmo0200* to *lmo0207*). Coexpression edges are displayed overlaid on a circular representation of the EGD-e genome.

Downloaded from http://msystems.asm.org/ on April 13, 2017 by guest

The Listeriomics database contains 74 proteomics data sets (Fig. 2D). All these data can be accessed through a summary panel with a search interface. Users can visualize the selected data sets on a heat map showing the presence or absence of each *Listeria* protein under a specific biological condition. Contrary to transcriptomics data sets, relative protein expression values are not available.

A webpage form is available to inform of the need to integrate a specific data set in the Listeriomics database. Before the publication of new genomes, transcriptomes, or proteomes to the Listeriomics website, the data sets must first be uploaded on referent repositories such as the Sequence Read Archive (61) or ArrayExpress (21). This process ensures that all the data sets are formatted in accordance with international standards and that minimal information on the experimental design of each study has been provided.

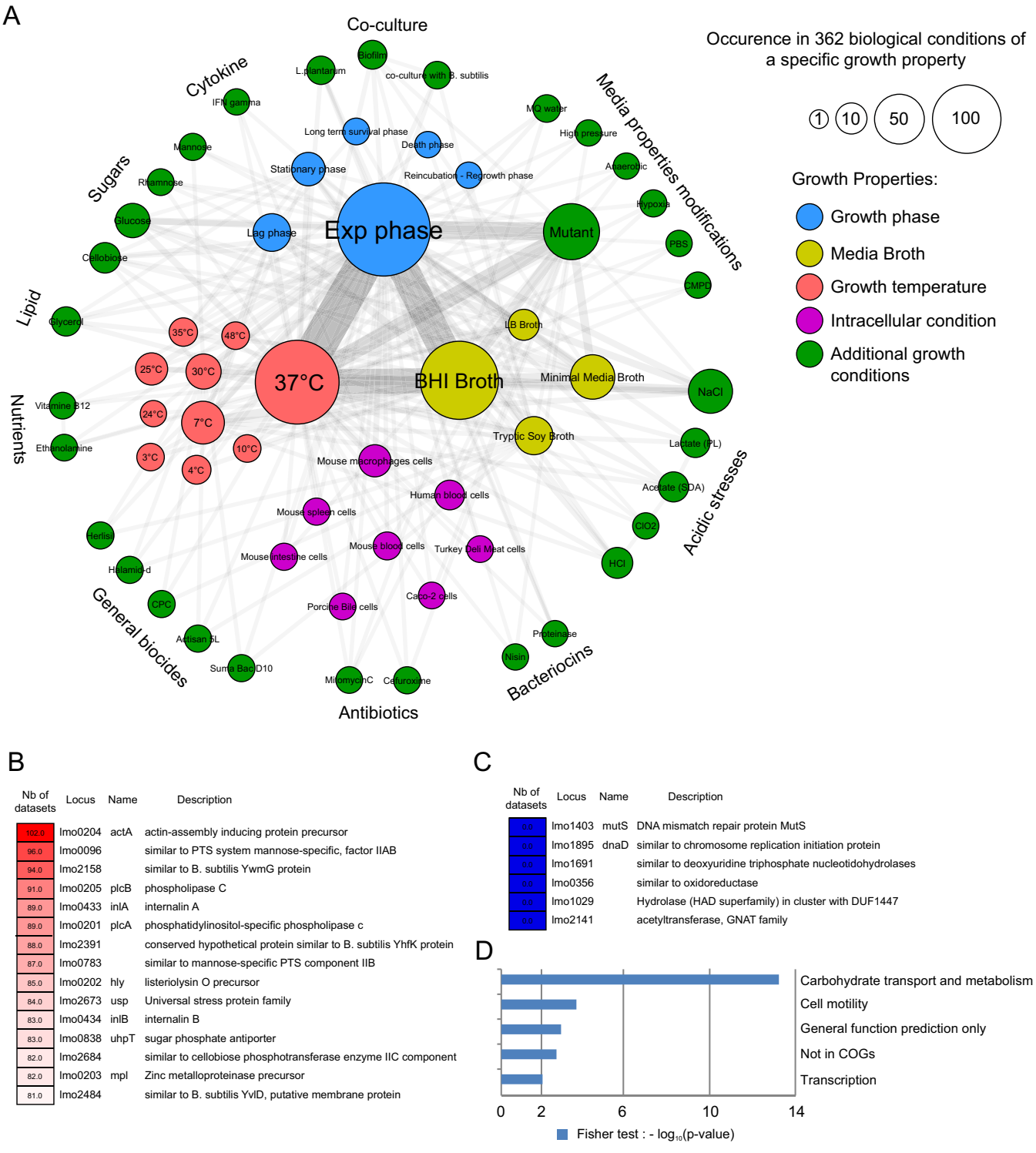
The multi-omics genome viewer. One of the key features of the Listeriomics interface is the multi-omics genome viewer. Figure 3A shows a variety of omics data sets produced for *L. monocytogenes* EGD-e grown at 37°C to the exponential phase in brain heart infusion (BHI) medium. RNA-Seq (62), tiling array (56), gene expression array (56), TSS (22), transcription termination site and ribosome profiling (63), and proteomics (36) data sets are displayed. To our knowledge, this is the first time that such a variety of omics scales can be browsed together through a genome viewer for a prokaryotic organism. Thus, the user can visualize the correlation of the genome annotation with transcription and translation for a specific coding RNA or sRNA. The genome viewer is dynamic, with zoom-in and zoom-out capabilities. The viewer also has search capability to access a specific position in the selected *Listeria* strain genome. Every omics data set present in the Listeriomics database can be added to this genome viewer.

Easy-access buttons on the home page of the Listeriomics interface allow quick access to three genome viewers with preloaded reference omics data sets. The first genome viewer is for *L. monocytogenes* EGD-e grown to the exponential phase at 37°C in BHI (Fig. 3A), the second is for the same condition but in stationary phase, and the last is for *L. monocytogenes* EGD-e grown in mouse macrophage cells.

Meta-analysis of *Listeria* transcriptomic data sets. A meta-analysis of the diversity of transcriptomic data sets available in the Listeriomics database offers a striking overview of the variety of studies performed by the *Listeria* research community. This wide range of studies covers a majority of the different living environments in which *Listeria* species have been observed to grow. As shown in Fig. 4A, several studies on the effects of different growth environments on *Listeria* have been performed, i.e., cold environments, acidic environments, specific gene deletions, common biocides, sugar availability, and intracellular growth. This variety of biological conditions can be browsed easily on the Listeriomics website. Notably, the most frequently used growth condition is bacterial growth to the exponential phase at 37°C in BHI. This biological condition is used as a reference condition in most of the studies.

We extracted the list of genes of *L. monocytogenes* EGD-e found to be differently expressed in the highest number of data sets (Fig. 4B; see Table S6). Of the top 15 genes on this list, 7 are well-studied virulence genes (*actA*, *hly*, *plcA*, *plcB*, *mpl*, *inlA*, *inlB*, *uhpT*) all regulated by the PrfA protein. We identified *actA* as the most variable gene. This gene is differently expressed in 102 of the 279 available data sets for this *Listeria* strain. Among several other functions (64–66), the ActA protein is responsible for actin nucleation and intracellular propulsion. The other genes on the list are involved in either sugar metabolism (*lmo0096*, *lmo2391*, *lmo0783*, *lmo2684*) or stress response (*lmo2158*, *lmo2673*). Finally, a not-yet-described membrane protein, Lmo2484, is differently expressed in 81 of the 279 available data sets. This protein is conserved with >90% similarity in all 83 *Listeria* strains present in the Listeriomics database. We also found six genes that have never been demonstrated to be differently expressed in a data set (Fig. 4C; see Table S6). As expected, these genes are involved in general bacterial physiology factors like DNA mismatch repair or reductase and transferase enzymes.

Downloaded from http://msystems.asm.org/ on April 13, 2017 by guest



B

Nb of datasets	Locus	Name	Description
102.0	lmo0204	actA	actin-assembling inducing protein precursor
96.0	lmo0096		similar to PTS system mannose-specific, factor IIAB
94.0	lmo2158		similar to B. subtilis YwmG protein
91.0	lmo0205	plcB	phospholipase C
89.0	lmo0433	inlA	internalin A
89.0	lmo0201	plcA	phosphatidylinositol-specific phospholipase c
88.0	lmo2391		conserved hypothetical protein similar to B. subtilis YhfK protein
87.0	lmo0783		similar to mannose-specific PTS component IIB
85.0	lmo0202	hly	listeriolysin O precursor
84.0	lmo2673	usp	Universal stress protein family
83.0	lmo0434	inlB	internalin B
83.0	lmo0838	uhpT	sugar phosphate antiporter
82.0	lmo2684		similar to cellobiose phosphotransferase enzyme IIC component
82.0	lmo0203	mpl	Zinc metalloproteinase precursor
81.0	lmo2484		similar to B. subtilis YvlD, putative membrane protein

C

Nb of datasets	Locus	Name	Description
0.0	lmo1403	mutS	DNA mismatch repair protein MutS
0.0	lmo1895	dnaD	similar to chromosome replication initiation protein
0.0	lmo1691		similar to deoxyuridine triphosphate nucleotidohydrolases
0.0	lmo0356		similar to oxidoreductase
0.0	lmo1029		Hydrolase (HAD superfamily) in cluster with DUF1447
0.0	lmo2141		acetyltransferase, GNAT family

D

FIG 4 Meta-analysis of the Listeriomics transcriptomic data sets. (A) Relational network built on the 362 transcriptomic biological conditions found in the Listeriomics database. Each node corresponds to a growth condition. The size of each node is proportional to the occurrence of each condition in the whole database. A link is drawn between two growth conditions if they are present in the same transcriptomic data set. (B) Heat map of the 15 genes with the highest ratio of differential expression. The value used for colorization is the number of data sets in which each gene has been found to be differently expressed. (C) Heat map of the six genes with no variability. (D) Pathway enrichment analysis of the 651 genes of *L. monocytogenes* EGD-e that are found differently expressed in >10% of the 279 data sets. We performed a pathway enrichment analysis by using COG information and the Fisher exact test *P* value.

Downloaded from http://msystems.asm.org/ on April 13, 2017 by guest

Finally, we extracted the 651 genes of *L. monocytogenes* EGD-e that were found to be differently expressed in >10% of the 279 data sets (see Table S6). We performed a pathway enrichment analysis by using COG (Clusters of Orthologous Groups) information and the Fisher exact test (67) (Fig. 4D). We found that the cluster of genes differently expressed in the highest number of data sets was associated with the carbohydrate transport and metabolism functions ($P = 4.15e-14$). This may be linked to the fact that, to survive in a wide variety of environments, *Listeria* bacteria have to be able to switch from one carbon source to another. The second most represented cluster is that of cell motility genes ($P = 2.18e-4$). The third ($P = 1.3e-3$) and fourth ($P = 2.1e-3$) most represented clusters include genes without any specific function, most of them being annotated as encoding hypothetical proteins. The latter result highlights the fact that many of the important genes of *Listeria* species gene regulatory networks remain to be described. The Listeriomics resource is an essential tool for their investigation.

Meta-analysis of *Listeria* proteomic data sets. An analysis of the variety of proteomic data sets available in the Listeriomics database shows that many reference biological conditions have been studied. As for transcriptomic data, there is a great number of data sets produced by bacteria grown to the exponential phase at 37°C in BHI. Users can also access data sets from intracellular growth, cold environments, and the stationary growth phase. The key information in these proteomic data sets is the protein extraction protocol used that indicates which compartment of the bacterial cells (cytoplasm, membrane, cell wall, and secretome) has been analyzed. The cytoplasmic compartment is the most studied. The second most studied is the secretome compartment, the focus of many publications, since secreted proteins are key components of the *Listeria* intracellular life cycle. The ability to visualize these proteomic data sets together at the Listeriomics website will help the *Listeria* community to investigate further these groups of proteins and their roles in the different compartments of bacterial cells.

We investigated the 58 proteomic data sets available for *L. monocytogenes* EGD-e. We counted the data sets in which each protein was analyzed (see Table S6). As expected, we found that most of the proteins analyzed in half of the data sets come from the translation machinery or the carbohydrate transport and metabolism pathway. Remarkably, the first known virulence factor on the list is the pore-forming toxin LLO (Lmo0202), which is detected in 22 of the 58 proteomic data sets present in the Listeriomics database. We also found that >80% (2,344/2,859) of the *L. monocytogenes* EGD-e proteins have been observed in one or more proteomic data sets (see Table S6). Remarkably, 512 proteins were never shown to be produced.

Systems-level analysis of the *L. monocytogenes* EGD-e virulence locus. We investigated the coexpression network of the *L. monocytogenes* EGD-e virulence locus (*lmo0200* to *lmo0206*), linking every gene with a Pearson correlation coefficient of >0.85 (Fig. 3B, and C). As part of the PrfA core regulon (68), we found *uhpT* (*lmo0838*) and *inlC* (*lmo1786*) coexpressed with the virulence locus. Strikingly, the *prfA* gene (*lmo0200*) did not appear to be strictly coexpressed with the other genes of the virulence locus. Neither the *inlA* (*lmo0433*) nor the *inlB* (*lmo0444*) gene, both of which are extensively studied virulence factors regulated by the PrfA protein, was found in the coexpression network. The two noncoding RNA elements rli51 and rli74 in the virulence locus were also found to be coexpressed. In addition, we found another gene, *lmo0752*, coexpressed with the virulence locus that has not been described in the PrfA core regulon (68). Nevertheless, this gene was previously described (69) as part of the bile tolerance locus (*lmo0745* to *lmo0755*).

DISCUSSION

The availability of curated information on genes, proteins, and cellular processes is essential not only for providing a better understanding of the *Listeria* genus but also as a key element in the development of our understanding of biological processes with respect to food industry and clinical applications. A systems biology approach requires the integration of a diversity of data collections, including, among others, genes, sRNAs,

Downloaded from http://msystems.asm.org/ on April 13, 2017 by guest

mRNAs, proteins, metabolites, protein-protein interactions, and protein-RNA interactions.

In the case of the *Listeria* genus, no such resource is currently available. At the species and strain levels, a first difficulty is indeed the recovery of these data sets in existing repositories, as is the case for *Listeria* proteomic data sets. In this way, the Listeriomics resource (http://listeriomics.pasteur.fr/) provides a unique, comprehensive, and up-to-date source of information on *Listeria*, including all the available complete genomes to date; a unified annotation of genes, proteins, and noncoding RNAs; and associated omics data and metadata relevant to genomic comparative and systems biology analyses. The curation process focused on the quality of the metadata provided. We completed the process by searching information in the corresponding publications. To our knowledge, the Listeriomics interface is the first bioinformatics resource designed to help scientists in *Listeria* pathogenesis research using a systems biology approach.

The information integrated into the Listeriomics database is provided through a useable and friendly interface that allows querying and visualization of *Listeria* genomic, phylogenetic, transcriptomic, and proteomic data sets, along with information on small RNAs and coexpression network of genes and small RNAs. The user experience and feedback from our collaborators using the Listeriomics interface for the last past 5 years were driving forces in organizing and improving the way to access data and tools (17, 22, 71). Indeed, through the Listeriomics resource, it is possible to summarize what has been discovered about a specific gene, including information on its distribution in the different *Listeria* genomes. Users can also identify the different transcriptomics data sets in which a specific gene or small RNA is differently expressed. Similarly, scientists can browse the different proteomics data sets to identify biological conditions in which a specific protein was detected by mass spectrometry. Finally, regulatory networks between genes and small RNAs can be inferred and explored.

The field of microbiology is considerably impacted by new technologies, and the role of databases in microbiology research will become even more important shortly (72). Until now, most of the model organism resources have focused on the widely studied bacterium *E. coli*, often with an emphasis on the manual curation of some particular aspect of omics information. Only a few of them, like the RegulonDB resource dedicated to the *E. coli* transcriptional regulatory network, really address the challenge of integrating knowledge based on experimental high-throughput omics data sets (50). For *B. subtilis*, SubtiWiki originally focused on manually curated gene annotation with a collection of pages providing interlinked pages on *B. subtilis* gene properties, including valuable information such as essentiality or sporulation, obtained from qualified members of the *Bacillus* community (73). Interestingly, the latest release of Subtiwiki integrates some omics data, with new modules allowing the linkage of pathway, interaction, and expression information (52). To our knowledge, none of these databases dedicated to model organisms such as *E. coli* or *B. subtilis* integrates as many data sets as the Listeriomics resource does. Moreover, Listeriomics is the only resource including such a variety of comparative and evolutionary analyses at the genomic level.

MATERIALS AND METHODS

Integration of *Listeria* species genomes and sRNAs. The complete *Listeria* genomes available in the NCBI RefSeq and GenBank databases (see Table S1) were downloaded and integrated into our the Listeriomics database. Information about serotype, lineage, and the origin of strain isolation was included when possible. We used the PasteurMLST tool (74) to search for the sequence type and clonal complex of each strain. For *L. monocytogenes* EGD-e genes, we integrated their predicted functional group (11), COG, operon prediction (56), and subcellular protein localization prediction (27) (see Fig. S1).

Many studies have been performed to identify sRNAs in *L. monocytogenes* (22, 55–58). Among these studies, only one concerned strain 10403S (57); the others focused on EGD-e. Altogether, 304 noncoding RNA elements have now been reported in *L. monocytogenes*, of which 154 are sRNAs, 104 are asRNAs, and 46 are *cis*-regulatory elements (cisRegs) including riboswitches (see Table S2 and Fig. S1). We included these elements in our database, along with supplementary information gathered from all related publications. Prediction of secondary structures at 37°C for each noncoding RNA were calculated by using UNAFold software (hybrid-ss-min default parameters on the whole RNA [76]).

Downloaded from http://msystems.asm.org/ on April 13, 2017 by guest

***Listeria* ortholog gene families.** Nucleic and amino acid sequences of all the annotated coding genes were produced for each complete *Listeria* genome by using GenBank files and a custom-made Python script. Amino acid sequences for each *Listeria* genome were aligned against those of other proteins of all the *Listeria* genomes by using BLASTP+ (53) with an E-value threshold of 0.01. We used PanOCT (78) to build families of *Listeria* orthologs. PanOCT is able to deal with recently diverged paralogs by using neighborhood gene information. The percentage of genomes needed for a cluster to be considered an orthologous gene family was set to 100%. The length ratio to eliminate shorter protein fragments when a protein is split because of a frameshift event was set to 1.33 as previously recommended (78). Ortholog gene families were finally extracted from panoct.pl output files by using the gene_order.pl script, which is included in the PanOCT archive.

Amino acid sequences of each cluster were aligned by using ProbCons version 1.12 (79) with default parameters. Resulting alignments were postprocessed to filter unreliable positions by using Gblocks version 0.91b (80) with the parameter settings as follows. The minimum number of sequences for a conserved position was set to (n/2) + 1 (where n is the total number of sequences in the aligned data set), the maximum number of contiguous nonconserved positions was set to 50, the minimum length required for a block was set to 5, and gap positions were not allowed. The corresponding nucleic acid sequence alignments were obtained from each cluster amino acid sequence alignment with a custom-made Perl script.

***Listeria* species reference phylogenomic tree.** Nucleic acid sequence alignments of ortholog gene families were concatenated into a single superalignment. This superalignment was used to compute a maximum-likelihood tree by using FastTree 2 (81), a parallelized and optimized software, to build maximum-likelihood trees. The following parameters were used. The generalized time-reversible model was chosen, the exhaustive search mode was selected to obtain a more accurate reconstruction, NNI and SPR heuristics were used to browse the tree space, and BIONJ weighting was chosen to join events during tree space browsing. Support analyses were performed with the Shimodaira-Hasegawa test associated with 1,000 resampling steps of site likelihood.

Integration of transcriptomic data sets. We downloaded the 64 published *Listeria*-related ArrayExpress experiments (21) in MAGE-TAB standard (82) (Fig. 1; see Table S3). Every MAGE-TAB file included an IDF (investigation description format) file showing general information about the experiment and an SDRF (sample and data relationship format) file showing data relationship. We manually curated every SDRF file and integrated the different metadata of the data set by using the publication linked to each study when available. We grouped the metadata information into key biological parameters for each data set: growth, time point, temperature, mutant, media, strain used, and strain array (Fig. 2A to C).

We then added matrices of expression or comparison for each biological condition. In total, 32 different gene expression arrays and six RNA-Seq technologies were combined. For gene expression arrays, we downloaded the 32 ADF (array design format) files that describe the relationship between array probes and genes. Only processed data provided by ArrayExpress were used, and no new normalization was applied. For each gene, we calculated the median differences of log values for relative expression tables. In the case of RNA-Seq, alignments of raw reads were performed with the Bowtie (83), segemehl (84), or novoalignCS tool, depending on the sequencing technology. For each alignment file, we computed the number of reads per kilobase per million mapped reads for each genomic feature and the number of reads per million for genome-wide coverage. Differential expression analysis was performed with DESeq version 1.14.0 (85) on per-feature raw counts.

Altogether 492 files, 150 absolute value data, and 342 relative expression data were created and integrated into the Listeriomics website (Fig. 1 and 2; see Fig. S1). For every experiment, relative expression data were always available, whereas absolute expression data were found for only a quarter of the data.

Construction of an expression atlas. We designed an expression atlas to provide for each transcriptomic data set a list of the differently expressed genome elements based on log fold change values. A statistical analysis (Shapiro-Wilk and Lilliefors tests) showed that all the relative expression data sets have a Gaussian distribution with a mean equal to zero. We applied normalization directly to log fold change values by multiplying every value by 1 divided by the square root of sigma, where sigma is the estimated standard deviation of each data set. This normalization ensured that every log fold change data set will have a standard deviation equal to 1 (Fig. S2). Consequently, the user can apply a standard cutoff, like log fold change < 1.5, and extract the differently expressed genome elements for every data set. A standard deviation equal to 1 leads, on average, to a selection of 7.1% of the most differently expressed genome elements, without removing the existing differences in the number of elements differently expressed under different biological conditions.

Construction of coexpression networks. We selected 42 transcriptomic data sets for which the absolute expression of genes and small RNAs was available (21 tiling array and 21 RNA-Seq [see Table S4 and Fig. S1]). The Pearson correlation coefficient for each genome element across all biological conditions was calculated. A link in the coexpression networks was created only when the absolute Pearson correlation coefficient was >0.80.

Integration of proteomic data sets. We selected 23 publications by using the PubMed database, in which a mass spectrometry experiment with a *Listeria* strain was performed (see Table S5). We extracted all the available information on the biological conditions screened in each experiment from associated articles and supplementary files (Fig. 1). We also extracted the metadata information in the same format for the transcriptomic data sets. In all the experiments, a list of the proteins detected was available. In total, we extracted 102 proteomics files (74 absolute expression data sets and 28 relative-expression data sets [Fig. 2D; see Fig. S1]).

Downloaded from http://msystems.asm.org/ on April 13, 2017 by guest

Construction of the Listeriomics website and desktop versions. The Listeriomics website was built by using the BACNET development platform (unpublished data). This platform is based on Java and Eclipse e4 RCP/RAP API for building both web and desktop software versions and provides a rich user interface. This open-source platform is generic and can be used to set up a similar website for any organism.

Tutorials for the Listeriomics website can be found on the Listeriomics mediaWiki webpage. These tutorials can be accessed either from the home webpage or directly at http://wiki.listeriomics.com/.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/mSystems.00186-16.

FIG S1, EPS file, 1.1 MB.
FIG S2, EPS file, 1.5 MB.
TABLE S1, XLS file, 0.05 MB.
TABLE S2, XLS file, 0.1 MB.
TABLE S3, XLS file, 0.1 MB.
TABLE S4, XLS file, 0.03 MB.
TABLE S5, XLS file, 0.04 MB.
TABLE S6, XLS file, 1.5 MB.
TEXT S1, DOCX file, 0.1 MB.

ACKNOWLEDGMENTS

We thank Francis Impens, Jeffrey Mellin, and Nathalie Rolhion for their important feedback on the Listeriomics website, which helped us improve the user interface. We are grateful to Eric Westhof and Lilliana Radoshevich for their insights into the writing and structure of the manuscript. We also thank Daniel Dar, Omri Wurtzel, and Rotem Sorek of the Weizmann Institute for providing the TSS, RiboSeq, and TermSeq data sets.

This work received financial support from the European Research Council (BacCellEpi no. 670823), the French Agence Nationale de la Recherche (grant BACNET 10-BINF-02-01), the Human Frontier Science Program (grant RGP0011/2013), Laboratoire d'Excellence "Integrative Biology of Emerging Infectious Diseases" (ANR-10-LABX-62-IBEID), the Institut Pasteur, the Institut National de la Santé et de la Recherche Médicale, the Institut National de la Recherche Agronomique, the Pasteur-Weizmann Council grant, and the Fondation le Roch Les Mousquetaires. P.C. is a senior international research scholar of the Howard Hughes Medical Institute.

REFERENCES

1. Cossart P. 2011. Illuminating the landscape of host-pathogen interactions with the bacterium Listeria monocytogenes. Proc Natl Acad Sci U S A 108:19484–19491. https://doi.org/10.1073/pnas.1112371108.
2. Hamon M, Bierne H, Cossart P. 2006. Listeria monocytogenes: a multifaceted model. Nat Rev Microbiol 4:423–434. https://doi.org/10.1038/nrmicro1413.
3. Lecuit M. 2007. Human listeriosis and animal models. Microbes Infect 9:1216–1225. https://doi.org/10.1016/j.micinf.2007.05.009.
4. Chiara M, Caruso M, D'Erchia AM, Manzari C, Fraccalvieri R, Goffredo E, Latorre L, Miccolupo A, Padalino I, Santagada G, Chiocco D, Pesole G, Horner DS, Parisi A. 2015. Comparative genomics of Listeria sensu lato: genus-wide differences in evolutionary dynamics and the progressive gain of complex, potentially pathogenicity-related traits through lateral gene transfer. Genome Biol Evol 7:2154–2172. https://doi.org/10.1093/gbe/evv131.
5. Graves LM, Helsel LO, Steigerwalt AG, Morey RE, Daneshvar MI, Roof SE, Orsi RH, Fortes ED, Milillo SR, den Bakker HC, Wiedmann M, Swaminathan B, Sauders BD. 2010. Listeria marthii sp. nov., isolated from the natural environment, Finger Lakes National Forest. Int J Syst Evol Microbiol 60:1280–1288. https://doi.org/10.1099/ijs.0.014118-0.
6. Leclercq A, Clermont D, Bizet C, Grimont PAD, Le Flèche-Matéos A, Roche SM, Buchrieser C, Cadet-Daniel V, Le Monnier A, Lecuit M, Allerberger F. 2010. Listeria rocourtiae sp. nov. Int J Syst Evol Microbiol 60:2210–2214. https://doi.org/10.1099/ijs.0.017376-0.
7. Lang Halter E, Neuhaus K, Scherer S. 2013. Listeria weihenstephanensis sp. nov., isolated from the water plant Lemna trisulca taken from a freshwater pond. Int J Syst Evol Microbiol 63:641–647. https://doi.org/10.1099/ijs.0.036830-0.
8. Bertsch D, Rau J, Eugster MR, Haug MC, Lawson PA, Lacroix C, Meile L. 2013. Listeria fleischmannii sp. nov., isolated from cheese. Int J Syst Evol Microbiol 63:526–532. https://doi.org/10.1099/ijs.0.036947-0.
9. den Bakker HC, Warchocki S, Wright EM, Allred AF, Ahlstrom C, Manuel CS, Stasiewicz MJ, Burrell A, Roof S, Strawn LK, Fortes E, Nightingale KK, Kephart D, Wiedmann M. 2014. Listeria floridensis sp. nov., Listeria aquatica sp. nov., Listeria cornellensis sp. nov., Listeria riparia sp. nov. and Listeria grandensis sp. nov., from agricultural and natural environments. Int J Syst Evol Microbiol 64:1882–1889. https://doi.org/10.1099/ijs.0.052720-0.
10. Weller D, Andrus A, Wiedmann M, den Bakker HC. 2015. Listeria booriae sp. nov. and Listeria newyorkensis sp. nov., from food processing environments in the USA. Int J Syst Evol Microbiol 65:286–292. https://doi.org/10.1099/ijs.0.070839-0.
11. Glaser P, Frangeul L, Buchrieser C, Rusniok C, Amend A, Baquero F, Berche P, Bloecker H, Brandt P, Chakraborty T, Charbit A, Chetouani F, Couvé E, de Daruvar A, Dehoux P, Domann E, Domínguez-Bernal G, Duchaud E, Durant L, Dussurget O, Entian KD, Fsihi H, García-del Portillo F, Garrido P, Gautier L, Goebel W, Gómez-López N, Hain T, Hauf J, Jackson D, Jones LM, Kaerst U, Kreft J, Kuhn M, Kunst F, Kurapkat G, Madueno E, Maitournam A, Vicente JM, Ng E, Nedjari H, Nordsiek G, Novella S, de Pablos B, Pérez-Diaz JC, Purcell R, Remmel B, Rose M, Schlueter T, Simoes N, et al. 2001. Comparative genomics of Listeria species. Science 294:849–852.

Downloaded from http://msystems.asm.org/ on April 13, 2017 by guest

12. Nelson KE, Fouts DE, Mongodin EF, Ravel J, DeBoy RT, Kolonay JF, Rasko DA, Angiuoli SV, Gill SR, Paulsen IT, Peterson J, White O, Nelson WC, Nierman W, Beanan MJ, Brinkac LM, Daugherty SC, Dodson RJ, Durkin AS, Madupu R, Haft DH, Selengut J, Van Aken S, Khouri H, Fedorova N, Forberger H, Tran B, Kathariou S, Wonderling LD, Uhlich GA, Bayles DO, Luchansky JB, Fraser CM. 2004. Whole genome comparisons of serotype 4b and 1/2a strains of the food-borne pathogen Listeria monocytogenes reveal new insights into the core genome components of this species. Nucleic Acids Res 32:2386–2395. https://doi.org/10.1093/nar/gkh562.
13. Doumith M, Cazalet C, Simoes N, Frangeul L, Jacquet C, Kunst F, Martin P, Cossart P, Glaser P, Buchrieser C. 2004. New aspects regarding evolution and virulence of Listeria monocytogenes revealed by comparative genomics and DNA arrays. Infect Immun 72:1072–1083. https://doi.org/10.1128/IAI.72.2.1072-1083.2004.
14. Gilmour MW, Graham M, Van Domselaar G, Tyler S, Kent H, Trout-Yakel KM, Larios O, Allen V, Lee B, Nadon C. 2010. High-throughput genome sequencing of two Listeria monocytogenes clinical isolates during a large foodborne outbreak. BMC Genomics 11:120. https://doi.org/10.1186/1471-2164-11-120.
15. Orsi RH, den Bakker HC, Wiedmann M. 2011. Listeria monocytogenes lineages: genomics, evolution, ecology, and phenotypic characteristics. Int J Med Microbiol 301:79–96. https://doi.org/10.1016/j.ijmm.2010.05.002.
16. Kuenne C, Billion A, Mraheil MA, Strittmatter A, Daniel R, Goesmann A, Barbuddhe S, Hain T, Chakraborty T. 2013. Reassessment of the Listeria monocytogenes pan-genome reveals dynamic integration hotspots and mobile genetic elements as major components of the accessory genome. BMC Genomics 14:47. https://doi.org/10.1186/1471-2164-14-47.
17. Bécavin C, Bouchier C, Lechat P, Archambaud C, Creno S, Gouin E, Wu Z, Kühbacher A, Brisse S, Pucciarelli MG, García-del Portillo F, Hain T, Portnoy DA, Chakraborty T, Lecuit M, Pizarro-Cerdá J, Moszer I, Bierne H, Cossart P. 2014. Comparison of widely used Listeria monocytogenes strains EGD, 10403S, and EGD-e highlights genomic variations underlying differences in pathogenicity. mBio 5:e00969-14. https://doi.org/10.1128/mBio.00969-14.
18. Lechat P, Hummel L, Rousseau S, Moszer I. 2008. GenoList: an integrated environment for comparative analysis of microbial genomes. Nucleic Acids Res 36:D469–D474. https://doi.org/10.1093/nar/gkm1042.
19. Tan MF, Siow CC, Dutta A, Mutha NV, Wee WY, Heydari H, Tan SY, Ang MY, Wong GJ, Choo SW. 2015. Development of ListeriaBase and comparative analysis of Listeria monocytogenes. BMC Genomics 16:755. https://doi.org/10.1186/s12864-015-1959-5.
20. Bennett HJ, Pearce DM, Glenn S, Taylor CM, Kuhn M, Sonenshein AL, Andrew PW, Roberts IS. 2007. Characterization of relA and codY mutants of Listeria monocytogenes: identification of the CodY regulon and its role in virulence. Mol Microbiol 63:1453–1467. https://doi.org/10.1111/j.1365-2958.2007.05597.x.
21. Rustici G, Kolesnikov N, Brandizi M, Burdett T, Dylag M, Emam I, Farne A, Hastings E, Ison J, Keays M, Kurbatova N, Malone J, Mani R, Mupo A, Pedro Pereira R, Pilicheva E, Rung J, Sharma A, Tang YA, Ternent T, Tikhonov A, Welter D, Williams E, Brazma A, Parkinson H, Sarkans U. 2013. ArrayExpress update—trends in database growth and links to data analysis tools. Nucleic Acids Res 41:D987–D990. https://doi.org/10.1093/nar/gks1174.
22. Wurtzel O, Sesto N, Mellin JR, Karunker I, Edelheit S, Bécavin C, Archambaud C, Cossart P, Sorek R. 2012. Comparative transcriptomics of pathogenic and non-pathogenic Listeria species. Mol Syst Biol 8:583. https://doi.org/10.1038/msb.2012.11.
23. Pischimarov J, Kuenne C, Billion A, Hemberger J, Cemič F, Chakraborty T, Hain T. 2012. sRNAdb: a small non-coding RNA database for gram-positive bacteria. BMC Genomics 13:384. https://doi.org/10.1186/1471-2164-13-384.
24. Halbedel S, Reiss S, Hahn B, Albrecht D, Mannala GK, Chakraborty T, Hain T, Engelmann S, Flieger A. 2014. A systematic proteomic analysis of Listeria monocytogenes house-keeping protein secretion systems. Mol Cell Proteomics 13:3063–3081. https://doi.org/10.1074/mcp.M114.041327.
25. Desvaux M, Dumas E, Chafsey I, Chambon C, Hébraud M. 2010. Comprehensive appraisal of the extracellular proteins from a monoderm bacterium: theoretical and empirical exoproteomes of Listeria monocytogenes EGD-e by secretomics. J Proteome Res 9:5076–5092. https://doi.org/10.1021/pr1003642.
26. Renier S, Chambon C, Viala D, Chagnot C, Hébraud M, Desvaux M. 2013. Exoproteomic analysis of the SecA2-dependent secretion in Listeria monocytogenes EGD-e. J Proteomics 80:183–195. https://doi.org/10.1016/j.jprot.2012.11.027.
27. Renier S, Micheau P, Talon R, Hébraud M, Desvaux M. 2012. Subcellular localization of extracytoplasmic proteins in monoderm bacteria: rational secretomics-based strategy for genomic and proteomic analyses. PLoS One 7:e42982. https://doi.org/10.1371/journal.pone.0042982.
28. Trost M, Wehmhöner D, Kärst U, Dieterich G, Wehland J, Jänsch L. 2005. Comparative proteome analysis of secretory proteins from pathogenic and nonpathogenic Listeria species. Proteomics 5:1544–1557. https://doi.org/10.1002/pmic.200401024.
29. Calvo E, Pucciarelli MG, Bierne H, Cossart P, Albar JP, García-Del Portillo F. 2005. Analysis of the Listeria cell wall proteome by two-dimensional nanoliquid chromatography coupled to mass spectrometry. Proteomics 5:433–443. https://doi.org/10.1002/pmic.200400936.
30. Quereda JJ, Pucciarelli MG, Botello-Morte L, Calvo E, Carvalho F, Bouchier C, Vieira A, Mariscotti JF, Chakraborty T, Cossart P, Hain T, Cabanes D, García-Del Portillo F. 2013. Occurrence of mutations impairing sigma factor B (SigB) function upon inactivation of Listeria monocytogenes genes encoding surface proteins. Microbiology 159:1328–1339. https://doi.org/10.1099/mic.0.067744-0.
31. Wehmhöner D, Dieterich G, Fischer E, Baumgärtner M, Wehland J, Jänsch L. 2005. 'LaneSpector', a tool for membrane proteome profiling based on sodium dodecyl sulfate-polyacrylamide gel electrophoresis/liquid chromatography-tandem mass spectrometry analysis: application to *Listeria monocytogenes* membrane proteins. Electrophoresis 26:2450–2460. https://doi.org/10.1002/elps.200410348.
32. Dieterich G, Kärst U, Fischer E, Wehland J, Jänsch L. 2006. LEGER: knowledge database and visualization tool for comparative genomics of pathogenic and non-pathogenic Listeria species. Nucleic Acids Res 34:D402–D406. https://doi.org/10.1093/nar/gkj071.
33. Zhang CXY, Creskey MC, Cyr TD, Brooks B, Huang H, Pagotto F, Lin M. 2013. Proteomic identification of Listeria monocytogenes surface-associated proteins. Proteomics 13:3040–3045. https://doi.org/10.1002/pmic.201200449.
34. Lee JH, Choi CW, Lee T, Kim SII, Lee JC, Shin JH. 2013. Transcription factor σ^B plays an important role in the production of extracellular membrane-derived vesicles in Listeria monocytogenes. PLoS One 8:e73196. https://doi.org/10.1371/journal.pone.0073196.
35. Tiong HK, Hartson S, Muriana PM. 2015. Comparison of five methods for direct extraction of surface proteins from Listeria monocytogenes for proteomic analysis by Orbitrap mass spectrometry. J Microbiol Methods 110:54–60. https://doi.org/10.1016/j.mimet.2015.01.004.
36. Donaldson JR, Nanduri B, Burgess SC, Lawrence ML. 2009. Comparative proteomic analysis of Listeria monocytogenes strains F2365 and EGD. Appl Environ Microbiol 75:366–373. https://doi.org/10.1128/AEM.01847-08.
37. Donaldson JR, Nanduri B, Pittman JR, Givaruangsawat S, Burgess SC, Lawrence ML. 2011. Proteomic expression profiles of virulent and avirulent strains of Listeria monocytogenes isolated from macrophages. J Proteomics 74:1906–1917. https://doi.org/10.1016/j.jprot.2011.05.008.
38. Agoston R, Soni K, Jesudhasan PR, Russell WK, Mohácsi-Farkas C, Pillai SD. 2009. Differential expression of proteins in Listeria monocytogenes under thermotolerance-inducing, heat shock, and prolonged heat shock conditions. Foodborne Pathog Dis 6:1133–1140. https://doi.org/10.1089/fpd.2009.0286.
39. Cacace G, Mazzeo MF, Sorrentino A, Spada V, Malorni A, Siciliano RA. 2010. Proteomics for the elucidation of cold adaptation mechanisms in Listeria monocytogenes. J Proteomics 73:2021–2030. https://doi.org/10.1016/j.jprot.2010.06.011.
40. Pittman JR, Buntyn JO, Posadas G, Nanduri B, Pendarvis K, Donaldson JR. 2014. Proteomic analysis of cross protection provided between cold and osmotic stress in Listeria monocytogenes. J Proteome Res 13:1896–1904. https://doi.org/10.1021/pr401004a.
41. Van de Velde S, Delaive E, Dieu M, Carryn S, Van Bambeke F, Devreese B, Raes M, Tulkens PM. 2009. Isolation and 2-D-DIGE proteomic analysis of intracellular and extracellular forms of Listeria monocytogenes. Proteomics 9:5484–5496. https://doi.org/10.1002/pmic.200900503.
42. Zhou Q, Feng X, Zhang Q, Feng F, Yin X, Shang J, Qu H, Luo Q. 2012. Carbon catabolite control is important for Listeria monocytogenes biofilm formation in response to nutrient availability. Curr Microbiol 65:35–43. https://doi.org/10.1007/s00284-012-0125-4.
43. Folio P, Chavant P, Chafsey I, Belkorchia A, Chambon C, Hébraud M. 2004. Two-dimensional electrophoresis database of Listeria monocytogenes EGDe proteome and proteomic analysis of mid-log and stationary

Downloaded from http://msystems.asm.org/ on April 13, 2017 by guest

growth phase cells. Proteomics 4:3187–3201. https://doi.org/10.1002/pmic.200300841.
44. Abram F, Su WL, Wiedmann M, Boor KJ, Coote P, Botting C, Karatzas KAG, O'Byrne CP. 2008. Proteomic analyses of a Listeria monocytogenes mutant lacking sigmaB identify new components of the sigmaB regulon and highlight a role for sigmaB in the utilization of glycerol. Appl Environ Microbiol 74:594–604. https://doi.org/10.1128/AEM.01921-07.
45. Rosenbloom KR, Sloan CA, Malladi VS, Dreszer TR, Learned K, Kirkup VM, Wong MC, Maddren M, Fang R, Heitner SG, Lee BT, Barber GP, Harte RA, Diekhans M, Long JC, Wilder SP, Zweig AS, Karolchik D, Kuhn RM, Haussler D, Kent WJ. 2013. ENCODE data in the UCSC genome browser: year 5 update. Nucleic Acids Res 41:D56–D63. https://doi.org/10.1093/nar/gks1172.
46. Caspi R, Altman T, Billington R, Dreher K, Foerster H, Fulcher CA, Holland TA, Keseler IM, Kothari A, Kubo A, Krummenacker M, Latendresse M, Mueller LA, Ong Q, Paley S, Subhraveti P, Weaver DS, Weerasinghe D, Zhang P, Karp PD. 2014. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res 42:D459–D471. https://doi.org/10.1093/nar/gkt1103.
47. Orsi RH, Bergholz TM, Wiedmann M, Boor KJ. 2015. The Listeria monocytogenes strain 10403S BioCyc database. Database (Oxford) 2015. https://doi.org/10.1093/database/bav027.
48. Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, Gillespie JJ, Gough R, Hix D, Kenyon R, Machi D, Mao C, Nordberg EK, Olson R, Overbeek R, Pusch GD, Shukla M, Schulman J, Stevens RL, Sullivan DE, Vonstein V, Warren A, Will R, Wilson MJC, Yoo HS, Zhang C, Zhang Y, Sobral BW. 2014. PATRIC, the bacterial bioinformatics database and analysis resource. Nucleic Acids Res 42:D581–D591. https://doi.org/10.1093/nar/gkt1099.
49. Cloots L, Marchal K. 2011. Network-based functional modeling of genomics, transcriptomics and metabolism in bacteria. Curr Opin Microbiol 14:599–607. https://doi.org/10.1016/j.mib.2011.09.003.
50. Salgado H, Peralta-Gil M, Gama-Castro S, Santos-Zavaleta A, Muñiz-Rascado L, García-Sotelo JS, Weiss V, Solano-Lira H, Martínez-Flores I, Medina-Rivera A, Salgado-Osorio G, Alquicira-Hernández S, Alquicira-Hernández K, López-Fuentes A, Porrón-Sotelo L, Huerta AM, Bonavides-Martínez C, Balderas-Martínez YI, Pannier L, Olvera M, Labastida A, Jiménez-Jacinto V, Vega-Alvarado L, Del Moral-Chávez V, Hernández-Alvarez A, Morett E, Collado-Vides J. 2013. RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. Nucleic Acids Res 41:D203–D213. https://doi.org/10.1093/nar/gks1201.
51. Hu JC, Sherlock G, Siegele DA, Aleksander SA, Ball CA, Demeter J, Gouni S, Holland TA, Karp PD, Lewis JE, Liles NM, McIntosh BK, Mi H, Muruganujan A, Wymore F, Thomas PD, Altman T. 2014. PortEco: a resource for exploring bacterial biology through high-throughput data and analysis tools. Nucleic Acids Res 42:D677–D684. https://doi.org/10.1093/nar/gkt1203.
52. Mäder U, Schmeisky AG, Flórez LA, Stülke J. 2012. SubtiWiki—a comprehensive community resource for the model organism Bacillus subtilis. Nucleic Acids Res 40:D1278–D1287. https://doi.org/10.1093/nar/gkr923.
53. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. Blast+: architecture and applications. BMC Bioinformatics 10:421. https://doi.org/10.1186/1471-2105-10-421.
54. Lechat P, Souche E, Moszer I. 2013. SynTView---an interactive multi-view genome browser for next-generation comparative microorganism genomics. BMC Bioinformatics 14:277. https://doi.org/10.1186/1471-2105-14-277.
55. Mandin P, Repoila F, Vergassola M, Geissmann T, Cossart P. 2007. Identification of new noncoding RNAs in Listeria monocytogenes and prediction of mRNA targets. Nucleic Acids Res 35:962–974. https://doi.org/10.1093/nar/gkl1096.
56. Toledo-Arana A, Dussurget O, Nikitas G, Sesto N, Guet-Revillet H, Balestrino D, Loh E, Gripenland J, Tiensuu T, Vaitkevicius K, Barthelemy M, Vergassola M, Nahori MA, Soubigou G, Régnault B, Coppée JY, Lecuit M, Johansson J, Cossart P. 2009. The Listeria transcriptional landscape from saprophytism to virulence. Nature 459:950–956. https://doi.org/10.1038/nature08080.
57. Oliver HF, Orsi RH, Ponnala L, Keich U, Wang W, Sun Q, Cartinhour SW, Filiatrault MJ, Wiedmann M, Boor KJ. 2009. Deep RNA sequencing of L. monocytogenes reveals overlapping and extensive stationary phase and sigma B-dependent transcriptomes, including multiple highly transcribed noncoding RNAs. BMC Genomics 10:641. https://doi.org/10.1186/1471-2164-10-641.
58. Mraheil MA, Billion A, Mohamed W, Mukherjee K, Kuenne C, Pischimarov J, Krawitz C, Retey J, Hartsch T, Chakraborty T, Hain T. 2011. The intracellular sRNA transcriptome of Listeria monocytogenes during growth in macrophages. Nucleic Acids Res 39:4235–4248. https://doi.org/10.1093/nar/gkr033.
59. Bierne H, Cossart P. 2007. Listeria monocytogenes surface proteins: from genome predictions to function. Microbiol Mol Biol Rev 71:377–397. https://doi.org/10.1128/MMBR.00039-06.
60. Aravind L. 2000. Guilt by association: contextual information in genome analysis. Genome Res 10:1074–1077. https://doi.org/10.1101/gr.10.8.1074.
61. Leinonen R, Sugawara H, Shumway M, International Nucleotide Sequence Database Collaboration. 2011. The sequence read archive. Nucleic Acids Res 39:D19–D21. https://doi.org/10.1093/nar/gkq1019.
62. Wehner S, Mannala GK, Qing X, Madhugiri R, Chakraborty T, Mraheil MA, Hain T, Marz M. 2014. Detection of very long antisense transcripts by whole transcriptome RNA-Seq analysis of Listeria monocytogenes by semiconductor sequencing technology. PLoS One 9:e108639. https://doi.org/10.1371/journal.pone.0108639.
63. Dar D, Shamir M, Mellin JR, Koutero M, Stern-Ginossar N, Cossart P, Sorek R. 2016. Term-seq reveals abundant ribo-regulation of antibiotics [sic] resistance in bacteria. Science 352:aad9822. https://doi.org/10.1126/science.aad9822.
64. Le Monnier A, Autret N, Join-Lambert OF, Jaubert F, Charbit A, Berche P, Kayal S. 2007. ActA is required for crossing of the fetoplacental barrier by Listeria monocytogenes. Infect Immun 75:950–957. https://doi.org/10.1128/IAI.01570-06.
65. Yoshikawa Y, Ogawa M, Hain T, Yoshida M, Fukumatsu M, Kim M, Mimuro H, Nakagawa I, Yanagawa T, Ishii T, Kakizuka A, Sztul E, Chakraborty T, Sasakawa C. 2009. Listeria monocytogenes ActA-mediated escape from autophagic recognition. Nat Cell Biol 11:1233–1240. https://doi.org/10.1038/ncb1967.
66. Travier L, Guadagnini S, Gouin E, Dufour A, Chenal-Francisque V, Cossart P, Olivo-Marin JC, Ghigo JM, Disson O, Lecuit M. 2013. ActA promotes Listeria monocytogenes aggregation, intestinal colonization and carriage. PLoS Pathog 9:e1003131. https://doi.org/10.1371/journal.ppat.1003131.
67. Gold DL, Coombes KR, Wang J, Mallick B. 2007. Enrichment analysis in high-throughput genomics—accounting for dependency in the NULL. Brief Bioinform 8:71–77. https://doi.org/10.1093/bib/bbl019.
68. Scortti M, Monzó HJ, Lacharme-Lora L, Lewis DA, Vázquez-Boland JA. 2007. The PrfA virulence regulon. Microbes Infect 9:1196–1207. https://doi.org/10.1016/j.micinf.2007.05.007.
69. Begley M, Sleator RD, Gahan CG, Hill C. 2005. Contribution of three bile-associated loci, bsh, pva, and btlB, to gastrointestinal persistence and bile tolerance of Listeria monocytogenes. Infect Immun 73:894–904. https://doi.org/10.1128/IAI.73.2.894-904.2005.
70. Reference deleted.
71. Mellin JR, Tiensuu T, Bécavin C, Gouin E, Johansson J, Cossart P. 2013. A riboswitch-regulated antisense RNA in Listeria monocytogenes. Proc Natl Acad Sci U S A 110:13132–13137. https://doi.org/10.1073/pnas.1304795110.
72. Médigue C, Moszer I. 2007. Annotation, comparison and databases for hundreds of bacterial genomes. Res Microbiol 158:724–736. https://doi.org/10.1016/j.resmic.2007.09.009.
73. Flórez LA, Roppel SF, Schmeisky AG, Lammers CR, Stülke J. 2009. A community-curated consensual annotation that is continuously updated: the Bacillus subtilis centred Wiki SubtiWiki. Database (Oxford) 2009:bap012.
74. Ragon M, Wirth T, Hollandt F, Lavenir R, Lecuit M, Le Monnier A, Brisse S. 2008. A new perspective on Listeria monocytogenes evolution. PLoS Pathog 4:e1000146. https://doi.org/10.1371/journal.ppat.1000146.
75. Reference deleted.
76. Markham NR, Zuker M. 2008. UNAFold: software for nucleic acid folding and hybridization. Methods Mol Biol 453:3–31. https://doi.org/10.1007/978-1-60327-429-6_1.
77. Reference deleted.
78. Fouts DE, Brinkac L, Beck E, Inman J, Sutton G. 2012. PanOCT: automated clustering of orthologs using conserved gene neighborhood for pan-genomic analysis of bacterial strains and closely related species. Nucleic Acids Res 40:e172. https://doi.org/10.1093/nar/gks757.
79. Do CB, Mahabhashyam MSP, Brudno M, Batzoglou S. 2005. ProbCons: probabilistic consistency-based multiple sequence alignment. Genome Res 15:330–340. https://doi.org/10.1101/gr.2821705.
80. Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein se-

Downloaded from http://msystems.asm.org/ on April 13, 2017 by guest

quence alignments. Syst Biol 56:564–577. https://doi.org/10.1080/10635150701472164.
81. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. PLoS One 5:e9490. https://doi.org/10.1371/journal.pone.0009490.
82. Rayner TF, Rocca-Serra P, Spellman PT, Causton HC, Farne A, Holloway E, Irizarry RA, Liu J, Maier DS, Miller M, Petersen K, Quackenbush J, Sherlock G, Stoeckert CJ, White J, Whetzel PL, Wymore F, Parkinson H, Sarkans U, Ball CA, Brazma A. 2006. A simple spreadsheet-based, MIAME-supportive format for microarray data: MAGE-TAB. BMC Bioinformatics 7:489. https://doi.org/10.1186/1471-2105-7-489.
83. Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25. https://doi.org/10.1186/gb-2009-10-3-r25.
84. Hoffmann S, Otto C, Kurtz S, Sharma CM, Khaitovich P, Vogel J, Stadler PF, Hackermüller J. 2009. Fast mapping of short sequences with mismatches, insertions and deletions using index structures. PLoS Comput Biol 5:e1000502. https://doi.org/10.1371/journal.pcbi.1000502.
85. Anders S, Huber W. 2010. Differential expression analysis for sequence count data. Genome Biol 11:R106. https://doi.org/10.1186/gb-2010-11-10-r106.

Downloaded from http://msystems.asm.org/ on April 13, 2017 by guest