Journal of the Iowa Academy of Science: JIAS

Volume 121 | Number 1-4 Article 4

2014

Improved Functional Prediction of Hypothetical Proteins from *Listeria monocytogenes* 08-5578

Michael J. Lagier
Grand View University

Brittany Bowman
Grand View University

Kelsey Brend
Grand View University

Katherine Hobbs
Grand View University

Michael Foggia
Grand View University

See next page for additional authors

Follow this and additional works at: https://scholarworks.uni.edu/jias

 Part of the Anthropology Commons, Life Sciences Commons, Physical Sciences and Mathematics Commons, and the Science and Mathematics Education Commons

Improved Functional Prediction of Hypothetical Proteins from *Listeria monocytogenes* 08-5578

Authors

Michael J. Lagier, Brittany Bowman, Kelsey Brend, Katherine Hobbs, Michael Foggia, and Mark McDaniel

Jour. Iowa Acad. Sci. 121(1–4):16–27, 2014

Improved Functional Prediction of Hypothetical Proteins from *Listeria monocytogenes* 08-5578

MICHAEL J. LAGIER*, BRITTANY BOWMAN, KELSEY BREND, KATHERINE HOBBS, MICHAEL FOGGIA and MARK MCDANIEL

Department of Biology, Grand View University, Des Moines, IA, 50316

Listeria monocytogenes is a foodborne human pathogen responsible for listerosis. The genomes of several *L. monocytogenes* strains have been recently sequenced. The genome of *L. monocytogenes* 08-5578, which was in part responsible for a significant listerosis outbreak in 2008, contains an unexpectedly high percentage of protein-encoding genes (1,927 out of 3,161; 60.96%) autonomously annotated as hypothetical proteins. The aim of this study was to test whether a manual annotation strategy could be used to assign more meaningful functional names to the hypothetical proteins of 08-5578. A holistic, manual gene annotation strategy that utilized sequence homology, cellular localization predictions, structure-based evidence, phylogeny, and protein-protein interaction data was used to assign potential cellular roles to 79 out of 100 hypothetical proteins randomly selected from the genome of 08-5578. Of significance, 5 of the 79 hypothetical proteins assigned a more meaningful name may contribute to the virulence of *L. monocytogenes* 08-5578, by contributing to chemotaxis, cell surface protein sorting, cell wall biosynthesis, and cold adaptation. The findings here support the notion that manual annotations, using a combination of diverse bioinformatics tools, can improve the quality of genomic information provided by automated genome annotation methods alone.

INDEX DESCRIPTORS: *Listeria monocytogenes*, hypothetical protein, annotation.

INTRODUCTION

The Gram-positive bacterium *Listeria monocytogenes* is a pathogen that can cause disease in humans, mammals, and avians (Hernandez-Milian and Payeras-Cifre, 2014). In severe cases, the central nervous system can be infected, leading to meningitis and the formation of brain abscesses (Hernandez-Milian and Payeras-Cifre, 2014). The great majority of human listeriosis cases arise as a result of consuming contaminated food products including deli meats, smoked seafood, dairy products, fresh fruits, and raw vegetables (Ferreira et al., 2014). Overall, in the United States, listeriosis is the third leading cause of death from food poisoning, despite the widespread availability of antimicrobial therapy (Lomonaco et al., 2015).

Given the clinical significance of listeriosis, several genomes of *L. monocytogenes* strains have been sequenced in an effort to better understand the lifestyle and pathogenicity of the species, as well as the potential influence of strain variability on virulence (Reddy and Lawrence, 2014). Among those recently sequenced includes a strain, *L. monocytogenes* 08-5578 (serotype 1/2a), that was in part responsible for a listeriosis outbreak in Canada during 2008 (Gilmour et al., 2010). The Canadian outbreak was nationwide, associated with ready-to-eat meat products, and resulted in 22 deaths from a total of 57 illnesses (Gilmour et al., 2010).

The genomes of bacteria contain protein-encoding genes for which bioinformatics cannot assign a potential cellular function during the genome annotation process (Ijaq et al., 2015). Such genes are most commonly termed hypothetical proteins (HP's). Genome annotation is the process of attaching biological information to DNA, RNA, and/or protein sequences. Many bacterial genomes contain a significant (up to 40%) portion of genes that are annotated as HP's (Ijaq et al., 2015). Even the most well-known species, including *Escherichia coli* and *Bacillus subtilis*, encode HP's within their perspective genomes (Bork, 2000). Given the widespread occurrence of HP's in bacterial genomes, and proteomic studies that show that HP's are likely to be expressed as proteins, it is probable that many HP's have important biological roles (Ijaq et al., 2015). Hence, efforts to further define the biological roles of HP's are justified – especially with regard to better understanding significant human pathogens (Hernández et al., 2009). Progress has been made in using a combination of discrete bioinformatics tools (STRING, manual text-mining, BLAST, SignalP) to improve the annotation of HP's from pathogens (Silber and Pereira, 2012; Kumar et al., 2014). For example, a combination of STRING and manual text-mining was used to assign potential biological functions for more than 50% of the previously annotated HP's of *Mycobacterium tuberculosis* (Doerks et al., 2012).

In examining the available genomes of *L. monocytogenes*, it was noted that the important outbreak-associated strain – *L. monocytogenes* 08-5578 – had a significant portion of its genes (1,927 of 3,161 total genes) autonomously annotated as HP's. For *L. monocytogenes* 08-5578, the percentage of genes annotated as HP's is 60.96% (www.img.jgi.doe.gov). In comparison, genomes of additional *L. monocytogenes* strains have 15% to 20% of their genes assigned the classification of HP (https://img.jgi.doe.gov/cgi-bin/mer/main.cgi?section=FindGenomes). The reason for this discrepancy in HP's for strain 08-5578 is unknown.

In this study, a holistic, manual annotation strategy was used in an effort to better assign potential biological functions to a random sampling of 100 *L. monocytogenes* 08-5578 HP's. Using

* Corresponding author: mlagier@grandview.edu
515-263-2873

a combination of sequence similarity tools (BLASTp, CDD, T-Coffee, KEGG), cellular localization data (TMHMM, SignalP, PSORT-B, Phobius), structure-based evidence (Pfam, PDB), protein-protein interaction predictions (STRING), and phylogenetic placements (Phylogeny.fr); cellular roles were predicted for 79 of 100 HP's analyzed. Significantly, a total of 5 HP's assigned a more meaningful functional prediction are genes that may be related to the disease-causing abilities of *L. monocytogenes* 08-5578, by contributing to chemotaxis, cell surface protein sorting, alanine metabolism, and cold adaptation (Reddy and Lawrence, 2014).

MATERIALS AND METHODS

Selection and Retrieval of Hypothetical Proteins

Proteins annotated as hypothetical proteins (HP's) in the genome of *L. monocytogenes* 08-5578 were retrieved from the Integrated Microbial Genomes database (IMG, http://img.jgi.doe.gov/cgi-bin/edu/main.cgi). The 1,927 HP's were downloaded into Microsoft Excel, and 100 (5.19% of the total HP's) were randomly selected for further study.

BLASTp Screening of Hypothetical Protein Set

The 100 selected HP's were initially examined using standard BLASTp searches (Altschul et al., 1990), as summarized in Table 1. The order of the HP's presented in Table 1 represent the order in which genes were randomly selected and analyzed. Individual FASTA protein sequences of HP's, used as queries, that yielded significant E-values (less than 1.00E-3) to previously deposited database proteins are included in Table 1. The 5 *L. monocytogenes* 08-5578 HP's that displayed significant sequence identity and similarity to genes previously shown to play a role in the disease-causing phenotypes of bacterial pathogens (highlighted in Table 1) were further annotated.

Sequence Similarity Analyses

Selected HP sequences were aligned using T-Coffee (Magis et al., 2014). Conserved functional domains within HP's were identified using queries of the Conserved Domain Database (www.ncbi.nlm.nih.gov/cdd). The placement of HP's within previously described cellular pathways and/or structures was accomplished by examining the KEGG database (Kanehisa and Goto, 2000). All programs used for similarity analyses were run according to default parameters.

Cellular Localization Predictions

TMHMM and Phobius (Krogh et al., 2001; Käll et al., 2004) were used (default parameters) to identify transmembrane regions within *L. monocytogenes* 08-5578 HP's. The cellular locations of HP's were predicted using a combination of SignalP and PSORT-B, using default settings (Petersen et al., 2011).

Structure-based Annotation of Hypothetical Proteins

Retrieved protein FASTA files of HP's were used as queries for structure-based annotations. HP's sharing significant sequence homology with well-described protein families were identified by probing the Pfam database (http://pfam.xfam.org/search) using standard settings. HP's sharing significant homology with proteins for which a structure has been experimentally elucidated were identified using the default search function of PDB (www.rcsb.org/pdb/home/home.do).

Protein-Protein Interactions

The STRING tool was used to identify proteins that *L. monocytogenes* HP's may interact with *in vivo*. The STRING database contains information from experimental data and public text collections to predict protein-protein interactions (Szklarczyk et al., 2015). The basic interaction unit in STRING is the functional association, likely contributing to a common biological purpose. Predicted protein-protein interactions are derived from multiple sources including known experimental interactions (co-expression experiments), genome functional pathway knowledge from manually curated databases (databases), automated text-mining searches (text-mining), shared gene neighborhood locations across multiple genomes (neighborhood), and interactions observed in one organism and systematically transferred to other organisms, via pre-computed orthology relations (co-occurrence and gene fusion). Default parameters of STRING were used in this study.

Phylogenetic Placements

Phylogenetic trees were created for HP's using Phylogeny.fr (www.phylogeny.fr/). For each, 20 sequences sharing significant sequence identity and similarity (BLAST-EXPLORER, standard settings) with the given HP were aligned by T-Coffee. The resultant alignments were used to construct maximum-likelihood trees (default settings). The finalized, consensus trees were visualized with TreeDyn (Dereeper et al., 2008). For all trees presented, proteins from closely related species and groups, and with the same predicted functionality, clustered with the query protein (HP's). In all cases, the observed clustering patterns supported the predicted functionality of analyzed HP's.

RESULTS AND DISCUSSION

BLASTp Analysis of *Listeria monocytogenes 08-5578*

In examining the number and percentage of hypothetical proteins (HP's) per genome of *L. monocytogenes*, it was noted that clinical isolate 08-5578 had a significantly higher percentage of protein-encoded genes annotated as HP's: specifically 60.96%. To determine if this value was a result of incomplete gene annotation, or is a unique characteristic of 08-5578, 100 genes annotated as HP's were randomly selected for a standard BLASTp examination. A total of 79 out of 100 tested genes shared significant sequence identity and similarity to proteins annotated as non-HP's (Table 1). All HP's tested yielded significant sequence identity and similarity to previously deposited database proteins, as indicted by E-values of less than 1.00E-3. The 79 HP's assigned more meaningful, functional names are predicted to play roles in a variety of cellular processes including membrane transport, transcriptional regulation, binary fission, chemotaxis, and nucleic acid metabolism (Table 1).

For all 79 HP's, additional analyses (CDD data and Pfam searches) yielded data consistent with the functional names identified by BLASTp (data not shown). The remaining 21 HP's tested by BLASTp share significant sequence homology with proteins annotated as hypothetical proteins from related bacteria. Therefore, these 21 proteins cannot be assigned a more specific,

Table 1. BLASTp analysis of selected hypothetical proteins (HP's).

HP Query Identifier[a]	Improved Functional Name[b,c]	BLASTp Hit Species	GenBank Hit Number	E-value	Identity (%)	Similarity (%)	Hit Query Coverage (%)
LM5578_2142	Peptidoglycan-binding protein	*Listeria monocytogenes*	WP_003723588.1	2.00E-166	100	100	100
LM5578_1063	Mulitdrug transporter	*Listeria monocytogenes*	WP_003732434.1	0.00E+00	100	100	100
LM5578_0153	**Hypothetical protein**	*Listeria ivanovii*	WP_025280134.1	9.00E-44	97	100	99
LM5578_2982	PTS beta-glucoside transporter	*Listeria monocytogenes*	WP_009924399.1	0.00E+00	100	100	100
LM5578_2915	NADPH nitroreductase	*Listeria monocytogenes*	WP_012952087.1	9.00E-152	100	100	100
LM5578_2259	Heme A synthase	*Listeria monocytogenes*	WP_003731964.1	0.00E+00	100	100	100
LM5578_0329	ABC transporter	*Listeria monocytogenes*	WP_003722903.1	2.00E-49	100	100	99
LM5578_2500	**Hypothetical protein**	*Listeria seeligeri*	WP_003744985.1	8.00E-42	92	100	97
LM5578_2159	Ferrichrome ABC transporter	*Listeria monocytogenes*	WP_003733843.1	0.00E+00	99	100	99
LM5578_2004	DNA-binding protein	*Listeria monocytogenes*	WP_003723862.1	1.00E-69	100	100	100
LM5578_0836	ABC transporter	*Listeria monocytogenes*	WP_009925081.1	0.00E+00	100	100	100
LM5578_2528	**Hypothetical protein**	*Listeria monocytogenes*	WP_026747212.1	9.00E-95	99	100	98
LM5578_2620	Gycine cleavage system H protein	*Listeria monocytogenes*	WP_003723327.1	4.00E-81	100	100	100
LM5578_1666	N-acetylmuramoyl-L-alanine amidase	*Listeria monocytogenes*	WP_003723525.1	0.00E+00	100	100	99
LM5578_2591	**Hypothetical protein**	*Listeria innocua*	EFR89811.1	1.00E-45	99	100	98
LM5578_1493	Peptidase M24	*Listeria monocytogenes*	WP_003722481.1	0.00E+00	100	100	100
LM5578_0802	Methyl-accepting chemotaxis protein	*Listeria monocytogenes*	CCQ20072.1	0.00E+00	96	95	100
LM5578_1048	GTP pyrophosphokinase	*Listeria monocytogenes*	WP_003722788.1	5.00E-162	100	100	100
LM5578_1141	YktA protein	*Bacillus nealsonii*	WP_016200967.1	3.00E-26	53	100	96
LM5578_1482	Competence protein ComG	*Listeria monocytogenes*	WP_003722471.1	3.00E-62	100	100	100
LM5578_1403	Phage-related protein	*Listeria innocua*	EFR90990.1	6.00E-28	83	92	98
LM5578_0516	Hydrolase	*Listeria monocytogenes*	WP_010989494.1	0.00E+00	100	100	100
LM5578_0900	**Hypothetical protein**	*Listeria monocytogenes*	WP_045131607.1	3.00E-51	99	100	97
LM5578_0768	Chemotaxis protein	*Listeria monocytogenes*	WP_003721810.1	0.00E+00	100	100	100
LM5578_2438	NADH oxidase	*Listeria monocytogenes*	AGR02026.1	0.00E+00	99	100	95
LM5578_2640	Glycosyl hydrolase	*Listeria monocytogenes*	WP_025370653.1	0.00E+00	100	100	100
LM5578_1753	Cell division protein	*Listeria monocytogenes*	WP_012951662.1	0.00E+00	100	100	100
LM5578_2373	2-nitropropane dioxygenase	*Listeria monocytogenes*	WP_003722294.1	0.00E+00	100	100	100
LM5578_1015	Iron sulfur cluster binding protein	*Listeria monocytogenes*	WP_031645503.1	0.00E+00	100	100	100
LM5578_0021	Alcohol dehydrogenase	*Listeria monocytogenes*	WP_003723841.1	0.00E+00	100	100	100
LM5578_0980	Glucan biosynthesis	*Paenibacillus species*	WP_036691738.1	2.00E-06	51	68	88
LM5578_2409	Chaperone protein	*Listeria monocytogenes*	WP_012951904.1	0.00E+00	100	100	100
LM5578_2645	Exoribonuclease R	*Listeria monocytogenes*	WP_004892474.1	0.00E+00	99	100	100
LM5578_0767	Glycosyl transferase	*Listeria monocytogenes*	WP_012951346.1	0.00E+00	100	100	99
LM5578_0111	Pyridoxamine 5′-phosphate oxidase	*Listeria monocytogenes*	WP_003722108.1	3.00E-95	100	100	99
LM5578_0087	6-phospho-beta-glucosidase	*Listeria monocytogenes*	WP_012951080.1	0.00E+00	100	100	99
LM5578_0446	**Hypothetical protein**	*Listeria monocytogenes*	EFR85778.1	2.00E-29	87	88	85
LM5578_2421	**Hyothetical protein**	*Listeria monocytogenes*	WP_003731889.1	2.00E-100	100	100	100
LM5578_2598	**Hypothetical protein**	*Bacillus species*	WP_026574418.1	9.00E-111	44	64	87
LM5578_0140	Acyl-CoA ligase	*Listeria monocytogenes*	WP_012951097.1	0.00E+00	100	100	100
LM5578_2834	Oligo-1, 6-glucosidase	*Listeria monocytogenes*	WP 003733141.1	0.00E+00	99	100	99
LM5578_0183	Alpha-hydrolase	*Listeria monocytogenes*	WP 031645386.1	3.00E-05	99	100	98
LM5578_1631	RNA-binding protein	*Listeria monocytogenes*	WP 012681299.1	6.00E-60	99	100	98

Table 1. Continued

HP Query Identifier[a]	Improved Functional Name[b,c]	BLASTp Hit Species	GenBank Hit Number	E-value	Identity (%)	Similarity (%)	Hit Query Coverage (%)
LM5578_1156	ABC transporter	*Listeria monocytogenes*	WP 014600718.1	0.00E+00	99	100	98
LM5578_2363	Xylose isomerase	*Listeria monocytogenes*	WP 003729565.1	0.00E+00	99	100	98
LM5578_2586	Voltage-gated ion channel	*Listeria monocytogenes*	CBY52861.1	0.00E+00	99	100	98
LM5578_0765	Chemotaxis protein MotB	*Listeria marthii*	EFR88514.1	0.00E+00	99	100	100
LM5578_0728	**Hyothetical protein**	*Listeria monocytogenes*	WP 031645255.1	0.00E+00	99	100	99
LM5578_1059	Esterase family protein	*Listeria monocytogenes*	EGJ24480.1	2.00E-180	98	100	98
LM5578_0585	Phosphoglycerate mutase	*Listeria monocytogenes*	WP 039387077.1	2.00E-164	99	100	98
LM5578_1349	Hydrolase	*Listeria monocytogenes*	YP 009044871.1	8.00E-69	98	100	97
LM5578_1369	Holin protein	*Listeria monocytogenes*	WP 039380817.1	4.00E-53	99	100	97
LM5578_2486	**Hypothetical protein**	*Listeria monocytogenes*	CDM20221.1	2.00E-46	99	100	99
LM5578_2593	**Hyothetical protein**	*Listeria monocytogenes*	CDM17735.1	1.00E-17	97	100	96
LM5578_0440	Glyoxalase	*Listeria monocytogenes*	WP003731407.1	3.00E-86	99	100	100
LM5578_1412	DNA processing protein DprA	*Listeria monocytogenes*	WP 003723892.1	0.00E+00	100	100	100
LM5578_2285	Transcriptional regulator	*Listeria monocytogenes*	WP 031644767.1	2.00E-149	99	100	100
LM5578_2231	RNA-binding protein S4	*Listeria monocytogenes*	WP 003731984.1	0.00E+00	100	100	100
LM5578_0099	RNase III inhibitor	*Listeria monocytogenes*	WP 009930801.1	2.00E-123	99	100	99
LM5578_1740	Aminotransferase V	*Listeria monocytogenes*	WP 003723322.1	0.00E+00	100	100	99
LM5578_2251	ABC transporter	*Listeria monocytogenes*	WP_012951862	0.00E+00	100	100	99
LM5578_0886	Spermidine ABC transporter	*Listeria monocytogenes*	WP_003732945	0.00E+00	100	100	99
LM5578_1146	ABC transporter	*Listeria monocytogenes*	WP_003722690	3.00E-161	100	100	97
LM5578_0336	YycH protein	*Listeria marthii*	EFR88944	0.00E+00	97	98	94
LM5578_2771	Cation transporter	*Listeria monocytogenes*	WP_003723624	0.00E+00	100	100	100
LM5578_0457	Glucose uptake protein	*Listeria monocytogenes*	WP_009912760	0.00E+00	100	100	100
LM5578_0284	RNA-binding protein	*Listeria marthii*	EFR89004	5.00E-118	99	100	100
LM5578_p41	**Hypothetical Protein**	*Listeria monocytogenes*	WP_025370671	3.00E-150	99	99	99
LM5578_1217	Propanediol protein PduU	*Listeria monocytogenes*	WP_003721543	5.00E-76	100	100	99
LM5578_2812	**Hypothetical protein**	*Listeria monocytogenes*	EHY61881	3.00E-12	82	92	97
LM5578_1624	DNA competence protein	*Listeria monocytogenes*	EEW13475	0.00E+00	99	99	96
LM5578_1977	Phosphohydrolase	*Listeria monocytogenes*	WP_003723983	9.00E-143	100	100	100
LM5578_0587	Magnesium transporter	*Listeria monocytogenes*	WP_003721353	0.00E+00	100	100	100
LM5578_0963	**Hypothetical protein**	*Listeria monocytogenes*	WP_003721450	0.00E+00	100	100	100
LM5578_1010	Sortase family protein	*Listeria monocytogenes*	WP_003722751	2.00E-158	100	100	100
LM5578_2526	**Hypothetical protein**	*Listeria monocytogenes*	WP_009911346	6.00E-116	99	99	99
LM5578_1065	Peroxiredoxin	*Listeria monocytogenes*	WP_003722848	1.00E-111	100	100	100
LM5578_2974	Cytochrome O oxidase	*Listeria monocytogenes*	WP_003721665	3.00E-153	100	100	100
LM5578_2343	ABC transporter	*Listeria monocytogenes*	WP_009932067	0.00E+00	99	100	100
LM5578_2495	DNA-binding protein	*Listeria ivanovii*	EFR97160	4.00E-127	98	100	97
LM5578_0872	**Hypothetical protein**	*Lactobacillus pentosus*	CCC17531.1	8.00E-10	27	48	78
LM5578_0966	Alanine racemase	*Listeria monocytogenes*	WP_014601850.1	0.00E+00	99	99	99
LM5578_1101	DNA-binding protein	*Listeria monocytogenes*	WP_003722883.1	0.00E+00	99	100	100
LM5578_2524	DNA-binding protein	*Listeria monocytogenes*	WP_031660006.1	1.00E-49	99	100	99
LM5578_1157	Teichoic acid permease	*Listeria monocytogenes*	WP_038409749.1	0.00E+00	99	99	99
LM5578_1377	**Hypothetical protein**	*Listeria monocytogenes*	EFR84828.1	2.00E-130	94	94	90
LM5578_0907	Transposase protein	*Listeria monocytogenes*	WP_031673407.1	4.00E-125	100	100	100
LM5578_0401	Deferrochelatase	*Listeria monocytogenes*	WP_046335455.1	0.00E+00	99	100	98
LM5578_2942	**Hypothetical protein**	*Listeria monocytogenes*	WP_003772666.1	1.00E-96	99	99	98
LM5578_2465	HAD family hydrolase	*Listeria monocytogenes*	WP_010989935.1	0.00E+00	99	100	99
LM5578_0533	**Hypothetical protein**	*Listeria monocytogenes*	WP_009930754.1	4.00E-130	99	99	95

Table 1. Continued

HP Query Identifier[a]	Improved Functional Name[b,c]	BLASTp Hit Species	GenBank Hit Number	E-value	Identity (%)	Similarity (%)	Hit Query Coverage (%)
LM5578_0473	DNA-directed RNA polymerase	*Listeria monocytogenes*	WP_039119694.1	1.00E-113	99	100	93
LM5578_2147	Acyl-CoA hydrolase	*Listeria monocytogenes*	WP_003728769.1	1.00E-121	99	99	90
LM5578_2348	**Hypothetical protein**	*Listeria monocytogenes*	WP_014602158.1	0.00E+00	99	100	100
LM5578_1552	XRE transcriptional regulator	*Listeria monocytogenes*	WP_003721916.1	4.00E-45	100	100	100
LM5578_1409	Signal peptidase I	*Listeria monocytogenes*	WP_003727520.1	1.00E-124	99	100	96
LM5578_0138	Multidrug transporter	*Listeria monocytogenes*	WP_031674857.1	2.00E-176	99	99	95
LM5578_0692	Terminase protein	*Streptococcus pyogenes*	WP_011184601	5.00E-44	46	63	93
LM5578_1224	Microcompartment protein	*Listeria monocytogenes*	WP_003719356.1	7.00E-58	99	100	87
LM5578_0691	**Hypothetical protein**	*Listeria monocytogenes*	EXL15680.1	2.00E-12	70	76	98

[a]Locus tag (find gene function) at https://img.jgi.doe.gov/cgi-bin/mer/main.cgi
[b]Bolded HP's could not be assigned a more meaningful functional name following analysis
[c]Shaded HP's represent proteins potentially related to listeria virulence

potential cellular function based on BLASTp alone (Table 1, bolded text).

The observation that manual, default BLASTp searching assigned more meaningful functional names to 79% of randomly-selected HP's suggests that errors accumulated during the original, automated annotation of the *L. monocytogenes* 08-5578 genome (Gilmour et al., 2010), which made use of a combination of BASys, Glimmer, GenDB and Artemis software. Automated annotation pipelines, when not supported by comprehensive manual annotation efforts to verify the automated data, may lead to erroneous functional predictions (Schnoes et al., 2009; Poptsova and Gogarten, 2010; Promponas et al., 2015). Therefore, it does not appear that strain 08-5578 has an unusually high percentage of HP's, but rather that the percent of HP's originally observed (60.96%) was likely caused by automated annotation errors. In contrast, the sample of HP's examined by BLASTp suggests that about 21% of the protein-encoded genes of 08-5578 represent HP's, which is consistent with *L. monocytogenes* genomes from closely related species.

Interestingly, the BLASTp experiment identified 5 genes, previously assigned as HP's, as genes that may play a role in the pathogenicity of *L. monocytogenes* (Table 1, highlighted in gray). This finding in particular supports the need for significant manual annotation efforts in predicting cellular functions to likely protein-encoding genes. The 5 genes that may contribute to the pathogenicity of 08-5578 were subjected to a more thorough annotation process, beyond BLASTp analysis, to better understand their potential cellular roles. Recent studies have suggested that annotation efforts that examine protein-encoding genes from multiple perspectives, including sequence homology, phylogeny, structural features, subcellular localizations and protein interactions, may yield more robust functional predictions (Silber and Pereira, 2012; Kumar et al., 2014). An effort was also made to better annotate the 21 HP's that remained HP's following the initial BLASTp searches. Even with the more comprehensive annotation process, these 21 HP's appear to represent bona fide HP's, for which additional "wet laboratory" experiments are needed to gain any insight into their cellular roles (data not shown).

Manual Annotation of Select Hypothetical Proteins

The observation was made that 5 genes previously annotated as HP's share significant sequence identity and homology to proteins that may contribute to the disease-causing phenotypes of *L. monocytogenes.* The 5 genes, according to shared sequence homology (BLASTp), are alanine racemase, sortase A, a methyl-accepting chemotaxis protein (MCP), an exoribonuclease R protein, and a flagella-associated MotB protein (Table 2).

Alanine Racemase

Alanine racemase is a conserved bacterial enzyme that converts L-alanine to D-alanine. D-alanine is required for the synthesis of peptidoglycan, which is an essential component of bacterial cell walls (Radkov and Moe, 2014). Alanine racemase is considered a promising target for antimicrobials, as it appears to be lacking in eukaryotes and may play a role in virulence (Thompson et al., 1998; Strych et al., 2000; Radkov and Moe, 2014). Consistent with BLASTp, a multiple sequence alignment of LM5578_0966 using T-Coffee identified an amino acid residue (lysine, K^{39}) important to the enzymatic function of racemases (Watanabe et al., 1999), and conserved among known alanine racemases (Table 3).

In addition, queries of the PDB and Pfam using LM5578_0966 yielded results suggesting that LM5578_0966 is more likely to represent an alanine racemase than a HP (Table 3). LM5578_0966 is also expected to reside within the cytoplasm, according to TMHMM, SignalP, PSORT-B, and Phobius (Table 4). STRING predicts that LM5578_0966 functionally interacts with proteins that may play a role in alanine metabolism, including cysteine desulfurases (which remove sulfur groups from L-cysteine to produce L-alanine) and proteins involved in cell wall synthesis (Fig. 1A). Lastly, phylogeny clusters LM5578_0966 among alanine racemases from *Listeria species* (Fig. 1B).

Sortase A

Sortase A is a membrane-associated protein that is involved in the physical attachment of other proteins to the external surface of

Table 2. Summary of hypothetical proteins potentially related to virulence.

HP Identifier[a]	Predicted Function After Manual Annotation	DNA Gene Length (Bases)	Protein Length (Amino Acids)
LM5578_0802	Methyl-accepting chemotaxis protein (MCP)	1,806	601
LM5578_2645	Exoribonuclease R	2,382	793
LM5578_0765	Chemotaxis protein MotB	828	275
LM5578_1010	Sortase A	669	222
LM5578_0966	Alanine racemase	1,107	368

[a]Locus tag (find gene function) at https://img.jgi.doe.gov/cgi-bin/mer/main.cgi

prokaryotes. A previous mutagenesis study showed that sortase A from *L. monocytogenes* EGD is involved in virulence (Bierne et al., 2002). According to homology-based and structure-based analyses, LM5578_1010 represents a sortase A, not a HP (Table 3), including the conservation of a key cysteine residue (Bradshaw et al., 2015). Localization predictions of LM5578_1010 were consistent with known sortases being embedded within the cytoplasmic membrane of related bacteria (Table 4).

LM5578_1010, like other sortases, does not contain a typical signal peptide for membrane-targeting, but does contain a hydrophobic region at the N-terminus (the membrane-spanning region identified by Phobius and TMHMM, Table 4) that acts as a signal peptide (Bradshaw et al., 2015). STRING data also support LM5578_1010 annotated as a sortase A; specifically, LM5578_1010 is predicted to functionally interact with several proteins that are involved with the process of delivering and sorting proteins at the cell surface (for example, signal peptidases; Fig. 2A). Also of significance, STRING predicts that LM5578_1010 functionally interacts with internalin J. In *L. monocytogenes* EGD, internalin J is not only critical for entry of the bacteria into host cells during infection, the protein requires sortase A for attachment to the cell wall. Indeed, *L. monocytogenes* EGD lacking sortase A activity displays decreased virulence in mouse models – in part due to a decrease in surface-bound internalin J (Milohanic et al., 2000). As expected, LM5578_1010 clusters with closely related sortase A proteins when subjected to phylogenetic analysis (Fig. 2B).

Chemotaxis Proteins

According to manual annotation efforts in this study, LM5578_0802 and LM5578_0765 appear to encode for conserved proteins related to bacterial chemotaxis, rather than HP's (Table 2). Annotation according to shared sequence homology suggests that LM5578_0802 encodes for a methyl-accepting chemotaxis protein, and LM5578_0765 represents a flagellar-associated protein MotB (Table 3). In both cases, the proteins contain conserved amino acid residues – glutamate E^{575} (MCP), and arginine R^{279} (MotB, Table 3) – which are believed to be important to functionality (Baker et al., 2006).

The flagellar-associated protein MotB is a structural component of the flagellar motor, and contributes to generating the energy needed for movement (Baker et al., 2006). A previous study removing the functionality of MotB in *L. monocytogenes* 10403S showed that MotB was required for motility and efficient colonization of mouse intestines (O'Neil and Marquis, 2006). Thus, MotB appears to contribute to *L. monocytogenes* virulence.

Table 3. Features of virulence-related proteins according to homology and structural analyses.

HP Identifier[a]	Name	[a]CDD Domain Hit (name)	[b]Key Residue Identified (name)	[c]KEGG placement (name)	[d]Pfam Hit (name)	[e]PDB Hit (name)
LM5578_0802	MCP	cd11386 (MCP signal domain)	Yes (E^{575})	Yes (bacterial chemotaxis)	PF12729 (chemotaxis signaling domain)	3ZX6 (bacterial chemotaxis receptor)
LM5578_2645	Exoribonuclease R	cd04471 (Rnase R)	Yes (R^{572})	Yes (RNA degradosome)	PF08206 (Ribonuclease B OB domain)	4PMW (DIS3-like exonuclease 2)
LM5578_0765	MotB	cd07185 (OmpA_C-like)	Yes (R^{279})	Yes (flagellar assembly)	PF13677 (MotB_plug)	3SOY (domain of motB)
LM5578_1010	Sortase A	cd06165 (Sortase_A_1)	Yes (C^{188})	Yes (sortases)	PF04203 (Sortase)	2KW8 (Sortase structure)
LM5578_0966	Alanine Racemase	cd00430 (Racemase)	Yes (K^{39})	Yes (D-alanine metabolism)	PF01168 (Ala_racemase)	1BD0 (alanine racemase)

[a]Top hit with significant expected value according to http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi
[b]According to T-Coffee alignment at http://www.ebi.ac.uk/Tools/msa/tcoffee/
[c]Using KEGG at http://www.genome.jp/kegg/kegg1d.html
[d]Top hit with significant homology at http://pfam.xfam.org/
[e]Top hit with significant E-value using http://www.rcsb.org/pdb/home/home.do

Table 4. Localization properties of select hypothetical proteins.

HP Identifier[a]	Name	[a]Consensus Localization	[b]Membrane Spanning Regions (Number)	[c]Signal Peptide Detected	[d]PSORT-B Prediction (Score)
LM5578_0802	MCP	cell membrane	Yes (2)	No	cytoplasmic membrane (8.78)
LM5578_2645	Exoribonuclease R	cytplasm	No	No	cytoplasm (9.97)
LM5578_0765	MotB	cell membrane	Yes (1)	No	cytoplasmic membrane (8.78)
LM5578_1010	Sortase A	cell membrane	Yes (1)	No	cytoplasmic membrane (9.95)
LM5578_0966	Alanine Racemase	cytoplasm	No	No	cytoplasm (9.97)

[a]According to overall trends observed with TMHMM, Phobius, SignalP, and PSORT-B
[b]According to TMHMM (http://www.cbs.dtu.dk/services/TMHMM/) and Phobius (http://phobius.sbc.su.se/)
[c]Using SignalP at http://www.cbs.dtu.dk/services/SignalP/
[d]Using PSORT-B at http://www.psort.org/psortb/(score greater than 7.50 is considered significant)

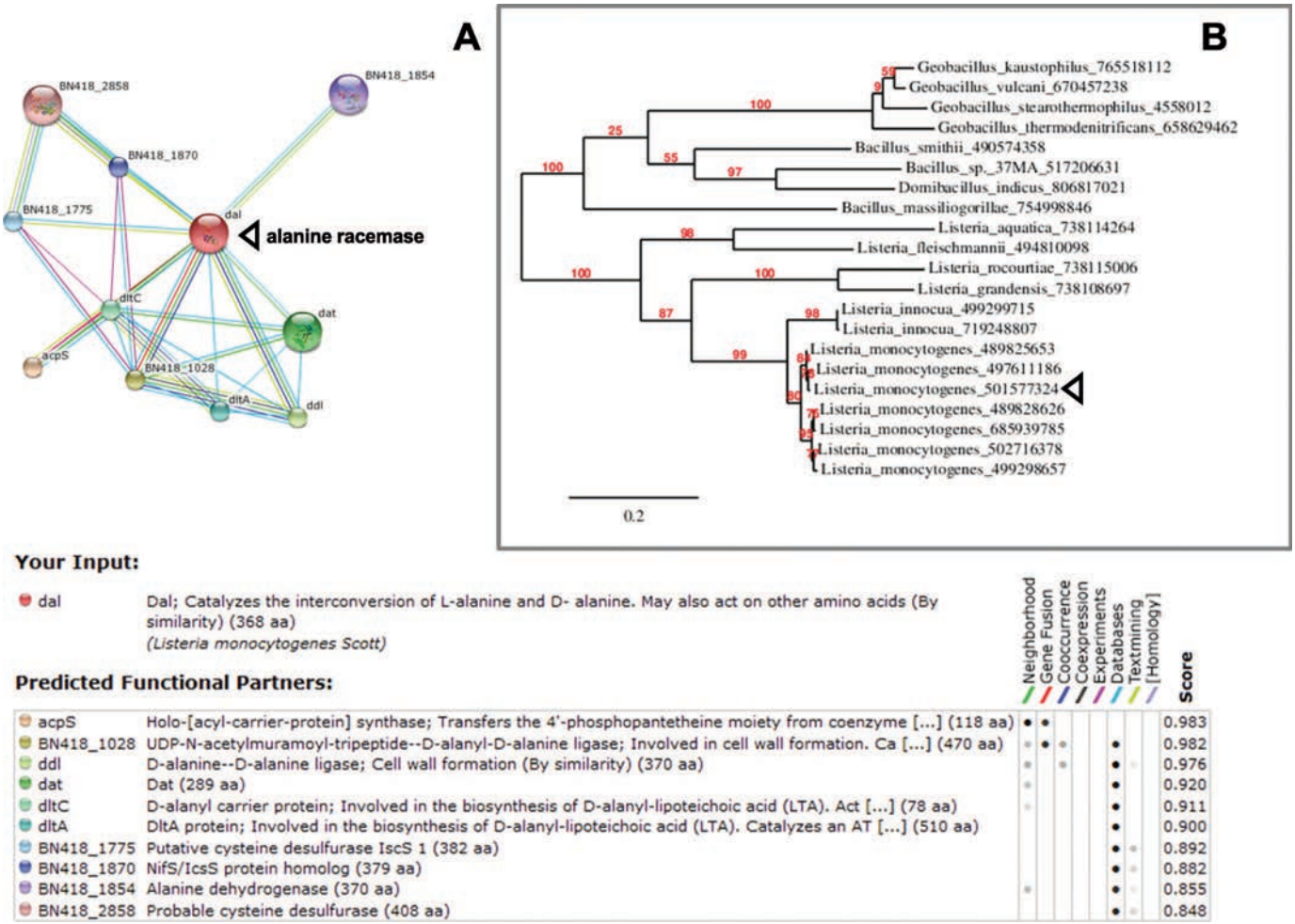
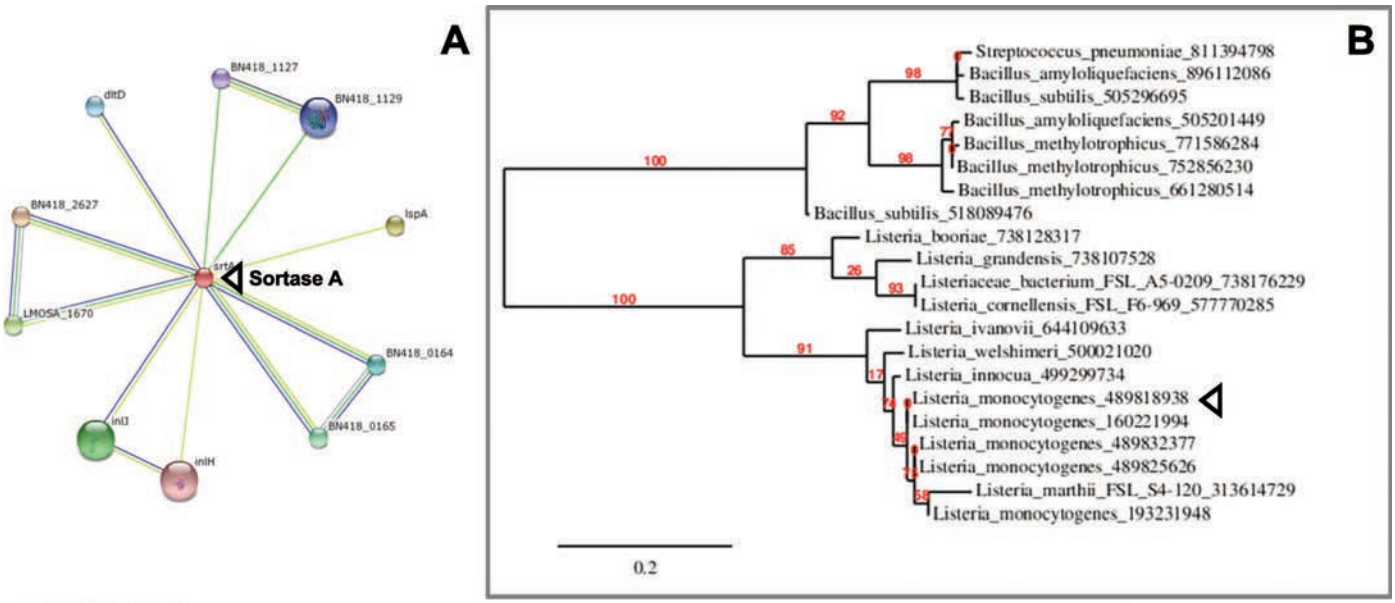


Fig. 1. STRING protein-protein interactions and phylogenetic placement of alanine racemase. A. STRING network map and summary table of predicted interactions. The category of interaction is shown in the table. Color saturation of the edges represents the confidence score of a given functional or physical association. The position of alanine racemase is indicated by an arrowhead. Shown interactions are considered high confidence, yielding confidence scores of greater than 0.7. Black and gray dots indicate the individual types of interactions (see Methods for more information about interaction types) predicted to occur between the query protein and the given functional partner. Black dots contribute to the overall confidence score to a greater degree than gray. B. Phylogeny of alanine racemase (marked with an arrowhead) and related proteins. Scale represents evolutionary distance. Values are branch support. All sequences shown are predicted alanine racemases.



Your Input:

		Neighborhood	Gene Fusion	Cooccurrence	Coexpression	Experiments	Databases	Textmining	[Homology]	Score
srtA	Sortase A (222 aa) *(Listeria monocytogenes Scott)*									

Predicted Functional Partners:

BN418_2627	Undefined function (1710 aa)	•		•				•		0.740
lspA	Signal peptidase II; This protein specifically catalyzes the removal of signal peptides from pr [...] (154 aa)							•		0.720
LMOSA_1670	Cell wall surface anchor family protein (1529 aa)	•		•				•		0.710
inlJ	Internalin J (916 aa)			•				•		0.678
BN418_0165	Wall-associated protein (583 aa)	•		•						0.667
BN418_0164	Wall-associated protein (705 aa)	•		•						0.634
dltD	DltD protein (424 aa)			•				•		0.632
BN418_1129	Uncharacterized protein yhfI (243 aa)	•								0.586
BN418_1127	Putative 3-methyladenine DNA glycosylase (207 aa)	•								0.579
inlH	InlC2 (548 aa)							•		0.574

Fig. 2. STRING interactions and phylogeny of sortase A. A. STRING map and table of interactions. The category of interaction is shown in the table. Color saturation represents confidence score of association. The position of sortase A is indicated by an arrowhead. Shown interactions are considered medium confidence, yielding scores of greater than 0.4. Scores higher than 0.7 are defined as high confidence. Black and gray dots indicate the individual types of interactions (see Methods for more information about interaction types) predicted to occur between the query protein and the given functional partner. Black dots contribute to the overall confidence score to a greater degree than gray. B. Phylogeny of sortase A (arrowhead) and related proteins. Scale represents evolutionary distance. Values are branch support. All sequences shown are predicted sortases.

By way of localization predictions, LM5578_0765 appears to be anchored into the cytoplasmic membrane by a single alpha-helix at the N-terminus of the protein. The remaining protein, downstream of the alpha-helix, extends to the exterior, away from the cytoplasmic membrane (Table 4). This topology is expected for MotB, as MotB homologs from related bacteria share an identical cellular localization and cytoplasmic membrane orientation (Kojima and Blair, 2004).

In terms of expected interactions, MotB is expected to interact with MotA, to help form the stator portion of the flagellular motor (Kojima and Blair, 2004). STRING analysis showed that LM5578_0765 likely physically interacts with MotA. In addition, LM5578_0765 functionally interacts, and is likely co-expressed, with additional flagellar proteins including those making up the structure of the flagellum (for example, flagellum hook protein FlgE) and aiding in flagellar movement (for example, flagellar motor switch protein FliG), as well as a two-component sensor histidine kinase, CheA, which helps control the direction of flagellar rotation via the relay of chemotactic signaling (Fig. 3A). Phylogenetic placement also favors LM5578_0765 as a MotB homolog in *L. monocytogenes* 08-5578 (Fig. 3B).

As mentioned, LM5578_0802 is likely to encode for a methyl-accepting chemotaxis protein according to homology-based annotation; including the presence of a conserved glutamate residue in the C-terminal region of the protein, which becomes methylated when a given MCP is bound to attractant molecules (Table 3). *In silico*-based localization experiments also support the notion of LM5578_0802 representing an MCP. In brief, methyl-accepting chemotaxis proteins (MCPs) are membrane-spanning proteins that are used by bacteria to detect and respond to extracellular molecules by either moving toward attractants or

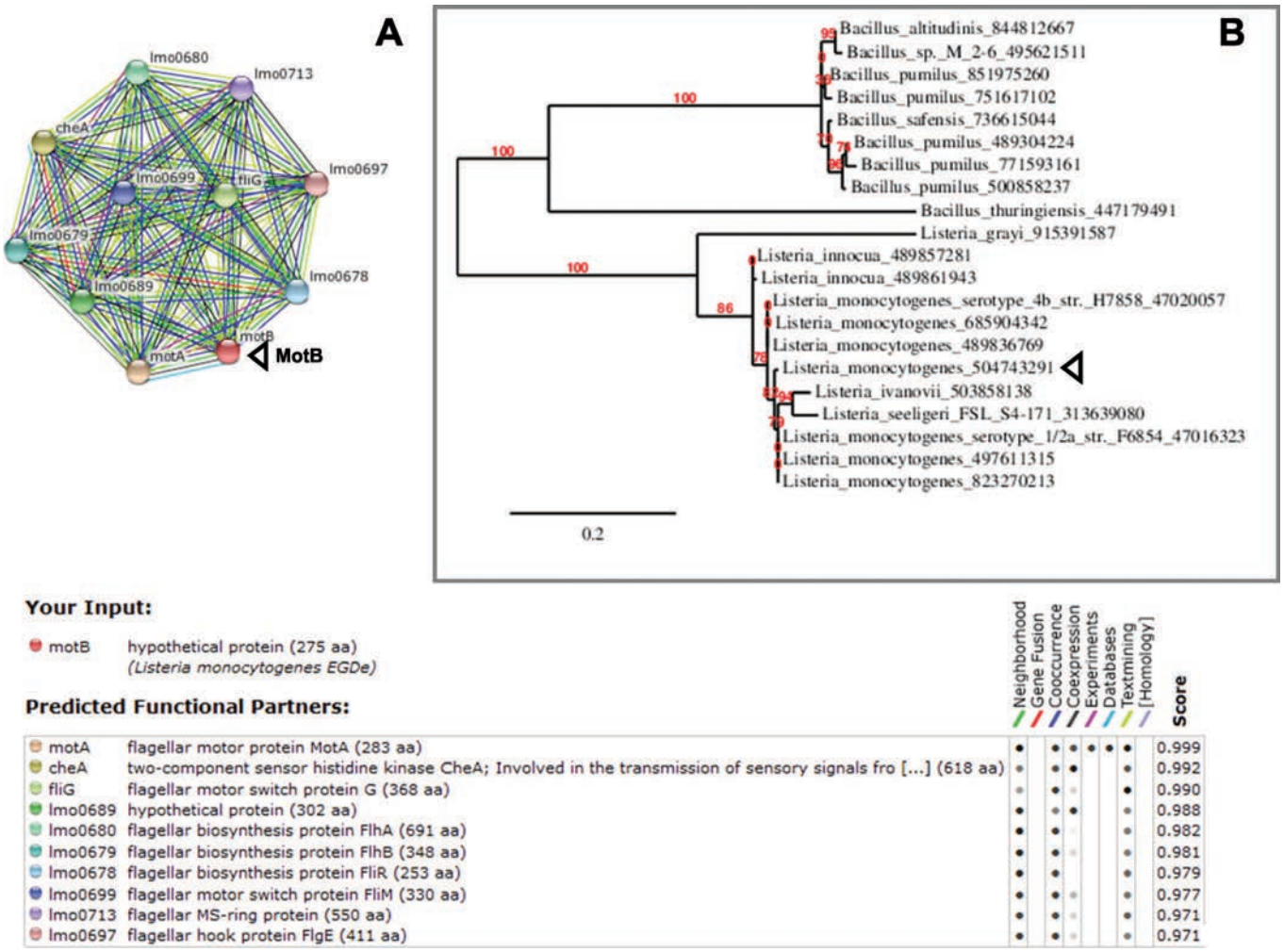


Fig. 3. STRING interactions and phylogeny of flagellar-associated protein MotB. A. STRING map and table of interactions. The category of interaction is shown in the table. Color saturation represents confidence score of association. The position of MotB is indicated by an arrowhead. Scores higher than 0.7 are defined as high confidence. Black and gray dots indicate the individual types of interactions (see Methods for more information about interaction types) predicted to occur between the query protein and the given functional partner. Black dots contribute to the overall confidence score to a greater degree than gray. B. Phylogeny of MotB (arrowhead) and related proteins. Scale represents evolutionary distance. Values are branch support. Sequences shown are predicted MotB proteins.

away from repellants (Wuichet et al., 2007). MCPs differ from one another by ligand specificity, which can include peptides, sugars, and metals. Although MCPs have not been directly linked to virulence of *L. monocytogenes*, the process of chemotaxis has been shown to contribute to the disease-causing capabilities of *L. monocytogenes* (Dons et al., 2004). However, in the related gastrointestinal pathogen *Campylobacter jejuni*, mutants lacking select MCPs showed a decreased ability to infect intestinal cell lines *in vitro* and to colonize chickens *in vivo* (Korolik, 2010).

As summarized in Table 4, LM5578_0802 is predicted to localize within the cytoplasmic membrane, and contains two membrane-spanning domains. The predicted membrane topology of LM5578_0802 is consistent with MCPs. Specifically, LM5578_0802 contains two alpha-helices linked together by an extracellular region (periplasmic-oriented, ligand-binding region), and followed by a cytoplasmic domain that is believed to interact with downstream chemotactic signaling proteins (Baker et al., 2006). STRING data (Fig. 4A) showed that LM5578_0802 is likely to physically interact with many proteins known to function in chemotaxis, including CheY (signal relay), CheA (transmission of signals to flagellum, also is co-expressed with LM5578_0802), and CheR (methylation of MCPs). Consistent with the annotation, phylogeny places LM5578_0802 among MCPs from related strains of *L. monocytogenes* (Fig. 4B).

Exoribonuclease R

Exoribonuclease R, also known as RNase R, has been shown to influence virulence in a number of bacteria, including *Aeromonas hydrophila*, *Shigella flexneri*, and *Salmonella enterica* (Cheng et al.,

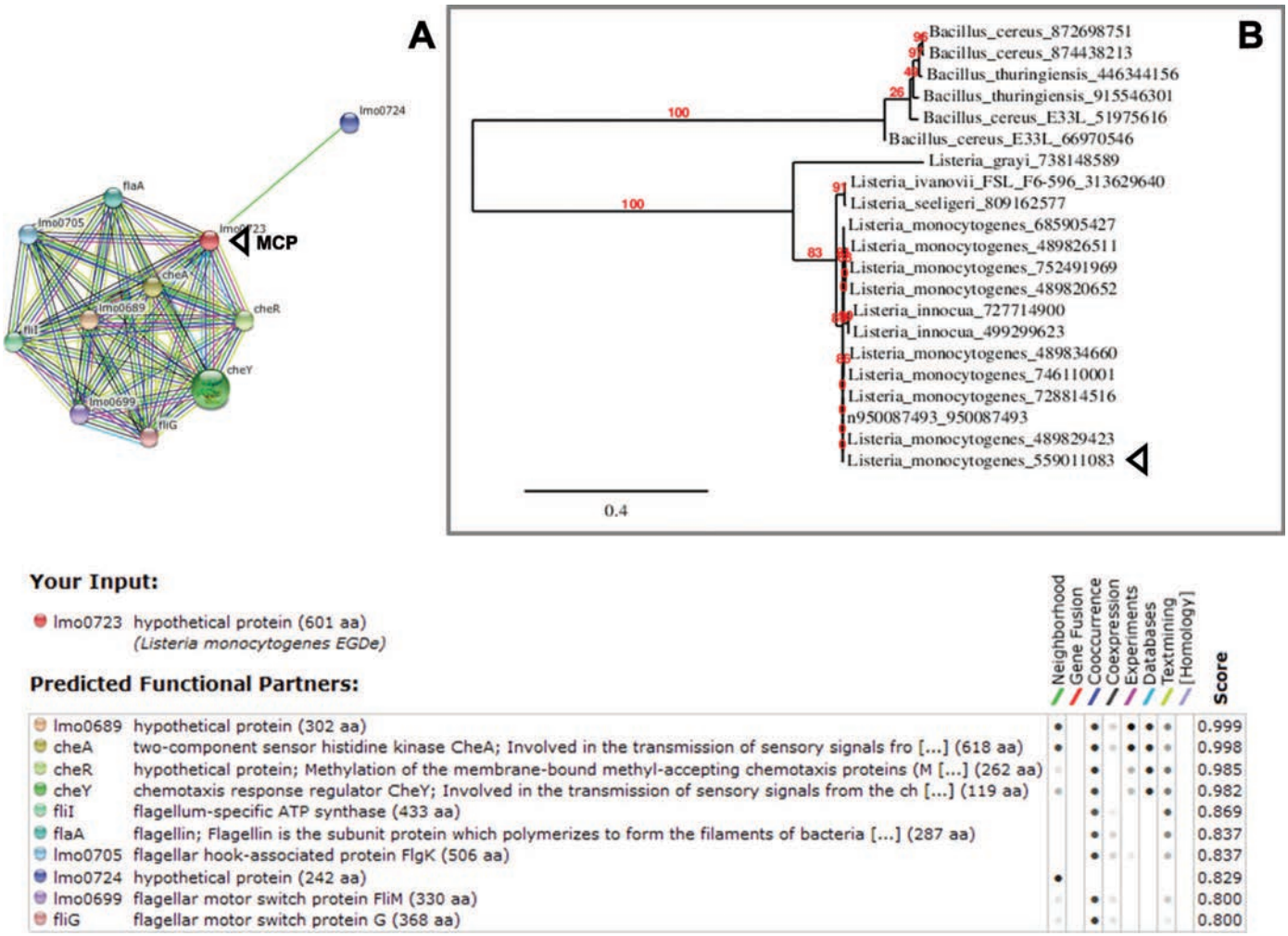
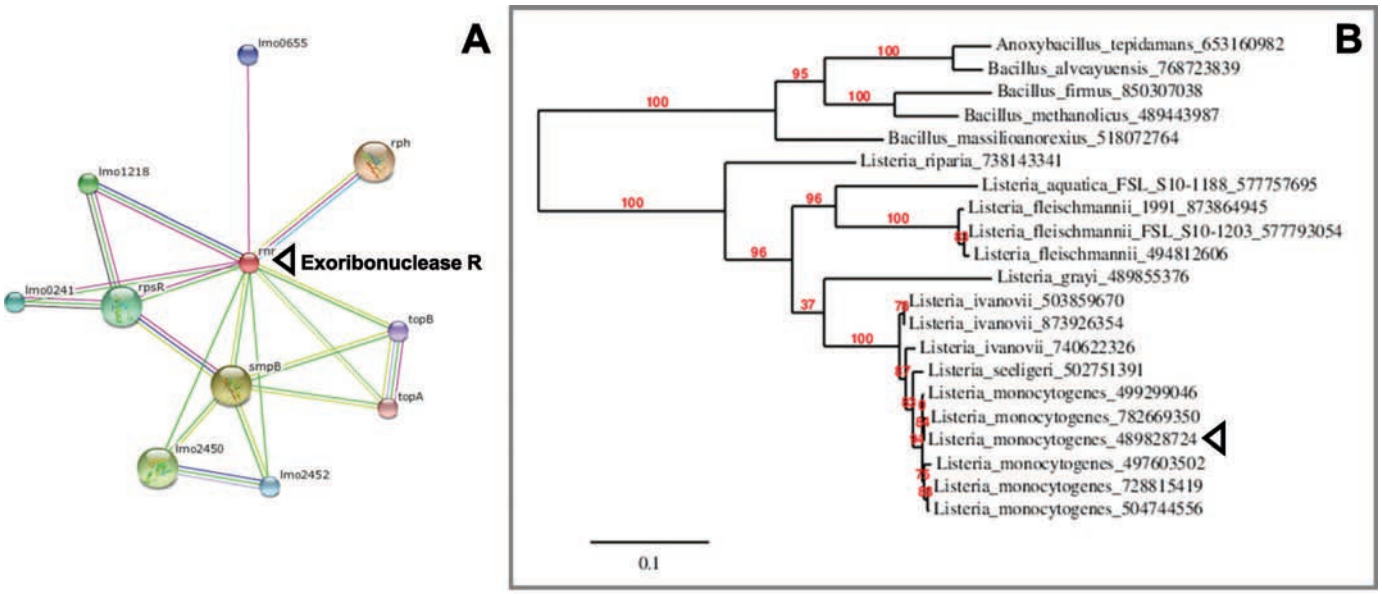


Fig. 4. STRING interactions and phylogeny of methyl-accepting chemotaxis protein (MCP). A. STRING map and table of interactions. The category of interaction is shown in the table. Color saturation represents confidence score of association. The position of MCP is indicated by an arrowhead. Scores higher than 0.7 are defined as high confidence. Black and gray dots indicate the individual types of interactions (see Methods for more information about interaction types) predicted to occur between the query protein and the given functional partner. Black dots contribute to the overall confidence score to a greater degree than gray. B. Maximum likelihood placement of MCP (arrowhead) and related proteins. Scale represents evolutionary distance and values are branch support. Sequences shown are predicted MCP's.

1998; Erova et al., 2008). In addition, exoribonuclease R appears to be a cold-shock protein that is needed for effective growth at low temperatures (Matos et al., 2014). Given observations that listeriosis can be transmitted to human hosts via refrigerated foods, exoribonuclease R may have significant potential as a virulence factor of *L. monocytogenes*. Biochemically, exoribonuclease R acts to regulate the degradation of mRNA and quality-control of rRNA, which are significant determinants of prokaryotic gene expression (Matos et al., 2014). LM5578_2645, designated as a HP, appears by shared sequence homology to represent an exoribonuclease R (Table 2), including the conservation of a residue (arginine, R^{572}) needed (Haddad et al., 2014) for protein functionality in known exoribonuclease R enzymes (Table 3). As anticipated, LM5578_2645 is predicted to be located in the cytoplasm and lacks identifiable signal peptides according to localization experiments (Table 4). LM5578_2645 also is expected to physically and functionally interact with additional proteins that contribute to RNA degradation; for example RNase PH (Matos et al., 2014), which has been recently shown to degrade cellular RNAs that contain significant secondary structure (Fig. 5A). The translated sequence of LM5578_2645 clusters, phylogenetically, with exoribonuclease R sequences from related strains of *L. monocytogenes*, which also supports previous data that LM5578_2645 likely represents an exoribonuclease R (Fig. 5B).

CONCLUSIONS

The findings here suggest that the annotation of hypothetical proteins of bacterial genomes can be improved by use of manual annotation strategies. Significantly, 5 HP's were found to encode genes that may impact the pathogenic nature of *L. monocytogenes*



Your Input:

		Neighborhood	Gene Fusion	Cooccurrence	Coexpression	Experiments	Databases	Textmining	[Homology]	Score
rnr	hypothetical protein; 3'-5' exoribonuclease that releases 5'-nucleoside monophosphates and is involved in maturation of structured RNAs (By similarity) (793 aa) *(Listeria monocytogenes EGDe)*									
Predicted Functional Partners:										
rph	ribonuclease PH; Phosphorolytic exoribonuclease that removes nucleotide residues following the [...] (248 aa)					•	•			0.981
smpB	SsrA-binding protein; Binds specifically to the SsrA RNA (tmRNA) and is required for stable ass [...] (154 aa)	•							•	0.957
lmo2450	hypothetical protein (248 aa)	•								0.918
lmo1218	hypothetical protein (252 aa)	•				•				0.827
rpsR	30S ribosomal protein S18; Binds as a heterodimer with protein S6 to the central domain of the [...] (79 aa)	•				•				0.807
lmo0241	hypothetical protein (251 aa)	•				•				0.796
lmo2452	hypothetical protein (247 aa)	•								0.782
lmo0655	hypothetical protein (235 aa)					•				0.741
topB	DNA topoisomerase III (717 aa)	•								0.720
topA	DNA topoisomerase I; Releases the supercoiling and torsional tension of DNA, which is introduce [...] (692 aa)	•								0.720

Fig. 5. Phylogeny and STRING analysis exoribonuclease R. A. STRING map and table of interactions. The category of interaction is shown in the table. Color saturation represents interaction confidence. The position of exoribonuclease R is indicated by an arrowhead. Scores higher than 0.7 are defined as high confidence. Black and gray dots indicate the individual types of interactions (see Methods for more information about interaction types) predicted to occur between the query protein and the given functional partner. Black dots contribute to the overall confidence score to a greater degree than gray. B. Maximum likelihood position of exoribonuclease R (arrowhead) and related proteins. Scale is evolutionary distance. Values are branch support. Sequences shown are exoribonucleases.

08-5578. Future studies will seek to further characterize the 5 genes potentially related to virulence via *in vitro* methodologies. The strategy used in this study makes use of a robust variety of readily available and no-cost bioinformatics tools that predict the functionality of protein-encoding genes using metrics known to influence functionality, including homology, structural features, phylogeny, cellular localization, and protein-protein interaction (genome context) data. Importantly, this strategy can likely be used to better understand the functions of hypothetical proteins in additional bacterial genomes.

ACKNOWLEDGEMENTS

Michael J. LaGier was supported in part by inclusion in the Microbial Genome Annotation Network, sponsored by the U.S. National Science Foundation (www.mgan-network.org/). We thank all members of the Biology Department at Grand View University for encouragement and project support, and Mr. Keith Daniels of Grand View University for computer support.

REFERENCES

ALTSCHUL SF, GISH W, MILLER W, MYERS EW, LIPMAN DJ. 1990. Basic local alignment search tool. J Mol Biol. 215(3):403–410.

BAKER MD, WOLANIN PM, STOCK JB. 2006. Systems biology of bacterial chemotaxis. Curr Opin Microbiol. 9:187–192.

BIERNE, H, MAZMANIAN SK, TROST M, PUCCIARELLI MG, LIU G, DEHOUX P, JÄNSCH L, GARCIA-DEL PORTILLO F, SCHNEEWIND O, COSSART P, et al. 2002. Inactivation of the srtA gene in *Listeria monocytogenes* inhibits anchoring of surface proteins and affects virulence. Mol Microbiol. 43(4):869–881.

BORK P. 2000. Powers and pitfalls in sequence analysis: the 70% hurdle. Genome Res. 10(4):398–400.

BRADSHAW WJ, DAVIES AH, CHAMBERS CJ, ROBERTS AK, SHONE CC, ACHARYA KR. 2015. Molecular features of the sortase enzyme family. FEBS J. 282(11):2097–2114.

CHENG ZF, ZUO Y, LI Z, RUDD KE, DEUTSCHER MP. 1998. The vacB gene required for virulence in *Shigella flexneri* and *Escherichia coli* encodes the exoribonuclease RNase R. J Biol Chem. 273(23):14077–14080.

DEREEPER A, GUIGNON V, BLANC G, AUDIC S, BUFFET S, CHEVENET F, DUFAYARD JF, GUINDON S, LEFORT V, LESCOT M, et al. 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. 36:W465–W469.

DOERKS T, VAN NOORT V, MINGUEZ P, BORK P. 2012. Annotation of the *M. tuberculosis* hypothetical orfeome: adding functional information to more than half of the uncharacterized proteins. PloS One. 7(4):e34302.

DONS L, ERIKSSON E, JIN Y, ROTTENBERG ME, KRISTENSSON K, LARSEN CN, BRESCIANI J, OLSEN JE. 2004. Role of flagellin and the two-component CheA/CheY system of *Listeria monocytogenes* in host cell invasion and virulence. Infect Immun. 72(6): 3237–3244.

EROVA TE, KOSYKH VG, FADL AA, SHA J, HORNEMAN AJ, CHOPRA AK. 2008. Cold shock exoribonuclease R (VacB) is involved in *Aeromonas hydrophila* pathogenesis. J Bacteriol. 190:3467–3474.

FERREIRA V, WIEDMANN M, TEIXEIRA P, STASIEWICZ MJ. 2014. *Listeria monocytogenes* persistence in food-associated environments: epidemiology, strain characteristics, and implications for public health. J Food Prot. 77(1):150–170.

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.

HADDAD N, MATOS RG, PINTO T, RANNOU P, CAPPELIER J-M, PRÉVOST H, ARRAIANO CM. 2014. The RNase R from *Campylobacter jejuni* has unique features and is involved in the first steps of infection. J Biol Chem. 289(40): 27814–27824.

HERNANDEZ-MILIAN A, PAYERAS-CIFRE A. 2014. What is new in listeriosis? BioMed Res Int. 2014:358051.

IJAQ J, CHANDRASEKHARAN M, PODDAR R, BETHI N, SUNDARARAJAN VS. 2015. Annotation and curation of uncharacterized proteins- challenges. Front Genet. 6:119.

KÄLL L, KROGH A, SONNHAMMER EL. 2004. A combined transmembrane topology and signal peptide prediction method. J Mol Biol. 338(5):1027–1036.

KANEHISA M, GOTO S. 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28(1):27–30.

KOJIMA S, BLAIR DF. 2004. The bacterial flagellar motor: structure and function of a complex molecular machine. Int Rev Cytol. 233:93–134.

KOROLIK V. 2010. Aspartate chemosensory receptor signalling in *Campylobacter jejuni*. Virulence. 1(5):414–417.

KROGH A, LARSSON B, VON HEIJNE G, SONNHAMMER EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 305(3):567–580.

KUMAR K, PRAKASH A, TASLEEM M, ISLAM A, AHMAD F, HASSAN MI. 2014. Functional annotation of putative hypothetical proteins from *Candida dubliniensis*. Gene. 543(1):93–100.

LOMONACO S, NUCERA D, FILIPELLO V. 2015. The evolution and epidemiology of *Listeria monocytogenes* in Europe and the United States. Infect Genet Evol. 35:172–183.

MAGIS C, TALY JF, BUSSOTTI G, CHANG JM, DI TOMMASO P, ERB I, ESPINOSA-CARRASCO J, NOTREDAME C. 2014. T-Coffee: Tree-based consistency objective function for alignment evaluation. Methods Mol Biol. 1079:117–129.

MATOS RG, BÁRRIA C, MOREIRA RN, BARAHONA S, DOMINGUES S, ARRAIANO CM. 2014. The importance of proteins of the RNase II/RNB-family in pathogenic bacteria. Front Cell Infect Microbiol. 4:68.

MILOHANIC E, PRON B, BERCHE P, GAILLARD JL. 2000. Identification of new loci involved in adhesion of *Listeria monocytogenes* to eukaryotic cells. European Listeria Genome Consortium. Microbiol. 146(Pt 3):731–739.

O'NEIL HS, MARQUIS H. 2006. *Listeria monocytogenes* flagella are used for motility, not as adhesins, to increase host cell invasion. Infect Immun. 74(12):6675–6681.

PETERSEN TN, BRUNAK S, VON HEIJNE G, NIELSEN H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 8(10):785–786.

POPTSOVA MS, GOGARTEN JP. 2010. Using comparative genome analysis to identify problems in annotated microbial genomes. Microbiol. 156:1909–1917.

PROMPONAS VJ, ILIOPOULOS I, OUZOUNIS CA. 2015. Annotation inconsistencies beyond sequence similarity-based function prediction - phylogeny and genome structure. Stand Genomic Sci. 10:108.

RADKOV AD, MOE LA. 2014. Bacterial synthesis of D-amino acids. Appl Microbiol Biotechnol. 98(12):5363–5374.

REDDY S, LAWRENCE ML. 2014. Virulence characterization of *Listeria monocytogenes*. Methods Mol Biol. 1157:157–165.

SCHNOES AM, BROWN SD, DODEVSKI I, BABBITT PC. 2009. Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. PLoS Comput Biol. 5(12):e1000605.

SILBER AM, PEREIRA CA. 2012. Assignment of putative functions to membrane "hypothetical proteins" from the *Trypanosoma cruzi* genome. J Membr Biol. 245(3):125–129.

STRYCH U, HUANG HC, KRAUSE KL, BENEDIK MJ. 2000. Characterization of the alanine racemases from *Pseudomonas aeruginosa* PAO1. Curr Microbiol. 41(4):290–294.

SZKLARCZYK D, FRANCESCHINI A, WYDER S, FORSLUND K, HELLER D, HUERTA-CEPAS J, SIMONOVIC M, ROTH A, SANTOS A, TSAFOU KP. et al. 2015. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 43:D447–D452.

THOMPSON RJ, BOUWER HG, PORTNOY DA, FRANKEL FR. 1998. Pathogenicity and immunogenicity of a *Listeria monocytogenes* strain that requires D-alanine for growth. Infect Immun. 66(8):3552–3561.

WATANABE A, KUROKAWA Y, YOSHIMURA T, KURIHARA T, SODA K, ESAKI N. 1999. Role of lysine 39 of alanine racemase from *Bacillus stearothermophilus* that binds pyridoxal 5'-phosphate. Chemical rescue studies of Lys39 –> Ala mutant. J Biol Chem. 274(7):4189–4194.

WUICHET K, ALEXANDER RP, ZHULIN IB. 2007. Comparative genomic and protein sequence analyses of a complex system controlling bacterial chemotaxis. Methods Enzymol. 422: 1–31.