Using Genome-Scale Metabolic Models to Compare Serovars of the Foodborne Pathogen *Listeria monocytogenes*

A THESIS
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Zachary P. Metz

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

Advisor: Dr. David Baumler

August 2016

Acknowledgements

I would first like to thank my advisor, Dr. David Baumler, for believing in my potential as a graduate student and giving me the opportunity to pursue this degree. Additionally, his assistance, guidance, and support were essential for the completion of this project.

I would also like to express my gratitude to the distinguished members of my committee, Dr. Tonya Schoenfuss and Dr. Ted Labuza. I greatly appreciate their insights and advice were invaluable while completing my research and writing my thesis.

Next, I acknowledge the MnDRIVE Global Food Ventures Program for awarding me a graduate student fellowship, providing me with much appreciated funding, and allowing me to gain experience outside of the laboratory and professional development that I almost certainly would not have had access to otherwise.

I also acknowledge and appreciate the Department of Food Science and Nutrition and the College of Food, Agricultural, and Natural Resource Sciences for giving me admittance to this program and for providing me with the funding not provided by anyone else.

A special thanks to all the members of the Baumler Lab. The graduate students became a close-knit team of researchers that was able to assist and support each other through all of the rigors associated with graduate school. Additionally, the undergraduate students provided much needed assistance that considerably lessened the research workload.

Finally, I want to thank my friends, my sisters, Kelsey and Jessie, my dad, Randy, and my mom, Mary. Without their love and support, none of this would have been possible.

Abstract

Listeria monocytogenes is a microorganism of great concern for the food industry, most notably because it is the 2nd most deadly bacterial foodborne pathogen. Therefore, it is important to study the organism in order to identify novel methods of control. Systems biology is one such approach. Using a combination of computational techniques and laboratory methods, genome-scale metabolic models (GEMs) can be created, validated, and used to simulate growth environments and discern metabolic capabilities of microbes of interest, including *L. monocytogenes.* The objective of the work presented here was to generate GEMs for six different strains of *L. monocytogenes*, and to both qualitatively and quantitatively validate these GEMs with experimental data.

Qualitative validation by comparison to phenotypic microarray data resulted in GEMs with nutrient utilization agreement similar to that of previously published GEMs. Additionally, aerobic batch growth experiments resulted in predictions for growth rate and growth yield that were strongly and significantly correlated with experimental values. These findings are significant because they show that these GEMs for *L. monocytogenes* are comparable in agreement between *in silico* predictions and *in vitro* results to published models of other organisms. Therefore, as with the other models, namely those for *Escherichia coli*, *Staphylococcus aureus*, *Vibrio vulnificus*, and *Salmonella* spp., they can be used to determine new methods of growth control and disease treatment. Additionally, the findings confirm the acceptability of using semi-automated tools, like those provided by KBase, to generate GEMs.

Table of Contents

List of Tables vi

List of Figures viii

General Introduction 1

1 Literature Review 3

1.1 History, Isolation, and Identification 3

1.2 Characteristics 4

1.3 Lineages and Serotypes 5

1.4 Prevalence in Food 7

1.5 Listeriosis 10

1.6 *L. innocua* as a Biological Indicator 10

1.7 Genomics 11

1.8 Systems Biology of Foodborne Pathogens 15

1.9 Biochemical Pathways 16

2 Generation of Genome-Scale Metabolic Models and Qualitative Validation 18

2.1 Abstract 18

2.2 Introduction 18

2.3 Materials and Methods 20

2.3.1 Comparative Genome Alignment 20

2.3.2 Stock Culture Maintenance 21

2.3.3 Generation of Genome-Scale Metabolic Models 21

2.3.4 *In Vitro* Nutrient Utilization 22

2.3.5 *In Silico* Nutrient Utilization Prediction and Reconciliation 23

2.3.6 Essential Reaction Predictions 25

2.4 Results 27
2.4.1 Genomic Comparison 27
2.4.2 Reconstruction of Genome-Scale Metabolic Models 29
2.4.3 Nutrient Phenotype Data 30
2.4.4 Comparison Between *In Silico* Predictions and Experimental Nutrient Utilization Results 41
2.4.5 Agreement Comparison to Other Genome-Scale Metabolic Models 47
2.4.6 Essential Reactions 51
2.5 Discussion 56
2.5.1 Genomic Differences 56
2.5.2 Genome-Scale Metabolic Model Properties 58
2.5.3 Nutrient Utilization 58
2.5.4 Nutrient Utilization Agreement Summary and Comparison 59
2.5.5 Essential Reactions 62
2.6 Conclusion 63
3 Quantitative Analysis of Genome-Scale Metabolic Models for Six *L. monocytogenes* Strains 65
3.1 Abstract 65
3.2 Introduction 65
3.3 Materials and Methods 66
3.3.1 Growth Assays 66
3.3.2 Batch Growth Experiments 67
3.3.3 Dry Cell Weight Measurements 69
3.3.4 Viable Cell Counts 69

3.3.5 *In Silico* Quantitative Adjustment 69
3.3.6 Phylogenetic Analysis 70
3.4 Results 70
3.4.1 Growth Curves 70
3.4.2 Conversion Factors 73
3.4.3 Batch Growth and Qualitative *In Silico* Refinement 73
3.4.4 *In Silico* Growth Predictions in Food Matrices at Various Temperatures 76
3.5 Discussion 81
3.5.1 Growth Assays 81
3.5.2 Conversion Factors 82
3.5.3 Batch Growth and Growth Rate and Yield Comparison 82
3.5.4 Utilization of Previous Data 85
3.6 Conclusion 86
4 Summary 87
References 89
Appendix A: PLOS One Article 94

List of Tables

Table 1: Completed genomes of *Listeria* spp. available at NCBI[a] 13
Table 2: The six strains of *L. monocytogenes* chosen for this study..... 20
Table 3: Nutrients present in the essential reaction simulation reflecting minimal media26
Table 4: Nutrients added to essential reaction simulations to reflect high-risk foods 26
Table 5: Numerical genome comparison of the six chosen strains of *L. monocytogenes* 29
Table 6: Gene, metabolite, and reactions contained in each version of the GEMs 30
Table 7: Enzymes catalyzing strain and serovar-specific metabolic reactions..... 30
Table 8: Summary of experimental carbon utilization data..... 31
Table 9: Summary of experimental nitrogen utilization data 34
Table 10: Summary of experimental phosphorous utilization data 36
Table 11: Summary of experimental sulfur utilization data 39
Table 12: Number of strain and serovar specific nutrients..... 40
Table 13: Unique carbon sources metabolized by a single *L. monocytogenes* strain or serovar..... 40
Table 14: Unique nitrogen sources metabolized by a single *L. monocytogenes* strain or serovar..... 41
Table 15: Unique sulfur/phosphorus sources metabolized by a single *L. monocytogenes* strain or serovar..... 41
Table 16: Carbon sources with agreement between *in silico* predictions and experimental results for all strains..... 42
Table 17: Numerical agreement summary for carbon source utilization, false negative and false positive values correspond to final models 46
Table 18: Numerical agreement summary for nitrogen source utilization 47
Table 19: Numerical agreement summary for phosphorus source utilization 47
Table 20: Numerical agreement summary for sulfur source utilization 47
Table 21: Essential reaction summary for the simulation of conditions representing glucose minimal media 52
Table 22: Essential reaction summary for the simulation of conditions representing queso fresco..... 52

Table 23: Essential reaction summary for the simulation of conditions representing chicken breast 52
Table 24: Essential reaction summary for the simulation of conditions representing smoked salmon 53
Table 25: Essential reaction summary for the simulation of conditions representing cantaloupe 53
Table 26: Essential reactions summary for the simulation of conditions representing romaine lettuce 53
Table 27: Enzymes catalyzing strain and serovar-specific essential reactions in simulations reflecting glucose minimal media 54
Table 28: Enzymes catalyzing strain and serovar-specific essential reactions in simulations reflecting queso fresco 54
Table 29: Enzymes catalyzing strain and serovar-specific essential reactions in simulations reflecting chicken breast 55
Table 30: Enzymes catalyzing strain and serovar-specific essential reactions in simulations reflecting smoked salmon 55
Table 31: Enzymes catalyzing strain and serovar-specific essential reactions in simulations reflecting cantaloupe 56
Table 32: Enzymes catalyzing strain and serovar-specific essential reactions in simulations reflecting romaine lettuce 56
Table 33: Experimentally determined conversion factors 73
Table 34: Comparison of experimental and *in silico* growth rates (h^{-1}) 74
Table 35: Comparison of experimental and *in silico* growth yields (g/g glucose) 75
Table 36: Scalar factors between experimental and *in silico* growth rates for beef frankfurters compared to scalars already introduced 77
Table 37: Scalar factors between experimental and *in silico* growth rates for sliced turkey breast compared to scalars already introduced 77

List of Figures

Figure 1: Number of food recalls associated with *L. monocytogenes* since 2010 8
Figure 2: Biolog™ Phenotypic Microarray experimental flow diagram 23
Figure 3: Mauve genome alignment for the six chosen strains of *L. monocytogenes* 28
Figure 4: Mauve backbone alignment for the six chosen strains of *L. monocytogenes*.... 28
Figure 5: Comparison of *in silico* predictions to experimental results for 39 individual carbon sources with at least one disagreement 43
Figure 6: Comparison of *in silico* predictions to experimental results for 62 individual nitrogen sources 44
Figure 7: Comparison of *in silico* predictions to experimental results for 22 phosphorus sources..... 45
Figure 8: Comparison of *in silico* predictions to experimental results for 11 sulfur sources 46
Figure 9: Comparison of carbon source utilization agreement between genome-scale metabolic models created in this study and previous genome-scale metabolic models ... 48
Figure 10: Comparison of nitrogen source utilization agreement between genome-scale metabolic models created in this study and previous genome-scale metabolic models ... 49
Figure 11: Comparison of phosphorus source utilization agreement between genome-scale metabolic models created in this study and previous genome-scale metabolic models 50
Figure 12: Comparison of sulfur source utilization agreement between genome-scale metabolic models created in this study and previous genome-scale metabolic models ... 51
Figure 13: Diagram of the batch sparger apparatus used in this study 68
Figure 14: Microwell growth assay of six *L. monocytogenes* strains in IMM at 37°C including standard deviations 71
Figure 15: Microwell growth assay of six *L. monocytogenes* strains in MWB at 37°C including standard deviations 72
Figure 16: Microwell growth assay of six *L. monocytogenes* strains in MWB supplemented with 3% BHI at 37°C including standard deviations 72

Figure 17: Comparison of experimental and *in silico* growth rates and growth yields including experimental standard deviations..... 75
Figure 18: Maximum likelihood phylogeny of *L. monocytogenes* strains used in this study and their associated growth rates and growth yields 76
Figure 19: Comparison of *in silico* predicted growth rate and experimental growth data for four strains of *L. monocytogenes* in beef frank slurries at 4°C including experimental standard deviations..... 78
Figure 20: Comparison of *in silico* predicted growth rate and experimental growth data for four strains of *L. monocytogenes* in beef frank slurries at 8°C including experimental standard deviations..... 78
Figure 21: Comparison of *in silico* predicted growth rate and experimental growth data for four strains of *L. monocytogenes* in beef frank slurries at 12°C including experimental standard deviations..... 79
Figure 22: Comparison of *in silico* predicted growth rate and experimental growth data for four strains of *L. monocytogenes* in sliced turkey breast slurries at 4°C including experimental standard deviations..... 80
Figure 23: Comparison of *in silico* predicted growth rate and experimental growth data for four strains of *L. monocytogenes* in sliced turkey breast slurries at 8°C including experimental standard deviations..... 80
Figure 24: Comparison of *in silico* predicted growth rate and experimental growth data for four strains of *L. monocytogenes* in sliced turkey breast slurries at 12°C including experimental standard deviations..... 81

General Introduction

Listeria monocytogenes is a foodborne pathogen of great concern for the food industry. This concern is due to several important characteristics of the organism. First, it has an unusually high mortality rate, with some estimates as high as 30%, leading it to become the second leading cause of death among bacterial foodborne pathogens in the United States [1]. Secondly, as a ubiquitous organism, *L. monocytogenes* can be found in virtually any environment and on virtually any food product. Finally, and perhaps most importantly, the organism's ability to grow, not merely survive, at refrigeration temperatures.

Due to the importance of *L. monocytogenes* to the food industry, there is a continuous interest in new methods of control and treatment. The development of rapid, relatively cheap genome sequencing techniques in recent years has led to the emergence of genomic tools for identifying more properties of microorganisms through the field of systems biology, and its associated computational techniques.

Genome-scale metabolic models (GEMs) are one of the newer techniques by which foodborne pathogens are being studied. These GEMs take the genetic information contained in the entire genome and convert it to a metabolic network that consists of metabolic reactions and their associated metabolites. This network is then converted to a system of algebraic equations. Using computing software, this system of equations can be used to calculate the flow of metabolites through the metabolic network and predict the growth of the organism under specific conditions. The model can then be adjusted by comparing the predictions to experimental data. A working model can then be used to study the metabolism of the organism and identify metabolic reactions that are essential for the growth and survival of the organism. These essential reactions provide ideal targets for new methods of treatment and control. This type of approach, using GEMs to identify new targets for control of foodborne pathogens has already been done for seven foodborne pathogens, including *Escherichia coli* O157:H7, *Salmonella*, *Vibrio vulnificus*, and *Staphylococcus aureus*. The work presented here describes the creation and validation of GEMs for six different strains of *Listeria monocytogenes*.

Most studies concerning *L. monocytogenes* utilizing comparative genomics are focused on conservation of virulence factor genes, which is not altogether unexpected when the organism's extreme pathogenicity is considered. However, there is a surprising lack of studies focused on examining the metabolic capabilities and associated genes to compare strains of this organism. The work presented here seeks to become the first in-depth genome-scale study to do this by creating six strain-specific GEMs, the first of their kind for *L. monocytogenes*.

1 Literature Review

1.1 History, Isolation, and Identification

In 1924, a bacterium that had not been previously described was isolated following an outbreak of mononucleosis in laboratory rabbits and guinea pigs [2]. This bacterium was given a succession of generic names, starting with "*Bacterium monocytogenes,*" until, in 1927, the name *Listeria* was chosen in honor of the discoverer of antisepsis—Dr. Lister [3]. Evidence exists that *Listeria* had been observed, and even cultivated and described, prior to its discovery in 1924. However, due to the lack of preserved cultures, it is impossible to definitively link this previous work to *Listeria* [3]. The oldest strain of *Listeria*; however, is not the culture isolated in 1924. It is a clinical isolate deposited at the Pasteur Institute from a case of human meningitis in 1921 [3].

Listeria is ubiquitous in nature, and can be found in a wide variety of areas. It has been isolated from soil, plants, water, animals, and can even be carried asymptomatically in humans [4]. It can also colonize virtually any surface, and is a frequent problem in the food industry. Historically, the primary method of isolating strains of *Listeria* is by plating on tryptic soy base agar. In samples from humans, blood needs to be added to the growth agar as well. Typically, for isolation of *Listeria* from food products, plating on selective media is preceded by an enrichment step, which is usually carried out in selective liquid media. If rapid detection of *Listeria* is desired, the preferred method is polymerase chain reaction (PCR); however, immunoassay-based kits can also be used [4].

The genus *Listeria* is composed of seventeen species: *L. monocytogenes, L. innocua, L. ivanovii, L. seeligeri, L. welshimeri, L. grayi, L. rocourtiae, L. marthii, L. weihenstephanensis, L. fleischmannii, L. aquatica, L. riparia, L. grandensis, L. cornellensis, L. floridensis, L. newyorkensis,* and *L. booriae* [5]. *L. ivanovii* is further categorized into two subspecies: *L. ivanovii* subsp. *ivanovii* and *L. ivanovii* subsp. *londoniensis* [4]. The genus is identified using biochemical reactions, Gram staining, motility observation, and other common techniques. Identification at the species level is done by testing biochemical properties including hemolysis, the Christie–Atkins–Munch-

Petersen (CAMP) reaction, acid production on different carbon sources, hippurate hydrolysis, and reduction of nitrate [4].

1.2 Characteristics

Bacteria of the genus *Listeria* are unencapsulated, Gram-positive, and do not form spores [4, 6]. *Listeria* DNA contains low (36-42%) amounts of guanine and cytosine, and *Listeria* is related to *Enterococcus*, *Staphylococcus*, *Streptococcus*, *Bacillus*, and *Clostridium* [4, 6]. Listeriae are facultative anaerobes that grow well *in vitro*, particularly on brain heart infusion (BHI) media [4]. Nutrients essential for *Listeria* growth include select carbohydrates; the vitamins riboflavin, thiamine, thioctic acid, and biotin; and the amino acids glutamine, valine, cysteine, arginine, methionine, leucine, and isoleucine [7, 8]. Other nutrients—such as iron, histidine, and tryptophan—while not essential, have been show to enhance growth of *Listeria* [7, 8]. Additionally, some nutrients—including activated charcoal and cellobiose—have been shown to impact the expression of virulence factors [4].

The motility of *Listeria* is dependent on temperature. At lower temperatures (< 30°C) *Listeria*, like many other bacteria, use their flagella for motility [4, 6]. However, at higher temperatures (> 30°C) *Listeria* flagellin is not expressed. At these higher temperatures, if inside of a host cell, *Listeria* relies on actin-based motility [4]. This method of motility is executed by constant polymerization and depolymerization of actin filaments found within host cells. The *Listeria* cell creates a comet-like tail out of these actin filaments, and uses the forces generated by the polymerization of the actin filaments to propel itself around the host cell's cytoplasm [4].

Listeria is a highly resistant genus of bacteria and can survive a wide range of adverse conditions. The optimal temperature for growth is 30-37°C, but growth can occur from 1-45°C, which is an unusually broad range [4, 6]. Additionally, while neutral pH and 0.5% NaCl are optimal for growth, *Listeria* exhibits significant resistance to both salinity and pH change, and are able to survive from pH 4.5-9 and up to 10% NaCl [4, 6].

1.3 Lineages and Serotypes

Listeria monocytogenes isolates can be grouped into four lineages based upon their ecological, genetic, and phenotypic characteristics [9]. The designated lineage is typically chosen using several phenotypic and genotypic techniques, such as pulse field gel electrophoresis (PFGE), multilocus sequence typing (MLST), and ribotyping. The majority of *L. monocytogenes* isolates are placed into lineages I and II, while isolates from lineages III and IV are extremely rare [9]. These lineages are further characterized into 13 serotypes, with approximately 95% of the strains isolated from humans coming from three major serotypes—1/2a, 1/2b, and 4b [4]. Serotypes 1/2b and 4b fall into lineage I, while serotype 1/2a belongs to lineage II.

Ecologically, there are several different trends observed between lineages I and II when the source of the isolate is considered. In most regions of the world, lineage I, in particular serotype 4b, is responsible for most of the human listeriosis outbreaks [9]. Conversely, most of the strains isolated from food products and food-related environments belong to lineage II, not lineage I [9]. There are some outbreaks and sporadic cases associated with lineage II, but lineage I is overrepresented when compared to its prevalence in food [9]. It has been speculated that these ecological differences are a result of increased virulence of lineage I strains, and an increased ability of lineage II strains to grow, survive, and persist in food and food-related environments [9]. Lineages I and II appear to be evenly represented in terms of animal isolates, and no definitive trends have been observed in regards to environmental isolates [9].

Phenotypically, strains from different lineages differ in their virulence, as well as in their ability to survive environmental stresses. For virulence, the increased prevalence of lineage I in human listeriosis outbreaks, which cannot be associated with increased prevalence in food, suggests that lineage I is more virulent than lineage II. This is supported by studies that show that lineage I is more virulent in mice, and demonstrates more capability to spread between mammalian cells, than lineage II [9]. Additionally, some studies have shown that a significant portion, possibly more than 30%, of lineage II strains contained premature stop codons in *inlA*, resulting in decreased invasion efficiency and less virulence [9]. One study showed that serotype 4b had more resistance

to heat treatment after cold storage than serotype 1/2a, as well as a shorter lag time [9]. Overall, specific stress response differences, such as resistance to phages and disinfectants, for specific lineages are inconsistent, but several studies indicate that the lineages respond differently to these types of stresses [9].

Most evidence suggests that the genomes of *L. monocytogenes* strains are highly conserved between lineages. The majority of genetic differences between lineages I and II are in genes related to proteins on the surface of the cell [9]. Another area of genetic difference between lineages I and II is metabolic function. Some studies have found that there are a number of lineage I-specific genes associated with sugar metabolism [10]. Other studies have found that there are genes that are unique to lineage II associated with a phosphoenolpyruvate-dependent phosphotransferase system and a 3-isopropylmalate dehydrogenase [11]. Additionally, slight differences between lineages have been noted for genes associated with virulence, encoding internalins, teichoic acid synthesis, stress response, and number of plasmids [9].

There are multiple methods used to subtype *L. monocytogenes* strains, which are separated into two groups [12]. The first, more traditional, group consists of ribotyping (RT) and pulse field gel electrophoresis (PFGE). Ribotyping is the fragmentation of the chromosomal DNA of *L. monocytogenes* using a restrictive endonuclease [12]. These fragments are then separated on an agarose gel, typically using PFGE, which results in a banding pattern that can be used to characterize some of the similarities of strains [12]. This technique can be used to compare a vast array of DNA fragment sizes and permits the possibility of using multiple enzymes to increase the discriminatory power of the technique [12]. This method has been the gold standard for subtyping, due to the fact that it can be used on every isolate of *L. monocytogenes*, it has good discriminatory capability, and because the scientific community has a standard, verified procedure that is used worldwide [12].

The second group, based on whole genome sequencing (WGS), has arisen since DNA sequencing has become faster and more affordable [12]. These techniques—including Multi Locus Variable Tandem Repeats Analysis (MLVA), DNA microarrays, Multi Locus Sequence Typing (MLST), and Single Nucleotide Polymorphism (SNP)

typing—provide great potential for more in-depth characterization of *L. monocytogenes* strains. However, due to the lack of consensus among the scientific community for standardized, validated procedures, RT and PFGE are still used for sub-typing *L. monocytogenes* [12].

1.4 Prevalence in Food

The Center for Disease Control and Prevention estimates that in the United States there are approximately 1,600 cases of human listeriosis every year, that result in approximately 255 deaths [1]. It is also estimated that 99% of these cases are of foodborne origin [1]. Studies have indicated that the main source for the contamination of food is the environment in which the food is processed [13]. The serotypes most prevalent in foods—and in the environments in which these foods are processed—are those of lineage II, most notably serotypes 1/2a, 1/2c, 3a, and 3c [13]. This is of interest because, as previously mentioned, the serotype most commonly responsible for cases of human listeriosis is 4b [9, 13]. The underrepresentation of serotype 4b in food and food processing environments suggests a higher degree of virulence, and that other serotypes may be better adapted to survive in these environments.

As of July 2016, there have been 288 recalls issued by the FDA and USDA due to contamination with *L. monocytogenes* since 2010. From 2010 through 2015 there were an average of 45 recalls per year, with the largest (57) and smallest (37) number of recalls for any single year in this span coming in 2012 and 2010, respectively. Over this timespan, dairy products were the most commonly recalled food products. Figure 1 depicts the food products involved in *L. monocytogenes* recalls since 2010.

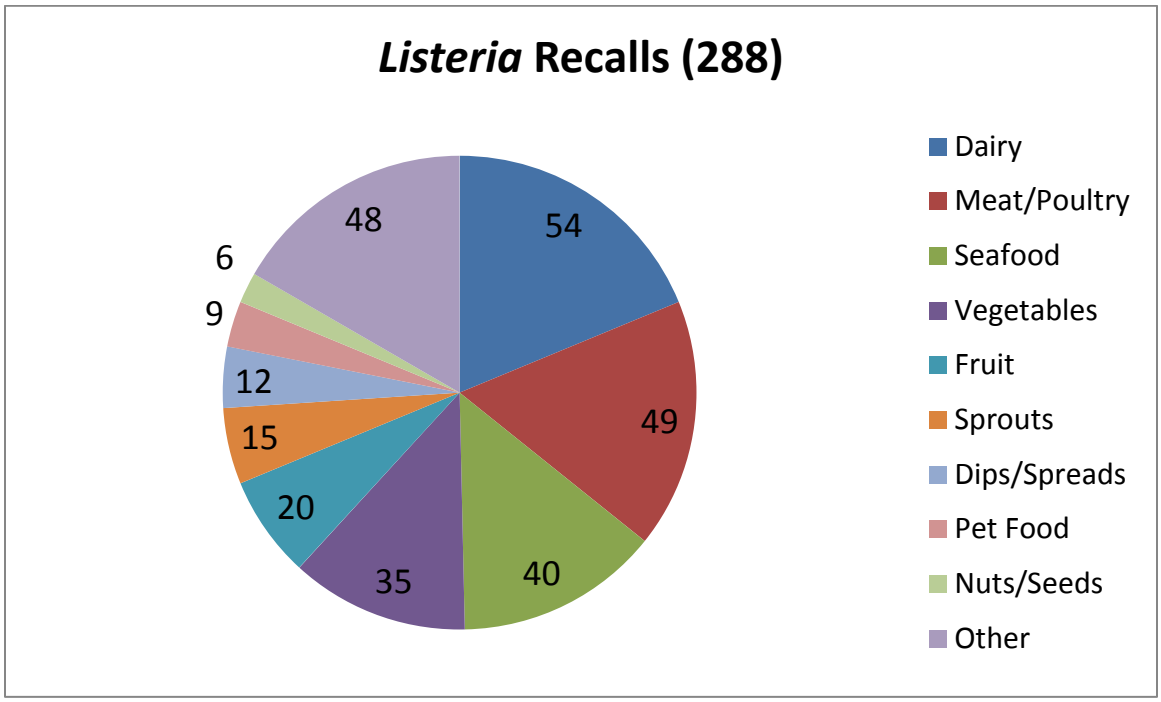

Figure 1: Number of food recalls associated with *L. monocytogenes* since 2010

Of the 288 recalls issued since 2010, 54 of them (18.8 %) were for dairy products. This is the largest single category of products recalled during this timeframe. The two most common types of dairy products recalled were cheeses (48) and ice cream (5). Frequently, the cheeses recalled were Hispanic cheeses, such as queso fresco, which are notoriously susceptible to contamination because of their comparatively high pH. Two of the recalls associated with dairy products that resulted in multi-state outbreaks of listeriosis are the Blue Bell Creameries recall in 2015 and the Oasis Brands, Inc. recall in 2014.

The Blue Bell Creameries recall resulted in ten cases of human listeriosis across four states: Kansas (5), Texas (3), Arizona (1), and Oklahoma (1). All ten of the patients were hospitalized, and three of the cases in Kansas resulted in death. The cases were linked to the consumption of Blue Bell Creameries ice cream products, in particular, the "Scoops" ice cream product and single serving food service ice cream cups. Upon investigation, the FDA found environmental *L. monocytogenes* samples in three Blue Bell facilities: one in Broken Arrow, Oklahoma, one in Brenham, Texas, and one in

Sylacauga, Alabama. The FDA findings noted that the sanitation practices in all three facilities were inadequate and too infrequent to prevent contamination. In addition, the facilities all had problems with insanitary equipment, most notably that condensate was allowed to drip onto product lines.

The Oasis Brands, Inc. recall resulted in five cases of human listeriosis spread across four states: Tennessee (2), Texas (1), New York (1), and Georgia (1). Four of the five people infected were hospitalized, and one of the two cases in Tennessee resulted in death. All of the patients were of Hispanic descent, and reported the consumption of Hispanic style cheese. The cases were then linked to the Hispanic style cheese and cheese products produced by Oasis Brands, Inc., most notably quesito casero and cuajada en hoja. While investigating the Miami, Florida production facility, the FDA discovered environmental samples of *L. monocytogenes*.

L. monocytogenes has been known to persist in food processing facilities for months or even years at a time [14]. Strains of *L. monoctyogenes* that are persistent are more likely to be found in food processing environments than they are in raw materials [14]. This is a major concern for food processing facilities, and makes it essential that cleaning procedures are routinely reviewed and validated to prevent the establishment of *L. monocytogenes* in an environmental niche. Persistent survival of *L. monocytogenes* in food processing facilities where cleaning and disinfection are common, might lead some to believe that *L. monocytogenes* forms a biofilm to aid in its protection and survival. However, the evidence is unclear and the studies that have tested the ability of *L. monocytogenes* to form biofilms have produced mixed results [14]. There is ample evidence that *L. monocytogenes* can rapidly and strongly adhere to a wide variety of materials that are common in food processing facilities [14]. However, a true biofilm requires that the cells involved produce an extracellular polysaccharide matrix, and many of the studies that attribute this strong adherence to the formation of a biofilm lack overwhelming evidence that this was the case [14].

1.5 Listeriosis

The human disease caused by *L. monocytogenes* is called listeriosis. There are an estimated 1,600 cases of listeriosis in the U.S. every year, and these cases result in approximately 255 deaths [1]. Listeriosis is primarily acquired through contaminated food, and has an abnormally high mortality rate (up to 30%) for a foodborne disease [15]. Like many foodborne illnesses, listeriosis is of highest concern for pregnant women, the elderly, and the immunocompromised. In fact, approximately 70% of listeriosis cases occur in individuals who are immunosuppressed to at least some degree, including cancer patients and transplant patients [15]. While some cases of listeriosis do occur in otherwise healthy individuals, they are far less frequent and usually less severe [15].

Since *L. monocytogenes* is a foodborne pathogen, the initial location of entry is usually the intestine. After passing through the wall of the intestine, its primary target is the liver [15]. When it reaches the liver, *L. monocytogenes* begins to multiply and spread. It accomplishes this by invading the normally non-phagocytic liver cells, quickly escaping a vacuole into the host cell's cytoplasm, rapidly multiplying while safe from the host's immune system, and directly spreading to nearby cells, wherein the cycle continues [15]. If the infection is unchecked, a distinct possibility in immunocompromised individuals, *L. monocytogenes* can then spread to secondary targets, such as the uterus in pregnant women, and the brain [15]. This can lead to manifestations of listeriosis that include meningitis, encephalitis, and abortions [15]. In otherwise healthy individuals, the most common manifestation of listeriosis is gastroenteritis [15].

1.6 *L. innocua* as a Biological Indicator

L. innocua is a species of *Listeria* that is not pathogenic to humans [4]. As such, it has potential to be used as a biological indicator in the food industry for the pathogenic species *L. monocytogenes*. Biological indicators can be used to determine the efficiency and effectiveness of sterilization procedures. However, to be effective, the indicator organism must exhibit the same, or superior, resistances to the treatment used [16].

Common treatments in the meat industry include heat, irradiation, and treatment with organic acids [17]. For *L. innocua* to be an effective indicator in meat, its resistance to these treatments must be similar to that of *L. monocytogenes*. Several studies have shown that *L. innocua* exhibits a similar response to heat to that of *L. monocytogenes* [16, 17]. The two species also display identical responses when exposed to gamma radiation [17]. It has also been shown that *L. innocua* and *L. monocytogenes* have similar tolerances for lactic acid, and that the effectiveness of lactic acid as a treatment is enhanced by the addition of sodium chloride [17]. Finally, treatment with sodium nitrite was not effective against either *L. innocua* or *L. monocytogenes* unless used in conjunction with either sodium chloride or low pH [17]. Since *L. innocua* displays the same responses to treatment as *L. monocytogenes*, it is evident that *L. innocua* would be an effective biological indicator for *L. monocytogenes* in meat.

L. innocua has also been determined to be an effective indicator in the production and ripening of low pH cheese [18]. Both species are inhibited by the initially low pH of the cheese, and showed a decline in population during the early stages of ripening. As the pH rises during the ripening process; however, the inhibition is lost and the populations of both species begin to recover [18]. Additionally, the two species displayed similar growth rates no matter where in the cheese blocks they were located [18]. Since the growth trends of *L. innocua* and *L. monocytogenes* were similar throughout the cheese making process, *L. innocua* can be considered an acceptable indicator for *L. monocytogenes* in cheese.

1.7 Genomics

The majority of the genomics work for *Listeria spp.* has been done on the pathogenic species *L. monocytogenes* and the closely related non-pathogenic species *L. innocua* [4, 19]. The genome of *L. monocytogenes* is approximately 2.9 Mbp, has a Guanine + Cytosine (G+C) content of approximately 38%, and contains approximately 2,900 Open Reading Frames (ORFs) [19]. In contrast, the *L. innocua* genome is slightly larger (3.0 Mbp), has a slightly lower G+C content (37%), and contains slightly more ORFs (approximately 3,000) [19]. Interestingly, the genome organization of *Listeria* spp.

is highly conserved between strains. Despite this, genomic and gene differences between strains are present, which is likely one reason for the variable pathogenicity of *Listeria* spp. [19].

In comparison to other microbial genomes, *Listeria* genomes are unusual in the percentage of genes that encode surface proteins. This percentage is 4.8% in *L. monocytogenes* EGDe, which is slightly higher than the 4.3% in *L. innocua* CLIP11626 [4, 19]. Another feature of *Listeria* genomes that differentiates them from other microbial genomes is the number of transport proteins encoded. For *L. monocytogenes* EGDe the percentage of genes that encode transport proteins is 11.6%, and for *L. innocua* CLIP11626 the percentage is 11.4% [4]. A third trait of *Listeria* genomes is the percentage of the genome dedicated to regulatory proteins: 7.3% for *L. monocytogenes* EGDe, and 7.1% for *L. innocua* CLIP11626 [4]. The high numbers of transport and regulatory proteins encoded in the genomes of *Listeria* spp. is likely the reason for its ubiquitous nature, its ability to adapt to a variety of environments, and its high tolerance of different types of stress.

Within the species *L. monocytogenes*, several studies have shown that strain specific genes cluster in such a way that they are able to be segregated into their respective lineages [19]. Interestingly, in many of the studies, genes known to encode virulence factors were very highly conserved between different strains of *L. monocytogenes*, which indicates that some other factor is likely responsible for the strain to strain differences in pathogenicity [19]. It is possible that this factor is related to surface protein encoding genes. Studies have shown that only a small fraction of the genes that encode surface proteins are present in all strains of *L. monocytogenes* [4]. Additionally, the surface protein genes present in each strain vary widely from strain to strain, with strains from the same lineages tending to group more closely together [4]. In contrast, the genes that encode transport and regulatory proteins are much more homogenous, which indicates that the surface protein encoding genes are much more likely to be responsible for some of the phenotypic differences between strains [4].

As of July 2016, The National Center for Biotechnology Information (NCBI) genome database contains 77 complete *Listeria* genomes, which are listed below in Table

1. Additionally, the database contains 701 *Listeria* genomes (data not shown) on which partial sequencing and assembly has begun, but not been completely finished.

Table 1: Completed genomes of *Listeria* spp. available at NCBI[a]

Name	Size (Mb)	GC%	Modify Date
Listeria monocytogenes	3.01839	38	6/4/2016
Listeria monocytogenes J1-220	3.03227	37.9	6/3/2016
Listeria monocytogenes J1816	2.94746	38	6/3/2016
Listeria monocytogenes	2.98189	38	5/12/2016
Listeria monocytogenes	2.92775	38	4/7/2016
Listeria monocytogenes	2.90535	38	4/7/2016
Listeria monocytogenes	3.03922	37.9715	4/7/2016
Listeria monocytogenes	2.94251	38	4/6/2016
Listeria monocytogenes	2.83987	38.2	4/6/2016
Listeria monocytogenes	2.84204	38.3	4/6/2016
Listeria monocytogenes ATCC 19117	2.95082	38	4/6/2016
Listeria monocytogenes	2.95075	38	4/6/2016
Listeria monocytogenes	2.89472	38	3/24/2016
Listeria monocytogenes	3.10812	37.9495	3/9/2016
Listeria monocytogenes	2.93973	38	3/9/2016
Listeria monocytogenes EGD	2.90719	38	2/25/2016
Listeria monocytogenes	2.86429	38	2/9/2016
Listeria monocytogenes	3.10552	38.042	2/5/2016
Listeria monocytogenes	3.04308	37.9763	2/5/2016
Listeria monocytogenes	2.90553	38	1/16/2016
Listeria monocytogenes 6179	3.07283	37.8717	10/9/2015
Listeria monocytogenes	2.89719	38	8/21/2015
Listeria monocytogenes	2.94748	38	8/21/2015
Listeria monocytogenes	2.89659	38.1	8/21/2015
Listeria monocytogenes	2.9325	38	8/21/2015
Listeria monocytogenes	2.94729	38	8/21/2015
Listeria monocytogenes	2.90749	38	8/21/2015
Listeria monocytogenes	2.91817	38	8/20/2015
Listeria monocytogenes	2.91164	38.1	8/19/2015
Listeria monocytogenes	2.95891	38	8/18/2015
Listeria monocytogenes	3.2433	37.9541	8/18/2015
Listeria monocytogenes	3.00952	37.9425	8/18/2015
Listeria monocytogenes	3.00952	37.9425	8/18/2015
Listeria monocytogenes	3.00951	37.9425	8/18/2015

Listeria monocytogenes serotype 4b str. F2365	2.90519	38	8/17/2015
Listeria monocytogenes	2.9045	38	8/17/2015
Listeria monocytogenes	2.90423	38	8/17/2015
Listeria monocytogenes 08-5578	3.10968	37.9653	8/16/2015
Listeria monocytogenes R479a	3.03165	37.8743	8/16/2015
Listeria monocytogenes	2.90667	38	8/16/2015
Listeria monocytogenes	2.98959	38	8/16/2015
Listeria monocytogenes	3.0916	37.8646	8/15/2015
Listeria monocytogenes	2.98895	38	8/15/2015
Listeria monocytogenes	2.94322	38	8/15/2015
Listeria monocytogenes WSLC1001	2.95124	38	8/15/2015
Listeria monocytogenes WSLC1042	2.94217	38	8/15/2015
Listeria monocytogenes	2.90466	38	8/15/2015
Listeria monocytogenes SLCC2372	3.02291	37.9586	8/14/2015
Listeria monocytogenes SLCC2540	2.97696	37.9	8/14/2015
Listeria monocytogenes SLCC7179	2.88223	38	8/14/2015
Listeria monocytogenes SLCC2479	2.97217	38	8/14/2015
Listeria monocytogenes ATCC 19117	2.9518	38	8/14/2015
Listeria monocytogenes SLCC5850	2.90714	38	8/14/2015
Listeria monocytogenes SLCC2376	2.84018	38.3	8/14/2015
Listeria monocytogenes L312	2.91235	38.1	8/14/2015
Listeria monocytogenes SLCC2378	2.94136	38	8/14/2015
Listeria monocytogenes FSL R2-561	2.9738	38	8/13/2015
Listeria monocytogenes Finland 1998	2.87443	38.1	8/13/2015
Listeria monocytogenes J0161	3.00046	37.9	8/13/2015
Listeria monocytogenes 10403S	2.90311	38	8/13/2015
Listeria monocytogenes 07PF0776	2.90156	38	8/13/2015
Listeria monocytogenes	2.9185	38.1	8/13/2015
Listeria monocytogenes 08-5923	2.99905	38	8/12/2015
Listeria monocytogenes serotype 4b str. CLIP 80459	2.91269	38.1	8/12/2015
Listeria monocytogenes SLCC2755	3.02393	38.0599	8/2/2015
Listeria monocytogenes serotype 4b str. LL195	2.90466	38	7/31/2015
Listeria monocytogenes HCC23	2.97621	38.2	7/30/2015
Listeria monocytogenes L99	2.9792	38.2	7/30/2015
Listeria monocytogenes M7	2.97616	38.2	7/30/2015
Listeria monocytogenes	2.77652	38.1	2/27/2015
Listeria monocytogenes N53-1	2.77685	38.1	2/27/2015

Listeria ivanovii subsp. ivanovii PAM 55	2.92888	37.1	12/18/2015
Listeria ivanovii subsp. londoniensis	2.99349	37	8/16/2015
Listeria ivanovii subsp. londoniensis	3.04503	37.1	8/16/2015
Listeria ivanovii subsp. ivanovii	2.91955	37.2	8/16/2015
Listeria ivanovii WSLC3009	2.91954	37.2	8/15/2015
Listeria seeligeri serovar 1/2b str. SLCC3954	2.79764	37.4	7/30/2015
Listeria welshimeri serovar 6b str. SLCC5334	2.81413	36.4	12/16/2014

[a]National Center for Biotechnology Information [20]

1.8 Systems Biology of Foodborne Pathogens

Systems biology is a field in which the combination of computational techniques and laboratory experiments are used to observe the components of cells and their interactions in order to better understand complicated cellular behaviors. The computational techniques provide a way to predict differences between strains in regards to metabolism and survival in different environments [21]. The models used for the computations consist of equations, variables, and parameters obtained from genome-scale biological networks. These genome-scale metabolic models (GEMs) are able to investigate the metabolic capabilities of several different microbes and have become increasingly useful in the last decade [21-23].

Using GEMs to study microorganisms began with studying the bacterium *E. coli* K-12, and has since expanded to over 100 different GEMs covering a variety of organisms [21, 24]. Presently, there are a total of seven GEMs associated with foodborne pathogens, each generated for a different strain of four different genera. Four GEMs have been constructed for different strains of *E. coli* O157:H7, and there is one GEM each for *Staphylococcus aureus*, *Vibrio vulnificus*, and *Salmonella* [25-30]. By using these GEMs, researchers have been able to simulate interactions between pathogen and host, study the metabolism of pathogens, and determine new methods of growth control and disease treatment [29-31]. Researchers have also been able to study the differences in metabolism between evolutionarily related pathogenic and commensal strains of *E. coli* [25, 26].

Computational modeling of bacterial metabolism offers a promising approach to predict strain-to-strain variation in metabolic capabilities and microbial strategies used in different environments. Systems biology combines computational and experimental approaches to study the complexity of biological networks at a systems level, where the cellular components and their interactions lead to new understanding of complex cellular behaviors and new candidate targets to kill bacteria. To date, GEMs of foodborne pathogens have been used successfully to identify essential reactions leading to new targets and methods to control growth and survival of *Salmonella*, *Staphylococcus aureus*, and *Vibrio vulnificus* [29-31].

1.9 Biochemical Pathways

L. monocytogenes is a ubiquitous organism that has been isolated from a number of different environments. This implies that *L. monocytogenes* must have robust metabolic capabilities that allow it to thrive in a multitude of environmental conditions. Most notably, when food contaminated with the organism is consumed by humans, the organism is able to adjust its metabolism in order to survive and replicate intracellularly [32]. The genes involved in this metabolic switch are largely regulated by the transcriptional regulator PrfA [32]. Nutrient availability has a profound impact on the expression of these PrfA-regulated genes and their products. For example, easily metabolized carbohydrates, such as glucose, have a severe inhibitory effect on the expression of these gene products. Whereas phosphorylated sugars, such as glucose-6-phosphate, do not [32]. This indicates that nutrient availability (e.g. glucose in the environment or glucose-6-phosphate inside host cells) is used as a signal for the optimum metabolic state of *L. monocytogenes*.

Recently, a study was performed that tested approximately 1,500 *L. monocytogenes* mutants for their phenotypes in both rich media (BHI) and using a Caco-2 cell culture assay [33]. In this work, the researchers were able to identify 141 genes that are important for invasion of host cells or for intracellular replication. Schauer *et al*. discovered that 27 of these genes were associated with transporters and lipoproteins, including multiple responsible for the uptake of sugars [33]. Additionally, they

discovered that 47 of the genes were associated with the metabolism of carbohydrates, amino acids, lipids etc. Analysis of the metabolism genes identified indicates that during intracellular replication of *L. monocytogenes* the pentose phosphate pathway is the primary sugar metabolism pathway instead of glycolysis [33]. Analysis of all 141 identified genes, along with subsequent testing in mice and with metabolic models, indicated that *L. monocytogenes* uses distinctly different metabolic pathways during intracellular replication than it does when grown on rich media. Additionally, it appears that during intracellular replication *L. monocytogenes* uses a large number of slightly overlapping pathways to utilize storage or excess products of the host cell's metabolism [33].

2 Generation of Genome-Scale Metabolic Models and Qualitative Validation

2.1 Abstract

The objective of this study was to generate genome-scale metabolic models (GEMs) for six different strains of *L. monocytogenes*, and to compare nutrient utilization predictions of these models to *in vitro* results. Genomes for six different strains of *L. monocytogenes* were taken from the NCBI database and uploaded to KBase—a semi-automated program used to generate the GEMs. These models were then used to generate nutrient utilization predictions for sources of carbon, nitrogen, phosphorus, and sulfur using General Algebraic Modeling System (GAMS) software. *In silico* predictions were then compared to *in vitro* experiments performed using Biolog™ phenotypic microarray plates. Following this comparison, GEMs were refined to increase agreement between *in silico* predictions and *in vitro* results.

57 of the 95 carbon sources tested *in vitro* were present in the models, and; therefore, these were the compounds from which comparisons could be drawn. Of these 57 compounds, agreement between *in silico* predictions and *in vitro* results ranged from 80.7% to 91.2% between strains. For nitrogen, 62 of the 95 compounds were present, and agreement ranged from 59.7% to 66.1%. For phosphorus, 22 of the 59 compounds were comparable, and agreement ranged from 18.2% to 27.3%. 11 of the 35 sulfur compounds were comparable, and agreement ranged from 72.7%% to 81.8%. The carbon and sulfur utilization agreements are similar to the utilization agreement of previously published GEMs. However, the nitrogen and phosphorus utilization agreements fall short of the agreements seen by previous GEMs. This indicates that the GEMs created by this study can be put to similar uses as previous GEMs, but there is room for improvement, particularly in the investigation of nitrogen and phosphorous metabolic pathways.

2.2 Introduction

Advancements in sequencing and annotation technology have made it increasingly more practical to use fully sequenced genomes in their entirety to study organisms. One such approach is the creation of genome-scale metabolic networks and their translation into computational models (GEMs). This work describes the creation of

six GEMs for different strains of *L. monocytogenes* and their qualitative validation through experimental data.

The first step in obtaining a working GEM for the study of *L. monocytogenes* is the creation of the base version of that model. The traditional method of model creation involves manually mining available genomic literature for the organism, and compiling this information into a biological network linking all of the genes, proteins, and reactions, which is a process that can take years to complete. The work presented here investigates the validity of using semi-automated online tools to construct the draft GEM, which can be done in a span of hours. The semi-automated process uses publicly available genomic information alongside online database tools to generate draft GEMs. The draft GEMs are then curated and expanded through an iterative process to validate computational predictions by comparison to experimental data. The experimental data used for the qualitative validation in this study was the Biolog^{TM} Phenotypic Microarray (PM) plate system.

Biolog^{TM} PM plates are 96 well plates used to detect cellular respiration [34]. The inoculum of these plates consists of a suspension of cells combined with a tetrazolium redox dye, which detects respiration. At the bottom of each of the wells is a dried media that allows for the determination of the phenotype of the organism studied. This media consists of all but one of the nutrients required for cellular growth. Each well also contains one compound to be tested. For example, the PM1 and PM2 plates are used to test for carbon source utilization. This means that the basic media in each well contains all of the nutrients required for growth except for carbon, and the individual compound to be tested in each well is a different source of carbon. The same holds true for the other plates used in this experiment to test nitrogen sources (PM3) as well as phosphorus and sulfur sources (PM4) [34].

Utilization of the compounds of interest is detected colorimetrically by use of the tetrazolium redox dye, and also through the monitoring of absorbance using a spectrophotometer. If the organism has the necessary transport and metabolic reaction systems to utilize the compound of interest, respiration occurs. The electron transport resulting from cellular respiration will result in the reduction of the tetrazolium dye,

which changes its color from colorless to purple. The turbidity in the wells can then be measured spectrophotometrically, which gives another indication of the degree to which the compound of interest can be utilized by the organism [34].

The six *L. monocytogenes* strains chosen for this study all come from one of the three serovars most frequently responsible for human listeriosis cases: 1/2a, 1/2b, and 4b. Strains J2-031 and JO161 come from serovar 1/2a, which belongs to lineage II. Strains J2-064 and R2-502 belong to serovar 1/2b and lineage I. Finally, Strains F2365 and ScottA belong to serovar 4b and lineage I. The six GEMs created for this strain will provide a fresh perspective on the metabolism of *L. monocytogenes* at the genome level, an approach that has previously been unexplored.

2.3 Materials and Methods

2.3.1 Comparative Genome Alignment

The six strains chosen for this study represent a variety of isolation sources as well as each of the three most prevalent serovars in terms of human listeriosis cases (Table 2).

Table 2: The six strains of *L. monocytogenes* chosen for this study

***L. monocytogenes* strain (ILSI #)**	**Serovar**	**Source**
J2-031	**1/2a**	**Animal isolate (bovine) 1996**
JO161 (R2-499)	**1/2a**	**Human epidemic 2000**
R2-502	**1/2b**	**Food epidemic 1994 (Illinois)**
J2-064	**1/2b**	**Animal isolate (bovine) 1989**
F2365 (J1-110)	**4b**	**Food epidemic 1985 (L.A)**
ScottA (J1-225)	**4b**	**Human epidemic 1993**

Genome alignments were performed using Mauve version 2.3.1 [35, 36]. Alignments were conducted using GenBank genome files obtained from the National Center for Biotechnology Information (NCBI) database [10, 37-40], and were performed using progressive Mauve using default parameters.

2.3.2 Stock Culture Maintenance

Culture stocks were kept at -80°C for long term storage. Strains were streaked onto brain heart infusion (BHI) agar plates (Becton, Dickinson and Company, Sparks, MD) and stored at 4°C. The refrigerated cultures were re-streaked onto fresh BHI agar plates every two weeks as needed.

2.3.3 Generation of Genome-Scale Metabolic Models

Six *Listeria monocytogenes* strain genomes (F2365, J2-031, J2-064, JO161, R2-502, and Scott A) were downloaded in GenBank file format from the NCBI database. These genomes were then uploaded to KBase, an online genome-scale metabolic reconstruction pipeline sponsored by the U.S. Department of Energy, and draft GEMs were created using the semi-automated tools provided by KBase that is built with the same bioinformatics pipeline as the ModelSeed [41]. In the first step of the Model SEED or K-Base pipeline, the assembled genome sequence is annotated by the Rapid Annotation using Subsystem Technology (RAST) server [42] and imported into the SEED analysis system. Next, a preliminary GEM is generated consisting of intracellular and transport reactions associated with genes on the basis of RAST annotations, spontaneous reactions and an organism-specific biomass reaction. In the gap-filling step of the pipeline, additional intracellular and transport reactions are added to create an analysis-ready GEM capable of simulating biomass production using only transportable nutrients. Flux Balance Analysis is then used to generate phenotype predictions in the model analysis step. The final three steps of the pipeline involve manual curation for the removal and addition of reactions from the model to fit Biolog phenotyping array data (when available) and gene essentiality data (when available) to produce an optimized model.

To make GEMs functional, it is a common practice to identify missing reactions required for metabolic pathways that are known to be functional based on experimental evidence, but the gene encoding the reaction may have not yet been identified. This gap-filling process was required in this study to fill in missing metabolic reactions required for the utilization of D-glucose, a compound that is known to be able to be metabolized

by *L. monocytogenes* and nearly all free living bacteria. Therefore, using KBase, gap-filling was performed on the draft *L. monocytogenes* models for D-glucose under both aerobic and anaerobic conditions. Then the draft models were downloaded in systems biology markup language (SBML) file format, and these SBML files were then converted to General Algebraic Modeling System (GAMS) file format to enable their utilization by GAMS software. GAMS software was used to conduct computational analysis using established methods such as Flux Balance Analysis (FBA), dynamic FBA, and Gene/Reaction essentiality predictions from the COBRA toolbox [43].

2.3.4 *In Vitro* Nutrient Utilization

Biolog™ PM1, PM3, and PM4 plates (Biolog, Inc., Hayward, CA) were used for all *in vitro* nutrient utilization experiments. When preparing Biolog™ plates, three isolated colonies from each refrigerated culture plate were inoculated into separate test tubes containing 9 mL of BHI broth. The BHI tubes were then incubated at 37°C overnight. After incubation, two tubes for each strain were chosen and 300 µL from each tube were taken and inoculated onto separate Biolog Universal Growth plus sheep's Blood (BUG+B) agar plates in triplicate (6 plates per strain). The BUG+B plates were then incubated at 37°C overnight. After incubation, one BUG+B plate was harvested using a sterile cotton swab, and cells were placed in 15 mL conical tubes containing 1.8 mL of IF-0 solution, prepared as directed by the manufacturer. 150 µL of cell suspension was removed, placed in a new 15 mL conical tube, and the OD_{590} was then adjusted to 0.171 ± 0.020 by diluting with fresh IF-0 solution. After OD_{590} adjustment, 1.8 mL of the diluted cell suspension was added to 15 mL conical tubes containing 9 mL of IF-0+ solution that had been prepared according to manufacturer instructions. For PM3 and PM4 plates, 108 µL of 100X ferric citrate/sodium succinate solution was added to each conical tube containing cellular suspension. The resulting solution was then transferred to a sterile reservoir, and then transferred to separate wells of Biolog™ plates, separate plates for each replicate of each strain (12 total), in 100 µL aliquots. 640 nm absorbance values were then taken at 0, 12, 24, 48, and 72 hour time points using a ChroMate® Microplate Reader (Awareness Technology, Inc., Palm City, FL).

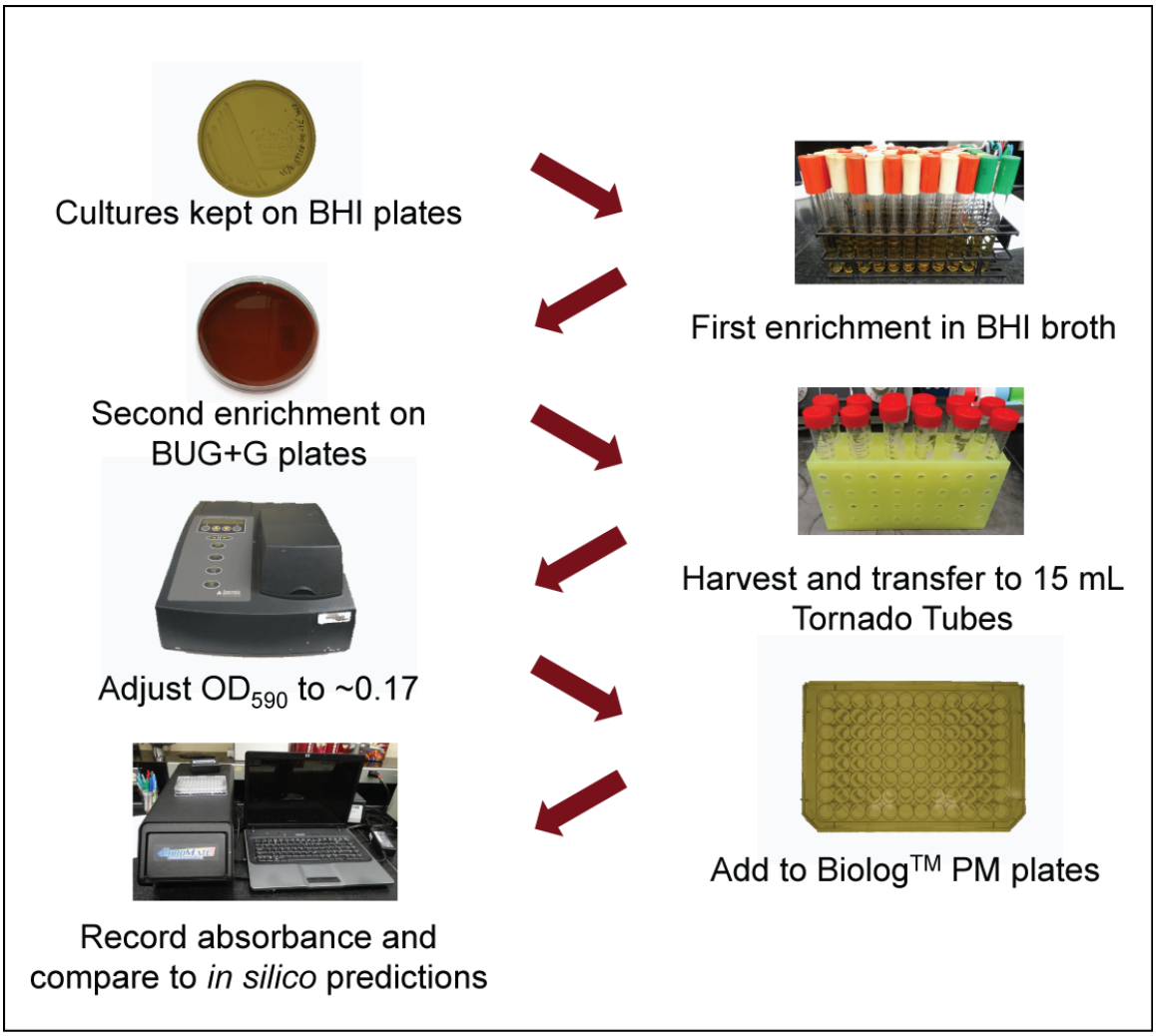

Figure 2: Biolog™ Phenotypic Microarray experimental flow diagram

2.3.5 *In Silico* Nutrient Utilization Prediction and Reconciliation

GAMS is an optimization software package and was used to perform flux balance analysis (FBA) on each of the models to determine predicted nutrient utilization capabilities using methods previously described [43]. FBA is a technique frequently used while studying GEMs. This is because FBA makes growth rate predictions possible by calculating the flow of metabolites through the network [44]. The foundation of FBA is the translation of a metabolic network into a matrix of stoichiometric coefficients, the S matrix. This matrix includes each metabolite in the network as a separate row and each metabolic reaction as a separate column. Representing the metabolic network in this

fashion places constraints on the system in the form of a steady state mass balance, that is, the rate of consumption of metabolites cannot exceed the rate of production. Additional constraints, such as maximum or minimum values for reaction fluxes, can also be implemented [44]. The function of these constraints is to restrict the possible range of reaction fluxes into a defined solution space. Growth rate predictions are enabled by representing the production of biomass as a metabolic reaction, which serves as the objective function, and adds a column to the S matrix. This allows the optimization of the objective function by calculating the flux distribution that results in the maximum flux through the biomass reaction [44]. In terms of nutrient utilization, when the optimized biomass reaction results in zero flux, it is determined that the organism cannot utilize that nutrient as a sole source of carbon, nitrogen, phosphorous, or sulfur.

After the initial nutrient utilization predictions, the metabolites contained in the GEMs were surveyed to identify compounds matching those present on Biolog™ plates that did not already have corresponding transport reactions in the draft GEMs. After identifying the missing compounds, the models were manually curated to add the required reactions in order to generate predictions for *in silico* nutrient utilization. These predictions were then compared to *in vitro* nutrient utilization experiments using Biolog™ plates.

Comparisons of *in silico* predictions to *in vitro* results generated a list of disagreements necessary to reconcile for validation of GEMs predictive accuracy. The GEMs were then manually curated to remove transport reactions associated with compounds for which *in vitro* results showed no growth but *in silico* predictions predicted growth (false positives). Compounds for which *in silico* simulations predicted no growth, but *in vitro* experiments showed growth (false negatives) were reconciled by first using KBase to gapfill the draft models on those compounds to identify any necessary missing metabolic reactions. The reactions in the resultant secondary draft models were then used as templates by which the manually curated models could be updated. This step was necessary because the manually curated models could not be re-uploaded to KBase and repeatedly gapfilled. The gapfilling, manual curation, and

Biolog™ experimental validation techniques presented here are common methods used to validate and optimize GEMs [25, 45-47].

2.3.6 Essential Reaction Predictions

Essential reactions are those required by a microorganism for viability and/or growth in a given environmental condition. Therefore, determination of essential reactions identifies metabolic reactions and corresponding genes as ideal targets for control of viability and growth. Predictions of essential reactions can be accomplished *in silico* by adding additional constraints to the FBA simulations of GEMs. The constraints introduced during essential reaction prediction simulations serve to restrict the flux through each metabolic reaction one at a time. FBA then re-optimizes the objective function (biomass production) while the single reaction is restricted. If, lacking the specific metabolic reaction, the predicted biomass is zero, the restricted reaction is deemed essential. This approach for predicting essential reactions with GEMs is well established [25, 48]. Using a similar approach, where essential metabolites were predicted Kim *et al.* were able to identify successful metabolite-mimicking drugs that had a bactericidal effect on the opportunistic foodborne pathogen, *Vibrio vulnificus* [30].

Essential reaction predictions were performed after final reconciliation between *in silico* predictions and *in vitro* experimental results. GAMS was used to simulate gene knockouts, eliminating one metabolic reaction at a time by constraining the flux to zero in order to identify which of those reactions were essential to the survivability and viability of each of the six strains of *L. monocytogenes*. Essential reactions were determined by a prediction of zero biomass generation associated with the removal of the reaction. Essential reaction predictions were first performed in conditions reflecting the nutrients present in a minimal media with glucose as the sole energy source (Table 3). Subsequently, the chemical composition of five of the foods most commonly recalled for *L. monocytogenes* contamination was obtained from the United States Department of Agriculture (USDA) nutrient database. Using this chemical composition, compounds that were both present in the foods and present in the models were added to the environmental simulations to see if any differences in essential reaction predictions arose

when growth was simulated on each food item. The five food items for which essential reaction predictions were generated were queso fresco, chicken breast, smoked salmon, cantaloupe, and romaine lettuce (Table 4).

Table 3: Nutrients present in the essential reaction simulation reflecting minimal media

Environment	Nutrients
Glucose Minimal Media	H_2O, PO_4^{3-}, CO_2, NH_3, Mn^{2+}, Zn^{2+}, SO_4^{2-}, Cu^{2+}, Ca^{2+}, H^+, Cl^-, Co^{2+}, K^+, Mg, cob(I)alamin, Na^+, Fe^{2+}, Fe^{3+}, O_2, and D-glucose

Table 4: Nutrients added to essential reaction simulations to reflect high-risk foods

Food	Additional Nutrients
Queso Fresco	L-Glutamate, Glycine, L-Lysine, L-Aspartate, L-Arginine, L-Glutamine, L-Serine, L-Methionine, L-Tryptophan, L-Phenylalanine, L-Tyrosine, L-Cysteine, Choline, L-Leucine, D-Alanine, L-Histidine, L-Proline, L-Asparagine, L-Valine, L-Threonine, Palmitate, Riboflavin, Thiamin, L-Isoleucine, Vitamin B12
Chicken Breast	L-Glutamate, Glycine, L-Lysine, L-Aspartate, L-Arginine, L-Serine, L-Methionine, L-Tryptophan, L-Phenylalanine, L-Tyrosine, L-Cysteine, Choline, L-Leucine, D-Alanine, L-Histidine, L-Proline, L-Valine, L-Threonine, Palmitate, Riboflavin, Thiamin, L-Isoleucine, Vitamin B12
Smoked Salmon	L-Glutamate, Glycine, L-Lysine, L-Aspartate, L-Arginine, L-Serine, L-Methionine, L-Tryptophan, L-Phenylalanine, L-Tyrosine, L-Cysteine, Choline, L-Leucine, D-Alanine, L-Histidine, L-Proline, L-Valine, L-Threonine, Palmitate, Riboflavin, Thiamin, L-Isoleucine, Vitamin B12
Cantaloupe	L-Glutamate, Glycine, L-Lysine, L-Aspartate, L-Arginine, L-Serine, L-Methionine, L-Tryptophan, L-Phenylalanine, L-Tyrosine, Sucrose, D-Fructose, L-Cysteine, Choline, L-Leucine, D-Alanine, L-Histidine, L-Proline, L-Valine, L-Threonine, Maltose, Palmitate, Riboflavin, Thiamin, L-Isoleucine
Romaine Lettuce	L-Glutamate, Glycine, L-Lysine, L-Aspartate, L-Arginine, L-Serine, L-Methionine, L-Tryptophan, L-Phenylalanine, L-Tyrosine, D-Fructose, L-Cysteine, Choline, L-Leucine, D-Alanine, L-Histidine, L-Proline, L-Valine, L-Threonine, Palmitate, Riboflavin, Thiamin, L-Isoleucine

2.4 Results

2.4.1 Genomic Comparison

As microorganisms evolve over millions of years, changes in the genome may occur, leading to differences in total gene content in modern strains. Therefore, we sought to determine genome level differences in the *L. monocytogenes* strains chosen for this work. The first step was to ascertain the degree to which the genomes of the strains differed. This was accomplished through the use of Mauve genome alignment software [35, 36]. The default output of Mauve differentiates the genomes into color-coordinated regions of homologous sequence that does not contain any rearrangements, termed Locally Collinear Blocks (LCBs) (Figure 3). Identical colors show areas of conservation between the genomes, and partial white lines in the LCBs indicate nucleotide sequence variation. 10 LCBs were determined to exist in each of the six genomes examined (Figure 3). In contrast, white regions that span the entire width of the LCB indicate areas of unique genomic content and contain ORFs unique to each strain.

An alternative Mauve output, termed the "backbone" view, makes all genomic content that is shared between all six strains a flat, mauve color (Figure 4). The remaining areas show conservation between subsets of genomes, depicted by similar colors in the applicable strains. For example, the sequence given a green-blue color near 2,700,000 bp in strain J2-031 and near 800,000 bp in strain JO161 is unique to serovar 1/2a. Another example is the green colored sequence near 1,000,000 bp in strain J2-064 and near 2,250,000 bp in strain R2-502, which is unique to serovar 1/2b. A third example is a sequence unique to serovar 4b, given a blue-green color near 2,270,000 bp in strain F2365, and near 520,000 bp in strain ScottA. Finally, a sequence given a brown color near 2,400,000 bp in strain JO161 and near the beginning of the genome in strain ScottA is unique to those two strains. The genomic islands unique to a single genome are also present, displayed in white.

A numerical comparison of the genomes of the six strains was accomplished using the RAST (Rapid Annotation using Subsystem Technology) database (Table 5) [42]. Each strain has approximately 3,000 genes, with 85% – 90% of those genes shared

among all of the strains. Less than 3% of the genes for each strain are unique, and none of the three serovars chosen had more than 53 genes unique to that serovar.

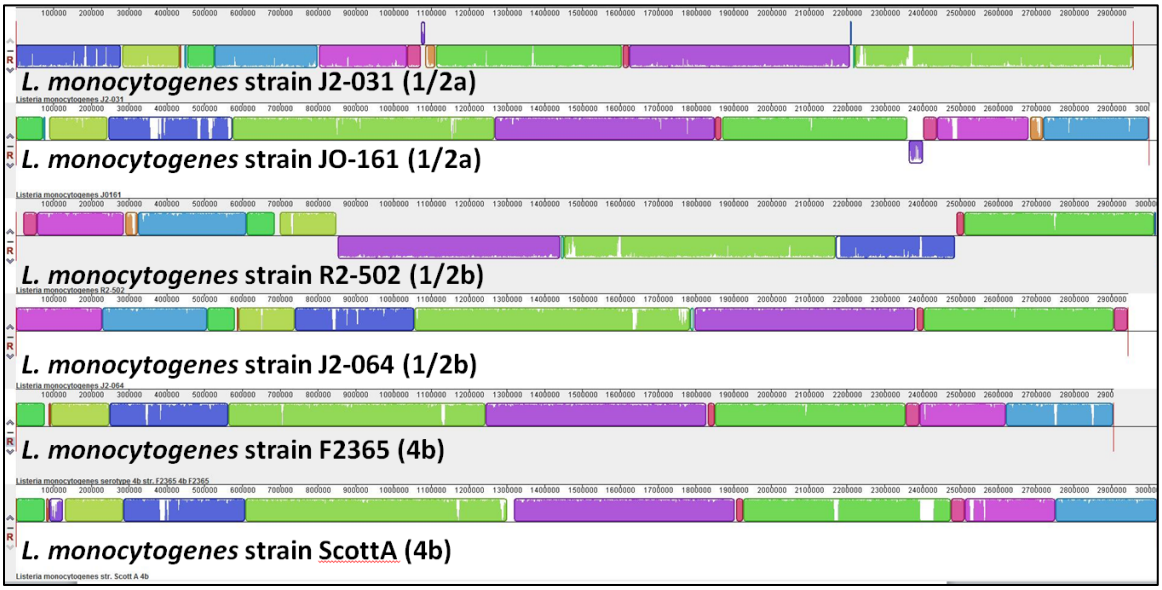

Figure 3: Mauve genome alignment for the six chosen strains of *L. monocytogenes*

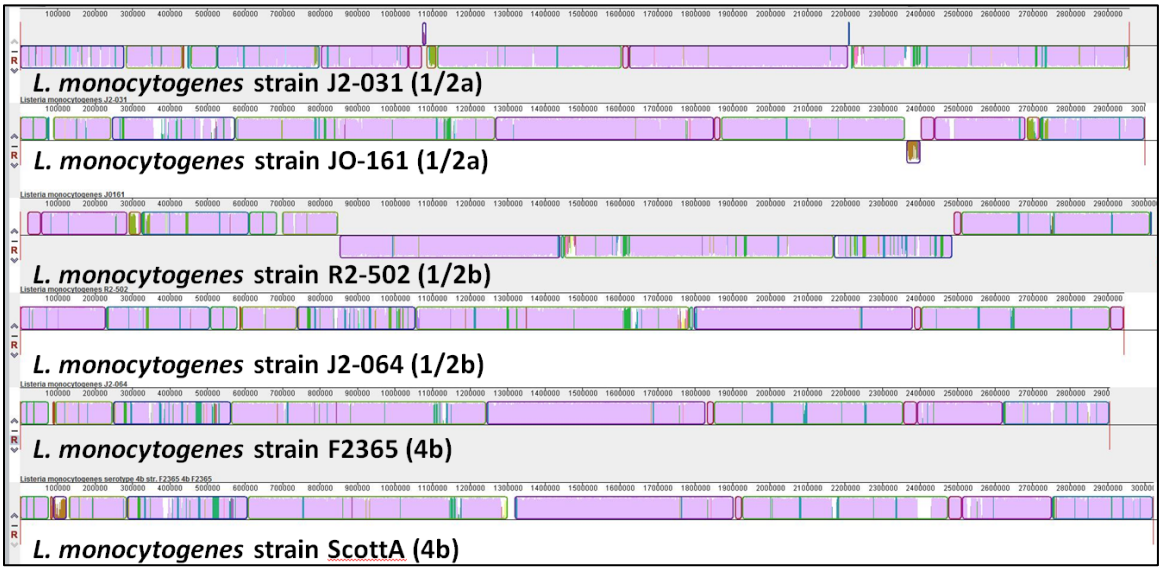

Figure 4: Mauve backbone alignment for the six chosen strains of *L. monocytogenes*

Table 5: Numerical genome comparison of the six chosen strains of *L. monocytogenes*

Strain	Genes (% Shared)	Unique Genes (% of Genome)	Serovar	Serovar Specific Genes
J2-031	3,009 (86.6%)	49 (1.63%)	1/2a	53
JO161	3,024 (86.1%)	65 (2.15%)		
J2-064	2,938 (88.7%)	16 (0.54%)	1/2b	13
R2-502	3,069 (84.9%)	57 (1.86%)		
F2365	2,907 (89.6%)	32 (1.10%)	4b	26
ScottA	3,016 (86.4%)	73 (2.42%)		

2.4.2 Reconstruction of Genome-Scale Metabolic Models

The number of genes, metabolites, and metabolic reactions present in GEMs are important characteristics (Table 6). The created GEMs contained just under 800 genes for each strain (slightly more than 25% of the ORFs contained in each genome). Each GEM also contained over 1,100 metabolites, and a cumulative total of 1,116 metabolites were contained in the six GEMs. 1,106 of these metabolites were shared between all six of the *L. monocytogenes* strains. There was also one metabolite unique to *L. monocytogenes* strain J2-064 (p-Hydroxybenzaldehyde), and one metabolite unique to serovar 1/2a (Toxopyrimidine). Draft GEMs contained over 1,050 reactions. Surveying the metabolites contained in the stoichiometric matrix of each of the GEMs for additional carbon sources added almost 200 reactions, and the final curation added approximately 100 reactions to each GEM. The cumulative number of reactions contained in all six GEMs was 1,335, and 1,300 of these reactions were shared among all six GEMs. Of the 1,335 possible reactions, one was unique to *L. monocytogenes* strain F2365, seven were unique to *L. monocytogenes* strain J2-031, two were unique to *L. monocytogenes* strain J2-064, and one reaction was unique to serovar 1/2a (Table 7).

Table 6: Gene, metabolite, and reactions contained in each version of the GEMs

Strain	Genes (% of Genome)	Metabolites	Draft GEM Reactions	Reactions After Validation with Carbon Data	Final Reactions
J2-031	770 (25.6%)	1,112	1,053	1,219	1,320
JO161	786 (26.0%)	1,115	1,055	1,221	1,320
J2-064	783 (26.7%)	1,114	1,056	1,221	1,320
R2-502	786 (25.6%)	1,114	1,057	1,223	1,322
F2365	780 (26.8%)	1,110	1,053	1,219	1,318
ScottA	779 (25.8%)	1,110	1,053	1,219	1,318

Table 7: Enzymes catalyzing strain and serovar-specific metabolic reactions

Strain/Serovar	Unique Enzymes
J2-031	S-Adenosyl-L-homocysteine hydrolase, rxn00257_c0, Isocitrate glyoxylate-lyase, alpha-D-Glucose-1-phosphate:alpha-D-glucose-1-phosphate, Xanthosine-5'-phosphate:L-glutamine amido-ligase (AMP-forming), rxn01615_c0, 4-amino-5-hydroxymethyl-2-methylpyrimidine synthetase_c0
J2-064	4-hydroxybenzaldehyde:NAD+ oxidoreductase, 4-hydroxy-benzyl-alcohol dehydrogenase
F2365	Adenosyl cobinamide kinase
1/2a	ATP:4-amino-5-hydroxymethyl-2-methylpyrimidine 5-phosphotransferase_c0

2.4.3 Nutrient Phenotype Data

Also of interest is the degree to which the different strains and different serovars of *L. monocytogenes* differ in their nutrient utilization. Biolog^{TM} PM plates are widely used to quickly generate large quantities of nutrient utilization data that can be used to qualitatively validate GEMs. Upon completion of the Biolog^{TM} PM experiments, strain and serovar specific nutrient utilization data became available for these six *L. monocytogenes* strains. 95 sources of carbon, 95 sources of nitrogen, 59 sources of phosphorous, and 35 sources of sulfur were analyzed, and the number of compounds capable of serving as the sole source of the respective nutrients was determined (Tables 8 – 11). The strain-specific and serovar-specific differences in nutrient utilization were

also examined for the same four nutrients (Table 12), and the unique nutrients for each strain and serovar were identified (Tables 13 – 15).

Table 8: Summary of experimental carbon utilization data

Substrate	J2-031	JO161	J2-064	R2-502	F2365	ScottA
L-Arabinose	-	+	-	+	-	+
N-Acetyl-DGlucosamine	+	+	+	+	+	+
D-Saccharic Acid	+	+	-	-	+	+
Succinic Acid	+	+	-	+	+	-
D-Galactose	+	+	-	+	+	-
L-Aspartic Acid	+	-	-	+	-	-
L-Proline	-	-	-	-	-	-
D-Alanine	-	-	-	-	-	-
D-Trehalose	+	+	+	+	+	+
D-Mannose	+	+	+	+	+	+
Dulcitol	-	-	-	-	-	-
D-Serine	-	-	-	-	-	-
D-Sorbitol	+	-	-	+	-	+
Glycerol	-	+	-	-	+	+
L-Fucose	+	+	+	+	+	-
D-Glucuronic Acid	+	-	+	+	+	+
D-Gluconic Acid	+	+	+	+	+	+
D,L-α-Glycerol-Phosphate	-	+	-	-	+	-
D-Xylose	+	-	-	+	+	-
L-Lactic Acid	+	+	-	-	+	-
Formic Acid	+	-	-	-	-	-
D-Mannitol	-	-	-	-	-	-
L-Glutamic Acid	-	-	-	-	-	-
D-Glucose-6-Phosphate	-	-	-	-	-	-
D-Galactonic acid-y-lactone	+	-	-	+	-	-
D,L-Malic Acid	+	+	-	+	+	+
D-Ribose	+	+	+	+	+	+
Tween 20	+	+	+	+	-	-
L-Rhamnose	+	-	-	+	-	+

D-Fructose	+	+	+	+	+	+
Acetic Acid	-	+	-	-	+	-
α-D-Glucose	+	+	+	+	+	+
Maltose	+	+	+	+	+	-
D-Melibiose	-	+	-	-	-	-
Thymidine	+	+	+	+	-	+
L-Asparagine	-	-	-	-	-	-
D-Aspartic Acid	+	-	-	-	-	-
D-Glucosaminic Acid	-	+	-	-	+	-
1,2-Propanediol	+	-	+	+	+	-
Tween 40	+	-	+	+	-	-
α-Keto-Glutaric Acid	-	-	+	-	-	+
α-Keto-Butyric acid	+	+	-	+	+	-
α-Methyl-DGalactoside	+	+	-	+	+	-
α-D-Lactose	-	+	+	+	-	-
Lactulose	+	-	-	-	+	-
Sucrose	-	-	-	-	+	-
Uridine	+	+	+	+	+	+
L-Glutamine	-	-	-	-	-	-
M-Tartaric Acid	+	-	-	-	-	-
D-Glucose-1-Phosphate	+	-	-	+	+	-
D-Fructose-6-Phosphate	+	-	+	+	+	+
Tween 80	+	-	+	-	-	-
α-Hydroxy Glutaric Acid-γ-Lactone	-	-	-	-	+	-
α-Hydroxy Butyric Acid	+	+	-	+	+	-
β-Methyl-DGlucoside	-	+	+	+	-	+
Adonitol	-	-	-	+	-	-
Maltotriose	+	+	+	+	+	+
2-Deoxy Adenosine	+	+	-	-	-	+
Adenosine	+	+	+	+	+	+
Glycyl-L-Aspartic Acid	-	-	-	-	-	-
Citric Acid	+	+	-	+	-	-
M-Inositol	-	-	-	+	-	-
D-Threonine	-	+	-	-	-	-

Fumaric Acid	-	-	-	-	+	-
Bromo Succinic	+	-	-	-	+	-
Propionic Acid	+	-	-	+	-	-
Mucic Acid	+	+	-	+	+	-
Glycolic Acid	+	-	-	+	+	-
Glyoxylic Acid	-	+	-	-	-	-
D-Cellobiose	+	+	+	+	+	+
Inosine	+	+	-	+	+	-
Glycyl-Lglutamic acid	-	-	-	-	-	-
Tricarballylic Acid	+	-	-	+	-	-
L-Serine	-	+	-	-	-	-
L-Threonine	+	+	-	+	+	-
L-Alanine	+	+	-	-	+	-
L-Alanyl-Glycine	+	+	-	-	+	-
Acetoacetic Acid	-	+	+	+	+	-
N-Acetyl-β-D-Mannosamine	-	-	-	-	-	-
Mono Methyl Succinate	-	-	-	-	-	-
Methyl Pyruvate	-	-	-	-	-	-
D-Malic Acid	-	-	-	-	-	-
L-Malic Acid	-	-	-	-	-	-
Glycyl-L-Proline	-	-	+	-	-	-
p-Hydroxy Phenyl Acetic Acid	-	-	+	+	-	-
m-Hydroxy Phenyl Acetic Acid	-	-	-	-	-	-
Tyramine	-	-	-	-	+	-
D-Psicose	-	-	-	-	-	-
L-Lyxose	-	-	-	+	-	-
Glucuronamide	-	-	-	-	-	-
Pyruvic Acid	-	-	-	-	-	-
L-Galactonic Acid-γ-Lactone	-	-	-	-	-	-
D-Galacturonic Acid	-	-	-	-	-	-
Phenylethylamine	-	-	-	-	-	-
2-Aminoethanol	-	-	-	-	-	-
Total Positives	49	42	26	46	43	23

Table 9: Summary of experimental nitrogen utilization data

Substrate	J2-031	JO161	J2-064	R2-502	F2365	ScottA
Ammonia	-	-	-	-	-	-
Nitrite	-	-	-	-	-	-
Nitrate	-	-	-	-	-	-
Urea	-	-	-	-	-	-
Biuret	-	-	-	-	-	-
L-Alanine	-	-	-	-	-	-
L-Arginine	-	-	-	-	-	-
L-Asparagine	-	-	-	-	-	-
L-Aspartate	-	-	-	-	-	-
L-Cysteine	-	-	-	-	-	-
L-Glutamate	-	-	-	-	-	-
L-Glutamine	-	-	-	-	-	-
Glycine	-	-	-	-	-	-
L-Histidine	+	+	-	-	+	+
L-Isoleucine	-	-	-	-	-	-
L-Leucine	-	-	-	-	-	-
L-Lysine	-	-	-	-	-	-
L-Methionine	-	-	-	-	-	-
L-Phenylalanine	-	-	-	-	-	-
L-Proline	-	-	-	-	-	-
L-Serine	-	-	-	-	-	-
L-Threonine	-	-	-	-	-	+
L-Tryptophan	+	+	+	+	+	-
L-Tyrosine	+	+	+	+	+	+
L-Valine	+	+	+	+	+	+
D-Alanine	-	-	-	-	-	-
D-Asparagine	+	-	+	-	-	+
D-Aspartic Acid	-	-	-	-	-	-
D-Glutamate	+	+	+	+	+	+
D-Lysine	+	+	+	+	+	+
D-Serine	-	-	-	-	-	-
D-Valine	-	-	+	+	-	-
Citrulline	+	+	+	+	+	+
L-Homoserine	-	-	-	-	-	-
Ornithine	-	-	-	-	-	-

N-Acetyl-L-glutamate	-	-	-	-	-	-
N-Phthaloyl-L-Glutamic Acid	+	+	+	+	+	+
L-Pyroglutamic Acid	-	-	-	-	-	-
Hydroxylamine	-	-	-	-	-	-
Methylamine	+	+	+	+	+	+
N-Amylamine	-	-	-	-	-	-
N-Butylamine	-	-	-	-	-	-
Ethylamine	-	-	-	-	-	-
Aminoethanol	-	-	-	-	-	-
Ethylenediamine	-	-	-	-	-	-
Putrescine	-	-	-	-	-	-
Agmatine	-	-	-	-	-	-
Histamine	-	-	-	-	-	-
β-Phenylethylamine	-	-	-	-	-	-
Tyramine	-	-	-	-	-	-
Acetamide	-	-	-	-	-	-
Formamide	-	-	-	-	-	-
Glucuronamide	-	-	-	-	-	-
D,L-Lactamide	-	-	-	-	-	-
D-Glucosamine	-	-	-	-	-	-
D-Galactosamine	-	-	-	-	-	-
D-Mannosamine	-	-	-	-	-	-
N-Acetyl-D-Glucosamine	-	-	-	-	-	-
N-Acetyl-D-Galactosamine	-	-	-	-	-	-
N-Acetyl-D-Mannosamine	-	-	-	-	-	-
Adenine	-	-	-	-	-	-
Adenosine	-	-	-	-	-	-
Cytidine	-	-	-	-	-	-
Cytosine	-	-	-	-	-	-
Guanine	-	-	-	-	-	-
Guanosine	-	-	-	-	-	-
Thymine	-	-	-	-	-	-
Thymidine	-	-	-	-	-	-
Uracil	-	-	-	-	-	-
Uridine	-	-	-	-	-	-

Inosine	+	+	+	-	-	+
Xanthine	-	-	-	-	-	-
Xanthosine	-	-	-	-	-	-
Uric Acid	-	-	-	-	-	-
Alloxan	-	-	-	-	-	-
Allantoin	-	-	-	-	-	-
Parabanic Acid	-	-	-	-	-	-
D,L-α-Amino-N-Butyric Acid	-	-	-	-	-	-
γ-Amino-N-Caproic Acid	-	-	-	-	-	-
ε-Amino-N-Caproic Acid	-	-	-	-	-	-
D,L-α-Amino-Caprylic Acid	-	-	-	-	-	-
δ-Amino-N-Valeric Acid	-	-	-	-	-	-
α-Amino-N-Valeric Acid	-	-	-	-	-	-
L-Alanyl-L-Aspartic Acid	-	-	-	-	-	-
Alanylglutamine	-	-	-	-	-	-
L-Alanyl-L-Glutamate	-	-	-	-	-	-
L-Alanylglycine	-	-	-	-	-	-
Alanylhistidine	-	-	-	-	-	-
Alanylleucine	-	-	-	-	-	-
L-Alanyl-L-Threonine	-	-	-	-	-	-
Glycyl-L-Asparagine	-	-	-	-	-	-
Glycylglutamine	-	-	-	-	-	-
Glycyl-L-Glutamate	-	-	-	-	-	-
Glycylmethionine	-	-	-	-	-	-
L-Methionyl-L-Alanine	-	-	-	-	-	-
Total Positives	11	10	11	9	9	11

Table 10: Summary of experimental phosphorous utilization data

Substrate	J2-031	JO161	J2-064	R2-502	F2365	ScottA
Phosphate	-	-	-	-	-	-
Pyrophosphate	-	-	-	-	-	-
Trimetaphosphate	-	-	-	-	-	-

Triphosphate	-	-	-	-	-	-
Triethyl Phosphate	-	-	-	-	-	-
Hypophosphite	-	-	-	-	-	-
Adenosine-2'-monophosphate	-	-	-	-	-	-
Adenosine-3'-monophosphate	-	-	-	-	-	-
Adenosine-5'-Monophosphate	-	-	-	-	-	-
Adenosine-2',3'-cyclic monophosphate	-	-	-	-	-	-
Adenosine-3',5'-cyclic monophosphate	-	-	-	-	-	-
Thiophosphate	-	-	-	-	-	-
Dithiophosphate	-	-	-	-	-	-
Glycerol-3-phosphate	-	-	-	-	-	-
β-Glycerol Phosphate	-	-	-	-	-	-
Carbamoylphosphate	-	-	-	-	-	-
2-Phospho-D-glycerate	-	-	-	-	-	-
3-Phosphoglycerate	-	-	-	-	-	-
Guanosine-2'-monophosphate	-	-	-	-	-	-
Guanosine-3'-monophosphate	-	-	-	-	-	-
Guanosine-5'-Monophosphate	-	-	-	-	-	-
Guanosine-2',3'-cyclic monophosphate	-	-	-	-	-	-
Guanosine-3',5'-cyclic monophosphate	-	-	-	-	-	-
Phosphoenolpyruvate	-	-	-	-	-	-
Phospho-Glycolic Acid	-	-	-	-	-	-
Glucose-1-phosphate	-	-	-	-	-	-
D-Glucose-6-Phosphate	-	-	-	-	-	-
2-Deoxy-D-Glucose-6-Phosphate	-	-	-	-	-	-
D-Glucosamine phosphate	-	-	-	-	-	-
6-Phospho-D-gluconate	-	-	-	-	-	-
Cytidine-2'-monophosphate	-	-	-	-	-	-

Cytidine-3'-monophosphate	-	-	-	-	-	-
Cytidine-5'-Monophosphate	-	-	-	-	-	-
Cytidine-2',3'-cyclic monophosphate	-	-	-	-	-	-
Cytidine-3',5'-cyclic monophosphate	-	-	-	-	-	-
D-Mannose1-phosphate	-	-	-	-	-	-
D-Mannose-6-Phosphate	-	-	-	-	-	-
Cysteamine-S-Phosphate	-	-	-	-	-	-
Phospho-L-Arginine	-	-	-	-	-	-
D-O-Phosphoserine	+	+	+	-	+	+
Phosphoserine	-	-	-	-	-	-
L-Threonine phosphate	-	-	-	-	-	-
Uridine-2'-monophosphate	-	-	-	-	-	-
Uridine-3'-monophosphate	-	-	-	-	-	-
Uridine-5'-Monophosphate	-	-	-	-	-	-
Uridine-2',3'-cyclic monophosphate	-	-	-	-	-	-
Uridine-3',5'-cyclic monophosphate	-	-	-	-	-	-
O-Phospho-D-Tyrosine	-	-	-	-	-	-
O-Phospho-L-Tyrosine	-	-	-	-	-	-
Phosphocreatine	+	-	+	+	+	+
Phosphoryl Choline	-	-	-	-	-	-
O-Phosphoryl-Ethanolamine	-	-	-	-	-	-
Acetylphosphate	-	-	-	-	-	-
2-Aminoethyl Phosphonic Acid	-	-	-	-	-	-
Methylene Diphosphonic Acid	-	-	-	-	-	-
Thymidine-3'-monophosphate	-	-	-	-	-	-
Thymidine-5'-monophosphate	-	-	-	-	-	-

Inositol Hexaphosphate	-	-	-	-	-	-
Thymidine-3',5'-cyclic monophosphate	+	+	+	+	+	+
Total Positives	3	2	3	2	3	3

Table 11: Summary of experimental sulfur utilization data

Substrate	**J2-031**	**JO161**	**J2-064**	**R2-502**	**F2365**	**ScottA**
Sulfate	-	-	-	-	-	-
Thiosulfate	-	-	+	+	-	-
Tetrathionate	-	-	+	+	-	-
Thiophosphate	-	-	-	-	-	-
Dithiophosphate	-	-	-	-	-	-
L-Cysteine	-	-	-	-	-	-
D-Cysteine	-	-	-	-	-	-
L-Cysteinyl-Glycine	-	-	-	-	-	-
L-Cysteic Acid	-	-	-	-	-	-
Cysteamine	-	-	-	-	-	-
L-Cysteine Sulfinic Acid	-	-	-	-	-	-
N-Acetyl-L-Cysteine	-	-	-	-	-	-
S-Methyl-L-Cysteine	-	-	-	-	-	-
Cystathionine	+	+	+	+	+	+
Lanthionine	-	-	-	-	-	-
Glutathione	-	-	-	-	-	-
D,L-Ethionine	-	-	-	-	-	-
L-Methionine	-	-	-	-	-	-
D-Methionine	-	-	-	-	-	-
Glycyl-L-Methionine	-	-	-	-	-	-
N-Acetyl-D,L-Methionine	-	-	-	-	-	-
L-Methionine S-oxide	-	-	-	-	-	-
L-Methionine Sulfone	-	-	-	-	-	-
L-Djenkolic Acid	-	-	-	-	-	-
Thiourea	-	-	-	-	-	-
1-Thio-beta-D-Glucose	+	+	+	+	+	+
Lipoamide	-	-	-	-	-	-
Taurocholic Acid	-	-	-	-	-	-

Taurine	-	-	-	-	-	-
Hypotaurine	-	-	-	-	-	-
p-Amino Benzene Sulfonic Acid	-	-	-	-	-	-
Butane Sulfonic Acid	-	-	-	-	-	-
2-Hydroxyethane Sulfonic Acid	-	-	-	+	-	-
Methane Sulfonic Acid	-	-	-	-	-	-
Tetramethylene Sulfone	-	-	-	-	-	-
Total Positives	2	2	4	5	2	2

Table 12: Number of strain and serovar specific nutrients

	Carbon Sources	Nitrogen Sources	Sulfur + Phosphorus Sources
All strains	11	7	3
No strains	24	82	87
Unique to J2-031	3	0	0
Unique to JO161	4	0	0
Unique to J2-064	1	0	0
Unique to R2-502	3	0	1
Unique to F2365	4	0	0
Unique to ScottA	0	1	0
Unique to 1/2b	1	1	2

Table 13: Unique carbon sources metabolized by a single *L. monocytogenes* strain or serovar

Strain/Serovar	Unique Nutrients
J2-031	Formic Acid, D-Aspartic Acid, M-Tartaric Acid
JO161	D-Melibiose, D-Threonine, Glyoxylic Acid, L-Serine
J2-064	Glycyl-L-Proline
R2-502	Adonitol, M-Inositol, L-Lyxose
F2365	Sucrose, α-Hydroxy Glutaric Acid-γ-Lactone, Fumaric Acid, Tyramine
1/2b	p-Hydroxy Phenyl Acetic Acid

Table 14: Unique nitrogen sources metabolized by a single *L. monocytogenes* strain or serovar

Strain/Serovar	Unique Compounds
ScottA	L-Threonine
1/2b	D-Valine

Table 15: Unique sulfur/phosphorus sources metabolized by a single *L. monocytogenes* strain or serovar

Strain/Serovar	Unique Compounds
R2-502	2-Hydroxyethane Sulfonic Acid
1/2b	Thiosulfate, Tetrathionate

2.4.4 Comparison Between *In Silico* Predictions and Experimental Nutrient Utilization Results

The previously described nutrient phenotype data can be compared to *in silico* predictions as part of the qualitative validation process. These comparisons allow the reconciliation of inaccuracies of *in silico* predictions, which results in optimized, validated GEMs. When comparing *in silico* predictions to experimental results for utilization of nutrients, there can be two types of disagreement: either the model predicts growth, but experimentally there was none (termed "false positive"), or the model predicts no growth, but experimentally there was growth (termed "false negative").

A comparison of agreement between draft and final models for carbon source utilization shows a drastic improvement in agreement following gapfilling and manual curation (Figure 5). The draft model contains the initial nutrient utilization predictions, while the final model contains predictions generated following manual curation and gapfilling. 57 carbon sources found in the PM1 plates were also present in the GEMs, which means comparison between *in silico* predictions and experimental data was possible. 18 of these carbon sources resulted in agreement between *in silico* predictions and experimental results for all strains (Table 16). It was found that draft GEMs agreements ranged from 48.3% to 69.0% (Table 17). For the final GEMs, agreements ranged from 80.7% to 91.2%, and each strain had a minimum of four false negatives and one false positive (Table 17).

After manual curation and gapfilling to improve the agreement between carbon utilization experimental data and the corresponding *in silico* predictions, a similar comparison was made for nitrogen (Figure 6). 62 nitrogen sources found in the PM3 plates were also present in the models, allowing comparisons to be made. It was found that agreement ranged from 59.7% to 66.1%, and all strains had at least five false negatives and at least 15 false positives (Table 18). In contrast to carbon, GEMs were not manually curated to improve agreement for nitrogen sources.

Comparisons between *in silico* predictions and experimental results for phosphorous and sulfur utilization were made in conjunction with the nitrogen comparisons (Figures 7 and 8). For phosphorous, 22 compounds enabled comparisons. For these 22 compounds, agreement ranged from 18.2% to 27.3%, and each strain had at least 16 false positives. However, none of the strains had any false negatives (Table 19). For sulfur 11 compounds were found in both the PM4 plates and the GEMs. For these 11 compounds, agreement ranged from 72.7% to 81.8%, and each strain had at least one false negative and one false positive (Table 20).

Table 16: Carbon sources with agreement between *in silico* predictions and experimental results for all strains

Nutrients
N-Acetyl-D-Glucosamine, L-Proline, D-Trehalose, D-Mannose, D-Serine, D-Gluconic Acid, D-Mannitol, L-Glutamate, D-Fructose, D-Glucose, Uridine, L-Glutamine, Amylotriose, Adenosine, Glycyl-L-Aspartate, D-Cellobiose, Glycyl-L-Glutamate, Glycyl-L-Proline

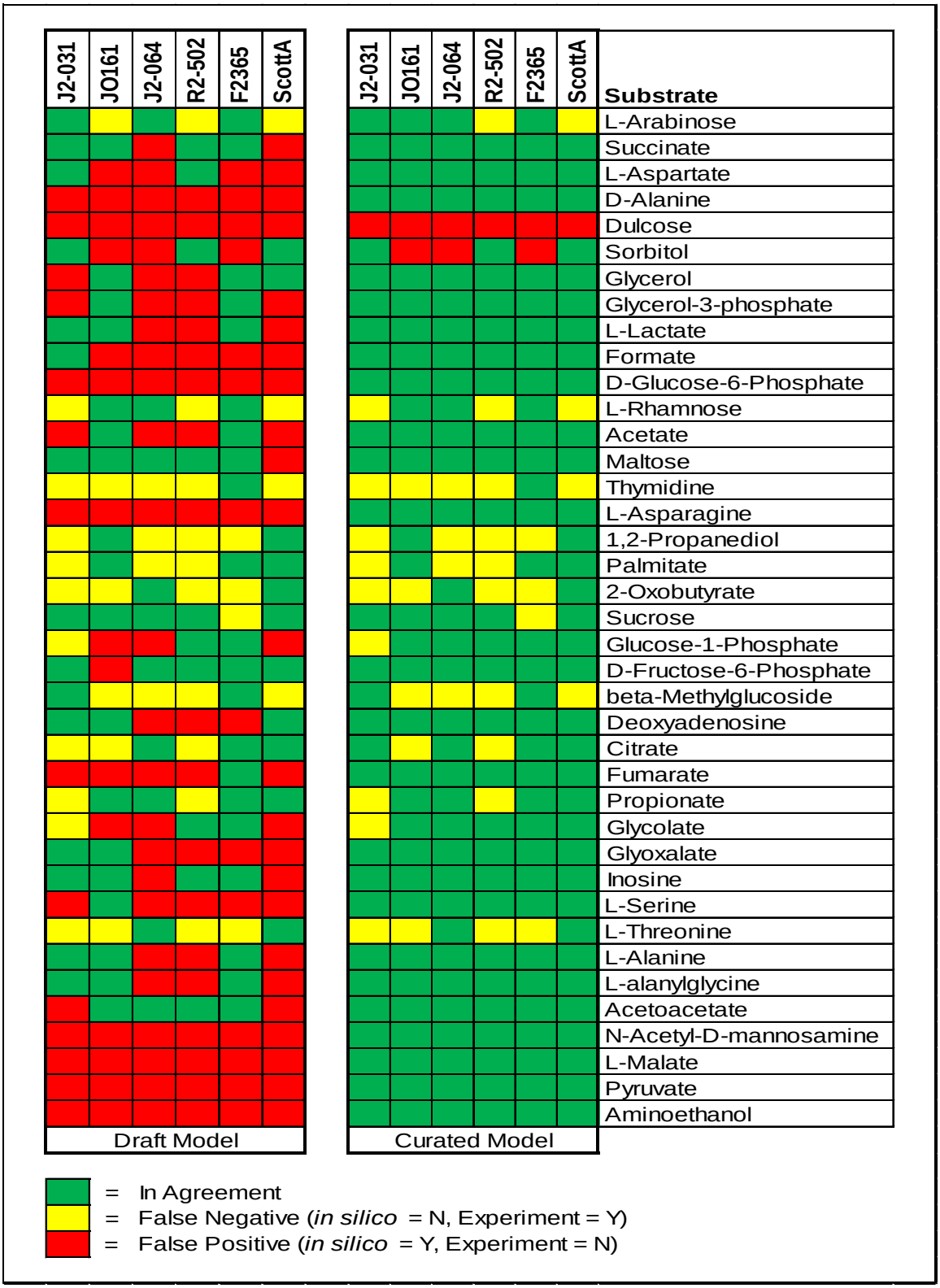

Figure 5: Comparison of *in silico* predictions to experimental results for 39 individual carbon sources with at least one disagreement

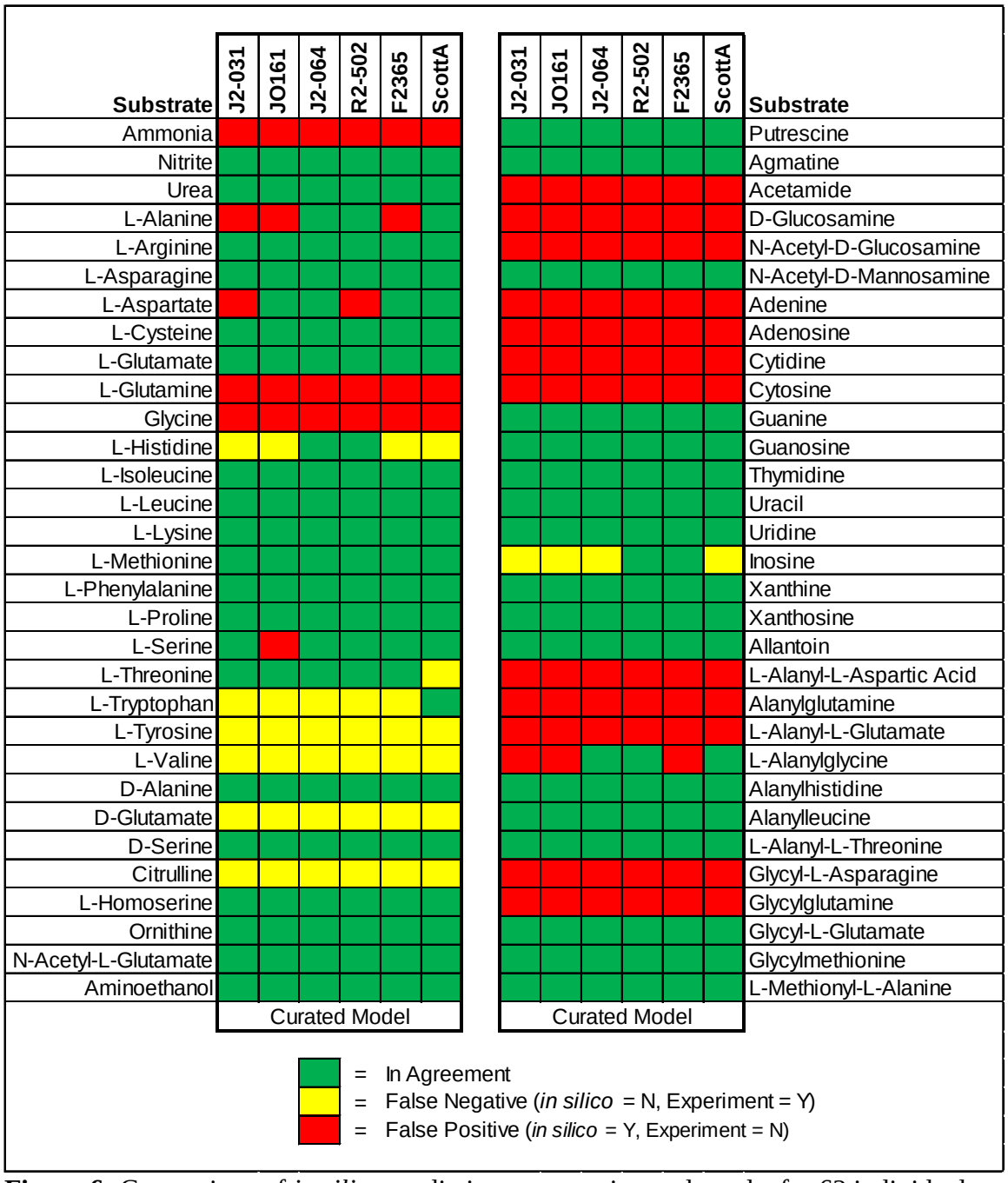

Figure 6: Comparison of *in silico* predictions to experimental results for 62 individual nitrogen sources

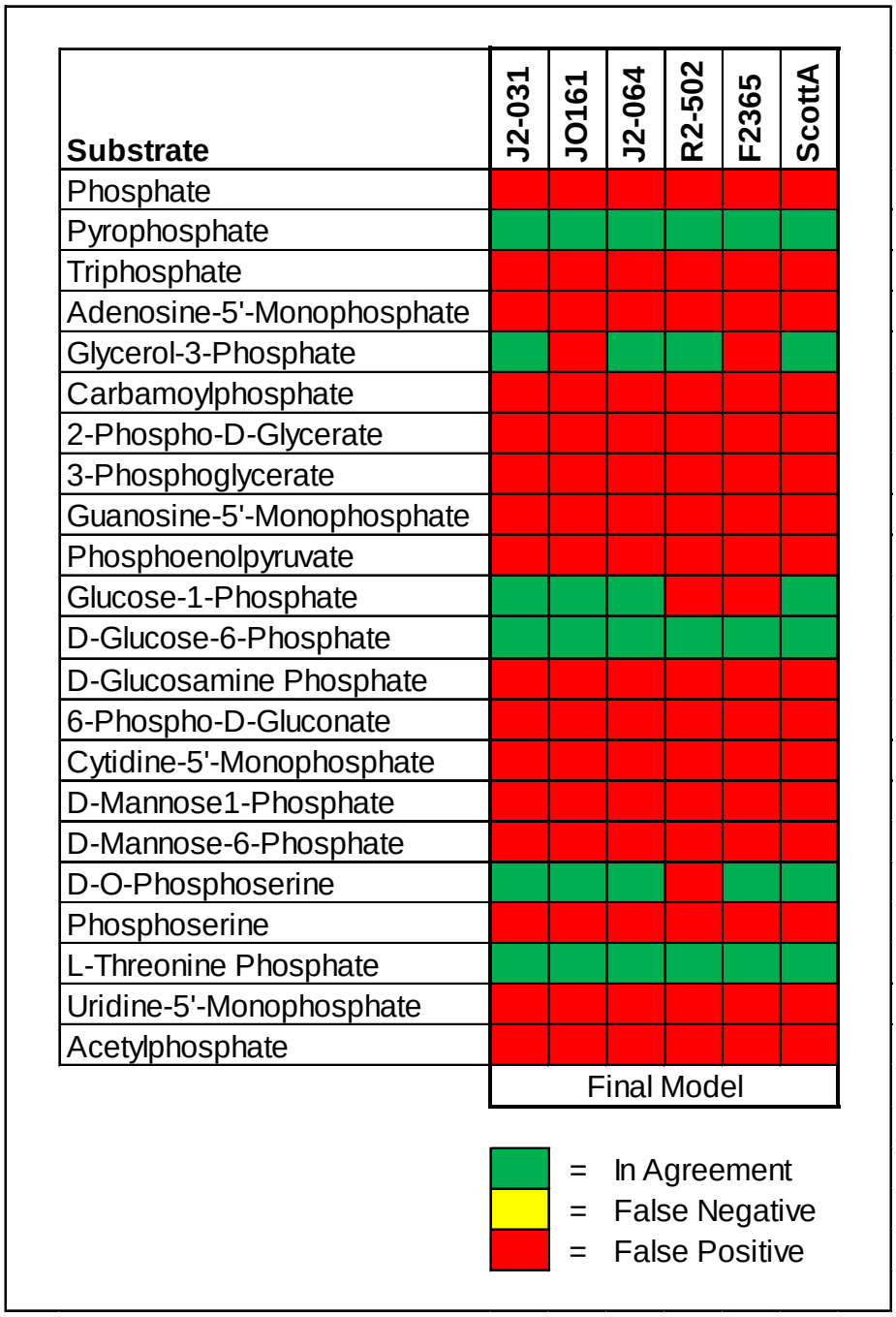

Figure 7: Comparison of *in silico* predictions to experimental results for 22 phosphorus sources

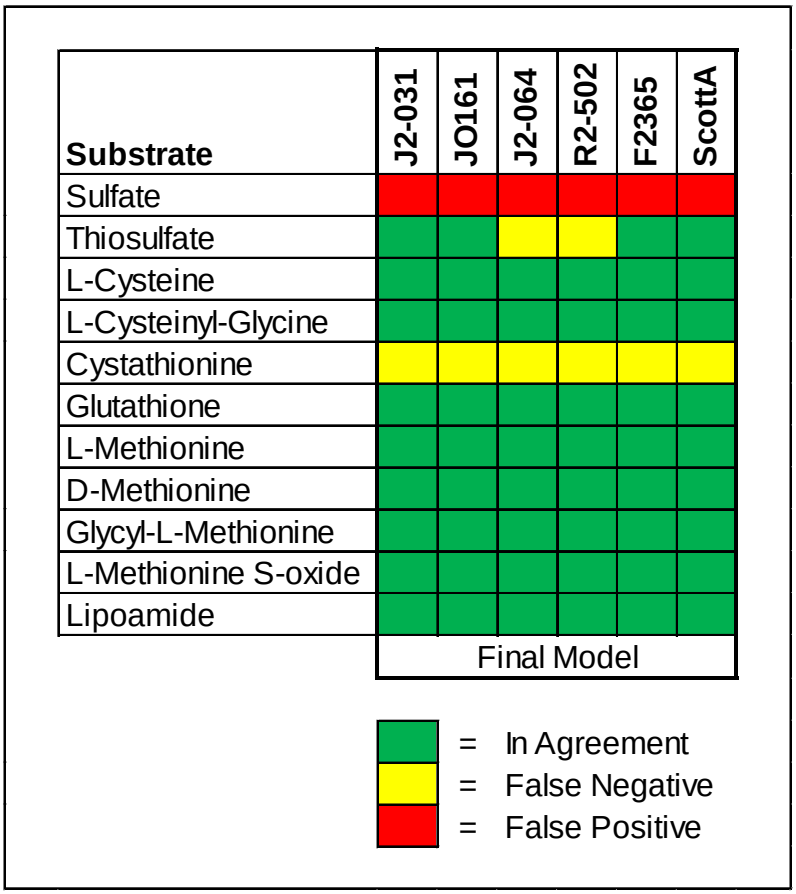

Figure 8: Comparison of *in silico* predictions to experimental results for 11 sulfur sources

Table 17: Numerical agreement summary for carbon source utilization, false negative and false positive values correspond to final models

Strain	Initial Agreement	Final Agreement	False Negatives	False Positives
J2-031	69.0%	82.5%	9	1
JO161	58.6%	87.7%	5	2
J2-064	48.3%	89.5%	4	2
R2-502	63.8%	80.7%	10	1
F2365	50.0%	89.5%	4	2
ScottA	51.7%	91.2%	4	1

Table 18: Numerical agreement summary for nitrogen source utilization

Strain	Agreement	False Negatives	False Positives
J2-031	59.7%	7	18
JO161	59.7%	7	18
J2-064	66.1%	6	15
R2-502	66.1%	5	16
F2365	62.9%	6	17
ScottA	64.5%	7	15

Table 19: Numerical agreement summary for phosphorus source utilization

Strain	Agreement	False Negatives	False Positives
J2-031	27.3%	0	16
JO161	22.7%	0	17
J2-064	27.3%	0	16
R2-502	18.2%	0	18
F2365	18.2%	0	18
ScottA	27.3%	0	16

Table 20: Numerical agreement summary for sulfur source utilization

Strain	Agreement	False Negatives	False Positives
J2-031	81.8%	1	1
JO161	81.8%	1	1
J2-064	72.7%	2	1
R2-502	72.7%	2	1
F2365	81.8%	1	1
ScottA	81.8%	1	1

2.4.5 Agreement Comparison to Other Genome-Scale Metabolic Models

It is worthwhile to compare the GEMs generated by this work to those previously created for other organisms in order to assess the validity of the semi-automated model creation approach. The carbon agreement of the six newly created *L. monocytogenes* GEMs was compared to the carbon agreement of 16 previously published GEMs (Figure 9), nitrogen agreement was compared to six previously published GEMs (Figure 10), and phosphorus and sulfur agreements were compared to four previously published GEMs (Figures 11 and 12) [25, 27, 49-56].

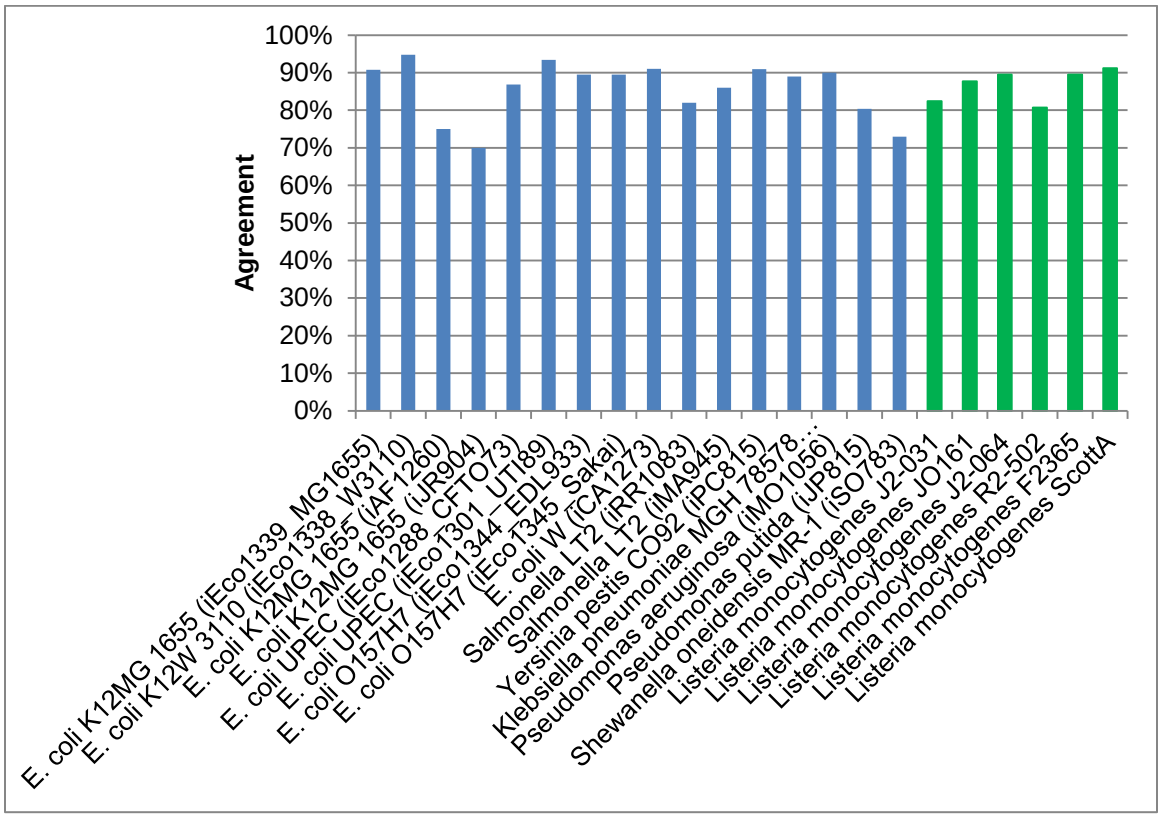

Figure 9: Comparison of carbon source utilization agreement between genome-scale metabolic models created in this study and previous genome-scale metabolic models

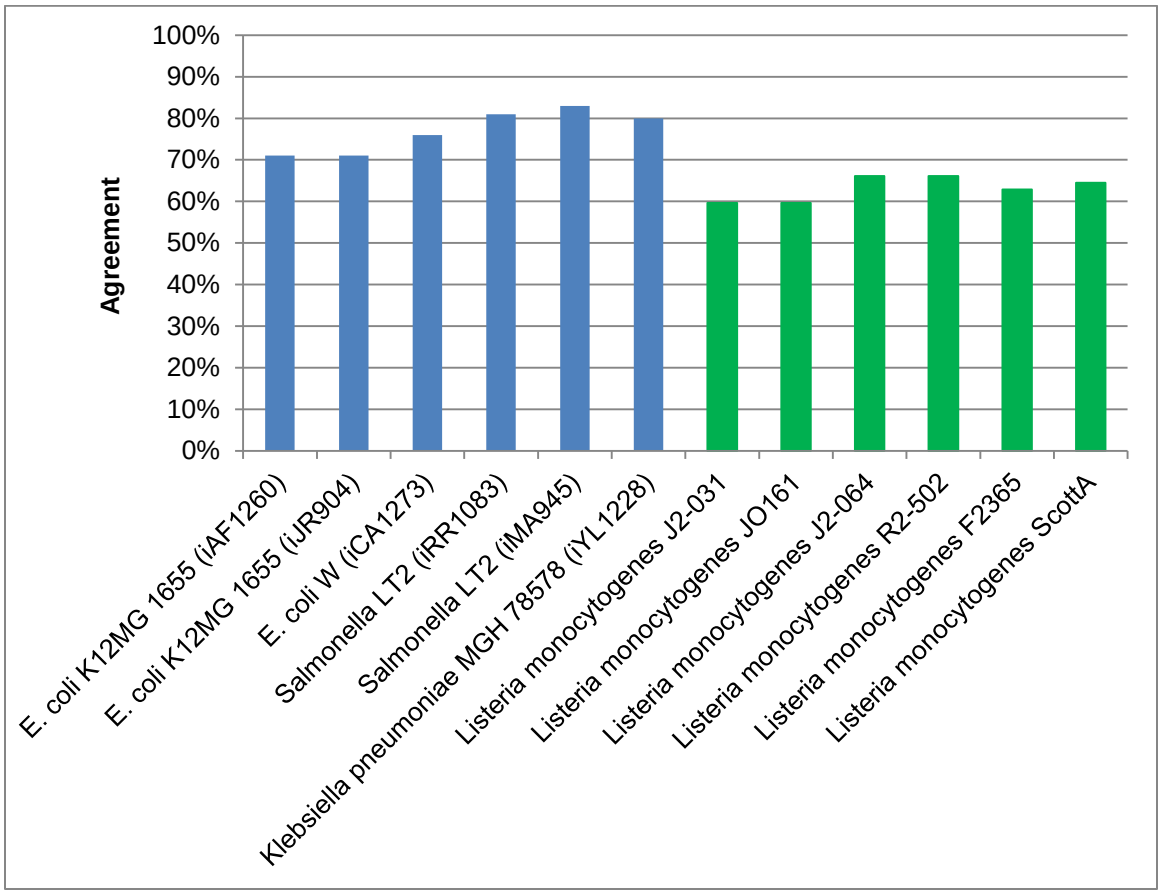

Figure 10: Comparison of nitrogen source utilization agreement between genome-scale metabolic models created in this study and previous genome-scale metabolic models

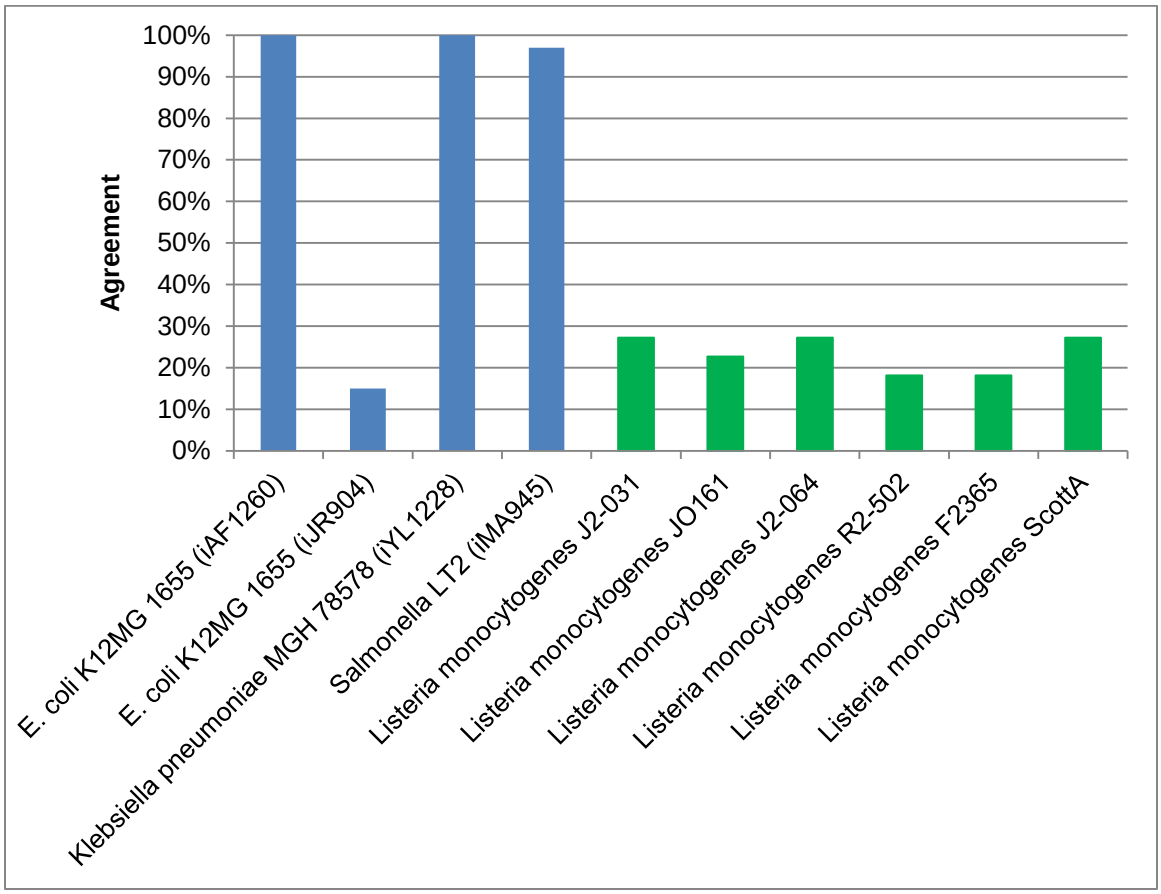

Figure 11: Comparison of phosphorus source utilization agreement between genome-scale metabolic models created in this study and previous genome-scale metabolic models

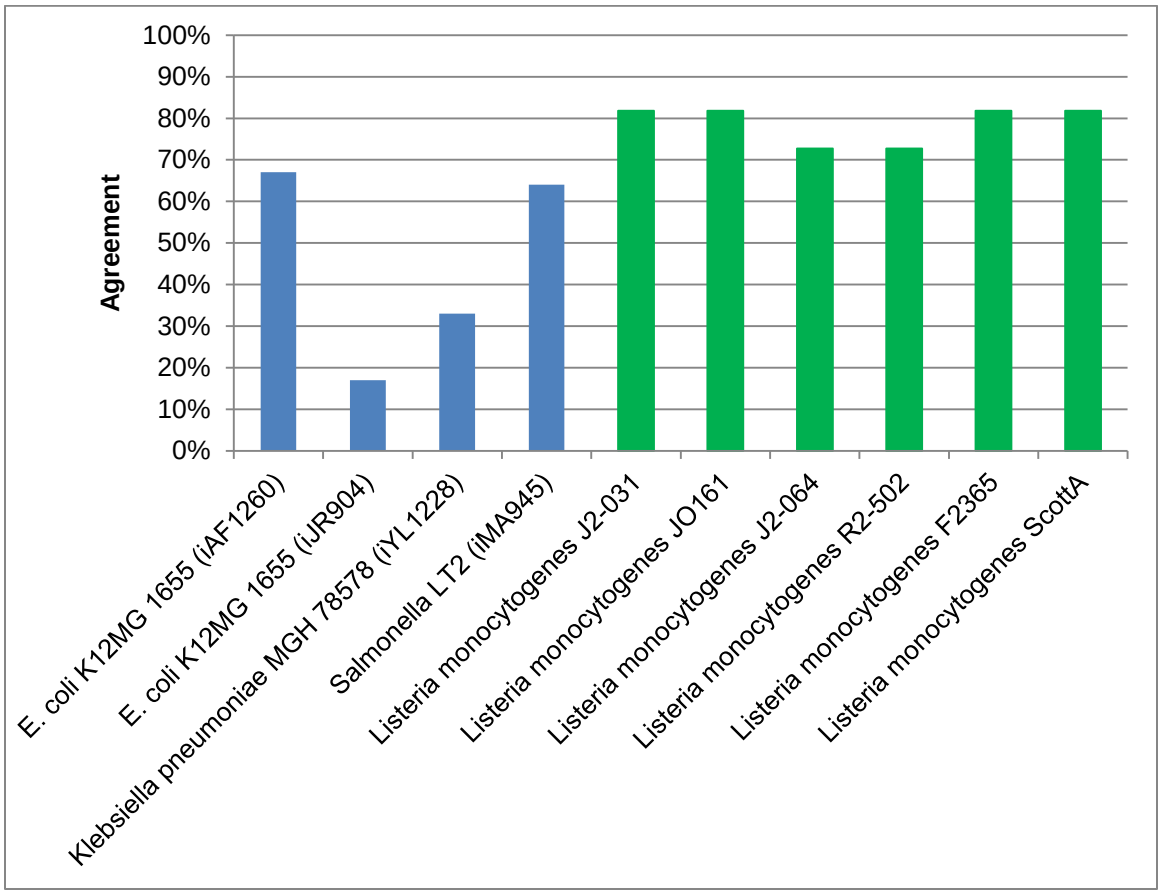

Figure 12: Comparison of sulfur source utilization agreement between genome-scale metabolic models created in this study and previous genome-scale metabolic models

2.4.6 Essential Reactions

Finally, the number of reactions essential to the survivability and viability of *L. monocytogenes* were predicted in simulations reflecting glucose minimal media as well as five high risk foods (Tables 21 – 26). The five high risk foods simulated were queso fresco, chicken breast, smoked salmon, cantaloupe, and romaine lettuce. There were approximately 370 predicted essential reactions when glucose minimal media was simulated, and this number decreased to approximately 300 in the high risk food simulations. Additionally, at least one unique essential reaction was predicted and identified in every simulation for *L. monocytogenes* strains J2-031, J2-064, and F2365, as well as serovar 1/2a (Tables 27 – 32).

Table 21: Essential reaction summary for the simulation of conditions representing glucose minimal media

Strain	Total (% Shared)	Unique (%)	Serovar	Serovar Specific Reactions
J2-031	369 (95.1%)	7 (1.90%)	1/2a	3
JO161	371 (94.6%)	0 (0%)		
J2-064	368 (95.4%)	2 (0.54%)	1/2b	0
R2-502	368 (95.4%)	0 (0%)		
F2365	367 (94.6%)	1 (0.27%)	4b	0
ScottA	367 (95.6%)	0 (0%)		

Table 22: Essential reaction summary for the simulation of conditions representing queso fresco

Strain	Total (% Shared)	Unique (%)	Serovar	Serovar Specific Reactions
J2-031	297 (93.6%)	9 (3.03%)	1/2a	1
JO161	295 (94.2%)	0 (0%)		
J2-064	306 (90.8%)	7 (2.29%)	1/2b	0
R2-502	294 (94.6%)	0 (0%)		
F2365	293 (94.9%)	1 (0.34%)	4b	0
ScottA	293 (94.9%)	0 (0%)		

Table 23: Essential reaction summary for the simulation of conditions representing chicken breast

Strain	Total (% Shared)	Unique (%)	Serovar	Serovar Specific Reactions
J2-031	297 (94.3%)	7 (2.36%)	1/2a	1
JO161	298 (94.0%)	0 (0%)		
J2-064	309 (90.6%)	7 (2.27%)	1/2b	0
R2-502	297 (94.3%)	0 (0%)		
F2365	296 (94.6%)	1 (0.34%)	4b	0
ScottA	296 (94.6%)	0 (0%)		

Table 24: Essential reaction summary for the simulation of conditions representing smoked salmon

Strain	Total (% Shared)	Unique (%)	Serovar	Serovar Specific Reactions
J2-031	297 (94.3%)	7 (2.36%)	1/2a	1
JO161	298 (94.0%)	0 (0%)		
J2-064	309 (90.6%)	7 (2.27%)	1/2b	0
R2-502	297 (94.3%)	0 (0%)		
F2365	296 (94.6%)	1 (0.34%)	4b	0
ScottA	296 (94.6%)	0 (0%)		

Table 25: Essential reaction summary for the simulation of conditions representing cantaloupe

Strain	Total (% Shared)	Unique (%)	Serovar	Serovar Specific Reactions
J2-031	295 (94.9%)	5 (1.69%)	1/2a	1
JO161	298 (94.0%)	0 (0%)		
J2-064	309 (90.6%)	7 (2.27%)	1/2b	0
R2-502	297 (94.3%)	0 (0%)		
F2365	296 (94.6%)	1 (0.34%)	4b	0
ScottA	296 (94.6%)	0 (0%)		

Table 26: Essential reactions summary for the simulation of conditions representing romaine lettuce

Strain	Total (% Shared)	Unique (%)	Serovar	Serovar Specific Reactions
J2-031	297 (94.3%)	7 (2.36%)	1/2a	1
JO161	298 (94.0%)	0 (0%)		
J2-064	309 (90.6%)	7 (2.27%)	1/2b	0
R2-502	297 (94.3%)	0 (0%)		
F2365	296 (94.6%)	1 (0.34%)	4b	0
ScottA	296 (94.6%)	0 (0%)		

Table 27: Enzymes catalyzing strain and serovar-specific essential reactions in simulations reflecting glucose minimal media

Strain/Serovar	Unique Enzymes
J2-031	S-Adenosyl-L-homocysteine hydrolase, Isocitrate glyoxylate-lyase, alpha-D-Glucose-1-phosphate:alpha-D-glucose-1-phosphate, Xanthosine-5'-phosphate:L-glutamine amido-ligase (AMP-forming), Maltose:orthophosphate 1-beta-D-glucosyltransferase, Adenosine:orthophosphate ribosyltransferase, beta-D-Glucose 1-phosphate 1,6-phosphomutase
J2-064	4-hydroxybenzaldehyde:NAD+ oxidoreductase, 4-hydroxy-benzyl-alcohol dehydrogenase
F2365	Adenosyl cobinamide kinase
1/2a	L-threonine ammonia-lyase, O-Acetyl-L-homoserine acetate-lyase (adding methanethiol), 2-Amino-4-hydroxy-6-hydroxymethyl-7,8-dihydropteridine:4

Table 28: Enzymes catalyzing strain and serovar-specific essential reactions in simulations reflecting queso fresco

Strain/Serovar	Unique Enzymes
J2-031	S-Adenosyl-L-homocysteine hydrolase, Oxalosuccinate:NADP+ oxidoreductase (decarboxylating), Isocitrate glyoxylate-lyase, alpha-D-Glucose-1-phosphate:alpha-D-glucose-1-phosphate, Xanthosine-5'-phosphate:L-glutamine amido-ligase (AMP-forming), Maltose:orthophosphate 1-beta-D-glucosyltransferase, Adenosine:orthophosphate ribosyltransferase, Isocitrate:NADP+ oxidoreductase (decarboxylating), beta-D-Glucose 1-phosphate 1,6-phosphomutase
J2-064	4-hydroxybenzaldehyde:NAD+ oxidoreductase, ATP:thiamine phosphotransferase, 2-Methyl-4-amino-5-hydroxymethylpyrimidine-diphosphate:4-methyl-5-, ATP:4-amino-2-methyl-5-phosphomethylpyrimidine phosphotransferase, 4-amino-2-methyl-5-phosphomethylpyrimidine synthetase, thiazole phosphate synthesis, 4-hydroxy-benzyl-alcohol dehydrogenase
F2365	Adenosyl cobinamide kinase
1/2a	2-Amino-4-hydroxy-6-hydroxymethyl-7,8-dihydropteridine:4

Table 29: Enzymes catalyzing strain and serovar-specific essential reactions in simulations reflecting chicken breast

Strain/Serovar	Unique Enzymes
J2-031	S-Adenosyl-L-homocysteine hydrolase, Isocitrate glyoxylate-lyase, alpha-D-Glucose-1-phosphate:alpha-D-glucose-1-phosphate, Xanthosine-5'-phosphate:L-glutamine amido-ligase (AMP-forming), Maltose:orthophosphate 1-beta-D-glucosyltransferase, Adenosine:orthophosphate ribosyltransferase, beta-D-Glucose 1-phosphate 1,6-phosphomutase
J2-064	4-hydroxybenzaldehyde:NAD+ oxidoreductase, ATP:thiamine phosphotransferase, 2-Methyl-4-amino-5-hydroxymethylpyrimidine-diphosphate:4-methyl-5-, ATP:4-amino-2-methyl-5-phosphomethylpyrimidine phosphotransferase, 4-amino-2-methyl-5-phosphomethylpyrimidine synthetase, thiazole phosphate synthesis, 4-hydroxy-benzyl-alcohol dehydrogenase
F2365	Adenosyl cobinamide kinase
1/2a	2-Amino-4-hydroxy-6-hydroxymethyl-7,8-dihydropteridine:4

Table 30: Enzymes catalyzing strain and serovar-specific essential reactions in simulations reflecting smoked salmon

Strain/Serovar	Unique Enzymes
J2-031	S-Adenosyl-L-homocysteine hydrolase, Isocitrate glyoxylate-lyase, alpha-D-Glucose-1-phosphate:alpha-D-glucose-1-phosphate, Xanthosine-5'-phosphate:L-glutamine amido-ligase (AMP-forming), Maltose:orthophosphate 1-beta-D-glucosyltransferase, Adenosine:orthophosphate ribosyltransferase, beta-D-Glucose 1-phosphate 1,6-phosphomutase
J2-064	4-hydroxybenzaldehyde:NAD+ oxidoreductase, ATP:thiamine phosphotransferase, 2-Methyl-4-amino-5-hydroxymethylpyrimidine-diphosphate:4-methyl-5-, ATP:4-amino-2-methyl-5-phosphomethylpyrimidine phosphotransferase, 4-amino-2-methyl-5-phosphomethylpyrimidine synthetase, thiazole phosphate synthesis, 4-hydroxy-benzyl-alcohol dehydrogenase
F2365	Adenosyl cobinamide kinase
1/2a	2-Amino-4-hydroxy-6-hydroxymethyl-7,8-dihydropteridine:4

Table 31: Enzymes catalyzing strain and serovar-specific essential reactions in simulations reflecting cantaloupe

Strain/Serovar	Unique Enzymes
J2-031	S-Adenosyl-L-homocysteine hydrolase, Isocitrate glyoxylate-lyase, alpha-D-Glucose-1-phosphate:alpha-D-glucose-1-phosphate, Xanthosine-5'-phosphate:L-glutamine amido-ligase (AMP-forming), Adenosine:orthophosphate ribosyltransferase
J2-064	4-hydroxybenzaldehyde:NAD+ oxidoreductase, ATP:thiamine phosphotransferase, 2-Methyl-4-amino-5-hydroxymethylpyrimidine-diphosphate:4-methyl-5-, ATP:4-amino-2-methyl-5-phosphomethylpyrimidine phosphotransferase, 4-amino-2-methyl-5-phosphomethylpyrimidine synthetase, thiazole phosphate synthesis, 4-hydroxy-benzyl-alcohol dehydrogenase
F2365	Adenosyl cobinamide kinase
1/2a	2-Amino-4-hydroxy-6-hydroxymethyl-7,8-dihydropteridine:4

Table 32: Enzymes catalyzing strain and serovar-specific essential reactions in simulations reflecting romaine lettuce

Strain/Serovar	Unique Enzymes
J2-031	S-Adenosyl-L-homocysteine hydrolase, Isocitrate glyoxylate-lyase, alpha-D-Glucose-1-phosphate:alpha-D-glucose-1-phosphate, Xanthosine-5'-phosphate:L-glutamine amido-ligase (AMP-forming), Maltose:orthophosphate 1-beta-D-glucosyltransferase, Adenosine:orthophosphate ribosyltransferase, beta-D-Glucose 1-phosphate 1,6-phosphomutase
J2-064	4-hydroxybenzaldehyde:NAD+ oxidoreductase, ATP:thiamine phosphotransferase, 2-Methyl-4-amino-5-hydroxymethylpyrimidine-diphosphate:4-methyl-5-, ATP:4-amino-2-methyl-5-phosphomethylpyrimidine phosphotransferase, 4-amino-2-methyl-5-phosphomethylpyrimidine synthetase, thiazole phosphate synthesis, 4-hydroxy-benzyl-alcohol dehydrogenase
F2365	Adenosyl cobinamide kinase
1/2a	2-Amino-4-hydroxy-6-hydroxymethyl-7,8-dihydropteridine:4

2.5 Discussion

2.5.1 Genomic Differences

Genome alignment tools, such as Mauve, can be extremely useful in identifying genomic differences between strains of the same organism. As can be seen in the Mauve

genome alignments, the genomes of *L. monocytogenes* are highly conserved between strains (Figures 3 and 4). This is represented in the default view by 10, large LCBs shared by all six strains. This indicates that not only genomic content, but genome organization is also very highly conserved between strains. This conservation is also represented in the "backbone view" by the large amount of the genome given the flat, mauve color, indicating that an overwhelming majority of the genomes is conserved between all six strains. Also in the "backbone" view, it is apparent that there are only a few, small sections of the genome are found in subsets of the strains. However, it is interesting to note that each of the three studied serovars had at least one serovar-specific segment of the genome.

When the content of predicted ORFs is compared, it is clear that much of the genome contents are shared by all six *L. monocytogenes* strains (Table 5). The genomes are all very similar in length, and greater than 85% of each genome is shared between all strains. This degree of conservation was expected, since most evidence suggests that the genomes of modern *L. monocytogenes* strains are highly conserved [19].

Interestingly, the amount of serovar specific genes is on the same scale as the amount of genes unique to each strain. This can be seen easily in Mauve backbone view (Figure 4) because there are very few regions of the genomes that are color coded so that they are only found in the genomes of the two strains of specific serovars. The amount of this two strain color coordination is comparable to the amount of white areas of unique genomic content. This is also represented by the fact that the number of serovar specific genes is similar to the number of unique genes (Table 5). The low frequency of serovar specific genes could possibly indicate that a new method of classification could be more representative of actual differences between strains. With the progression towards faster, cheaper sequencing techniques, perhaps genomic content would be a more specific differentiation practice than the PFGE standard. However, it is beyond the scope of this thesis to speculate on what, if any, method would improve the classification of *L. monocytogenes*.

2.5.2 Genome-Scale Metabolic Model Properties

As GEMs were curated and optimized through comparison to experimental data, the number of genes, metabolites, and metabolic reactions for each of the GEMs at each stage of their development increased (Table 6). The models were very consistent in the number of genes, metabolites, and reactions. The vast majority of the metabolites (99.1%) and reactions (97.4%) were shared between all six of the GEMs. However, three of the strains (*L. monocytogenes* strains J2-031, J2-064, and F2365) had at least one unique reaction, as did serovar 1/2a. These unique reactions may have important implications, because they may give these strains a competitive advantage in specific environments, or may explain the isolation source of the strain.

The level of diversity in the metabolic genes and corresponding content was not too surprising. This is because *L. monocytogenes* is a non-spore forming bacteria, and similar studies with other non-sporulating bacteria that compare different strains of the same organism, such as *E. coli*, saw similar levels of differentiation in metabolic gene/reaction content [25, 26]. In terms of the numbers of these components (genes, reactions, and metabolites), the GEMs created in this study are very comparable to other GEMs not based on *E. coli* [28, 30]. In contrast, when compared to models of *E. coli*, the models created in this study involve fewer genes, metabolites, and metabolic reactions, many of which are due to *E. coli* having an additional outer membrane and periplasm, which requires numerous additional transport and metabolic reactions for *in silico* inquiry [25, 26]. In addition, *E. coli* K12 is the most studied bacterium in the world, and approximately 70% of functions of the genes in the genome have been experimentally characterized in the last century. In contrast, a paucity of research exists for the characterization of genes encoding metabolic enzymes and transporters specific to *L. monocytogenes*.

2.5.3 Nutrient Utilization

The nutrient utilization experimental data generated as part of this study is some of the first large-scale metabolic data for numerous strains of *L. monocytogenes* (Tables 8-15). The compounds uniquely utilized by each strain could give some insight on why

that particular strain was isolated from its corresponding environment. For example, formic acid is a compound commonly used as a preservative in animal feed. Strain J2-031 was isolated from a bovine host, and was the only strain capable of metabolizing formic acid. It is possible that the utilization of formic acid as a carbon source from the animal's feed provided a competitive advantage for J2-031 in a bovine host. Another example of unique nutrient metabolism is L-serine utilization as a carbon source by strain JO161. L-serine is a nonessential amino acid in humans, meaning it can be synthesized by the body. Since strain JO161 was isolated as a clinical sample from a human patient, it could be that utilization of excess L-serine may provide strain JO161 a competitive growth advantage in its host environment. Other examples are sucrose, fumaric acid, and tyramine as carbon sources for strain F2365. Sucrose is added to many foods, tyramine is naturally present in many foods, and fumaric acid is an approved food additive. Since strain F2365 was isolated from a food epidemic, utilization of one or more of these compounds could have contributed to its successful survival and proliferation.

The nutrient utilization data also shows trends of metabolic similarities between strains. For example, of the 95 carbon sources tested, 35 had identical utilization across all of the strains, but just 15 were unique to one strain, and only one was serovar specific. This means that 44 of the compounds providing sole sources of carbon were used by some combination of strains. Analysis of these combinations, while beyond the scope of this study, could provide valuable insight into the metabolism of *L. monocytogenes*. In contrast, the metabolism of nitrogen, phosphorus, and sulfur was much more consistent across all six strains. 89 of the 95 nitrogen sources had the same utilization among all strains, and 90 of the 94 phosphorus and sulfur sources were consistent across strains, thus supporting a high level of conservation in the metabolic capabilities for these nutrients.

2.5.4 Nutrient Utilization Agreement Summary and Comparison

Using Biolog^{TM} phenotypic microarray plates to experimentally determine carbon source utilization phenotypic data is a very common method to validate GEMs for qualitative predictive accuracy. 57 of the carbon sources tested on Biolog^{TM} PM1 plates

were also present in the set of metabolites contained in the S matrix of the GEMs, thereby allowing comparisons of *in silico* and experimental nutrient utilization to be made. The draft versions of the GEMs created in this study through the use of the KBase genome-scale metabolic reconstruction pipeline had agreements between *in silico* predictions and experimental phenotype data ranging from 48.3% to 69.0%, with an average agreement of 56.9 ± 8.3% (Table 17). After manual curation added additional metabolic reactions and removed unnecessary transport reactions, this agreement increased to a range of 80.7% to 91.2%, with an average of 86.8 ± 4.3% (Table 17). This carbon source utilization agreement is slightly above the average agreement (86.6 ± 7.5%) of 16 published GEMs (Figure 9). In the case of *E. coli* K12, multiple iterations have expanded the contents of the series of GEMs for this organism, and each iterative metabolic reconstruction has led to improvement of the accuracy of nutrient utilization predictions from 70% to 98% [25, 49, 57]. This indicates that the models created in this study are comparable to those already published, and, as such, they can be used in similar ways that these other models have been used, for example, simulating complex environments such as foods or host niches.

Nitrogen source utilization experiments are another common method used to assess the validity of GEMs for qualitative growth predictions. Across the 62 nitrogen sources present in the GEMs created in this study, agreements ranged from 59.7% to 66.1%, with an average agreement of 63.2 ± 3.0%. This average is below that of the six GEMs used for comparison and falls within the third standard deviation from the average (77.0 ± 5.2%) (Figure 10). These results are not as promising as the carbon source utilization results; however, a more thorough review of the available genomic metabolism information could reveal possible opportunities for improvement since many of these nitrogen utilization pathways have not been extensively researched for *L. monocytogenes* strains. Because carbon source utilization is the most common form of model validation, the focus of this study was to maximize the agreement for carbon, but nitrogen, phosphorus, and sulfur were also included to investigate how well GEMs created with semi-automated tools can predict utilization of these nutrients. Therefore, less attention was paid to manual curation to improve the predictive accuracy of *L. monocytogenes*

GEMs for nitrogen metabolism, and these values could serve as a baseline for future studies to identify ways to improve genome-scale metabolic reconstruction pipelines.

In addition to carbon and nitrogen source utilization experiments, phosphorus and sulfur experiments are also performed and compared to *in silico* GEM predictions. Four previous GEMs were found that performed these experiments as part of the experimental validation process and their agreement can be seen in Figures 11 and 12. These GEMs had an average phosphorus agreement of 78.0 ± 42.0% and an average sulfur agreement of 45.3 ± 24.3%. In this study, there were 22 phosphorus and 11 sulfur sources for which comparisons were possible. Agreements for the GEMs created in this study ranged from 18.2% to 27.3%, with an average of 23.5 ± 4.5% for phosphorous, and from 72.7% to 81.8, with an average of 78.8 ± 4.7% for sulfur. This phosphorus agreement is well below the average of the four GEMs used for comparison (Figure 11). However, a comparably low phosphorus agreement (15%) was seen in one of the studies, so low phosphorus agreement is not unique to this study. Conversely, the sulfur agreement is considerably better than that of the four comparison GEMs, (Figure 12). Much like the nitrogen metabolism, phosphorus and sulfur metabolism received less vigorous curation, and, while the sulfur agreement is promising, the phosphorus agreement could potentially be improved. Both phosphorus and sulfur agreements, like nitrogen, could serve as baselines for future work.

While high-throughput phenotypic experiments like the Biolog^{TM} plates are extremely useful for the validation of GEMs, there are some drawbacks. Perhaps most notably, these plates detect cellular respiration, while the GEMs predict generation of biomass. While these functions are usually related, this is not always the case. Therefore, a possible reason for some of the false negatives seen throughout all of the nutrient utilization experiments is that the cells are capable of utilizing the compound for respiration, but do not obtain enough energy to grow from the nutrient as a sole source of carbon, nitrogen, phosphorous, or sulfur.

2.5.5 Essential Reactions

At the forefront of the possible applications for the use of GEMs is the identification of new methods for control for pathogens. This is primarily accomplished through the prediction of essential reactions. These metabolic reactions are those that are required for the organism to produce biomass and are likely necessary for viability and/or growth. They are identified by running *in silico* gene/reaction simulations: systematically simulating removal of one metabolic reaction at a time. These simulations attempt to optimize biomass production with the knocked out reaction. If no biomass is able to be produced, the reaction is considered essential. This approach has been successfully used with a GEM of *Vibrio vulnificus* to identify new control targets, and through screening of metabolic of metabolic analog libraries of compounds, led to new ways to kill and prevent the growth of *V. vulnificus* [30].

In this study, essential reaction simulations were first determined using the final, manually curated and validated GEMs in conditions simulating a glucose minimal media. Then, chemical compounds present the analytical composition of five of the foods most frequently recalled for *L. monocytogenes* contamination were added to the simulation medium, and essential reaction predictions were determined for each of these simulated environments. The five foods chosen were queso fresco, chicken breast, smoked salmon, cantaloupe, and romaine lettuce.

Summaries of the essential reaction predictions can be seen in Tables 21 – 26. Simulations run on the glucose minimal media predicted the most essential reactions (average = 368.3 ± 1.5). This was expected, since the food matrices that were simulated had additional nutrients available, which would allow the organism to compensate with alternative catabolic methods when certain metabolic reactions were eliminated. All of the essential reaction simulations done on foods predicted similar numbers of essential reactions (average = 298.3 ± 4.8). This was also expected, because the models did not contain all of the different nutrients available in each of the different foods, and; therefore, the compositions of the simulated media were very similar.

The predictions of essential reactions generated by this study can be used by future studies to identify new methods of control for a broad spectrum of *L.*

monocytogenes strains. The logical next step from this analysis is for future studies to experimentally validate the essentiality of the metabolic reactions. This can be accomplished by performing gene knockout experiments, which can be time consuming if mutant libraries are not readily available such as those for *E. coli* K12. Knocking out, or preventing the expression of, the genes associated with the predicted essential reactions should be lethal to the organism. Once the essentiality of the reactions has been confirmed, new drugs, such as chemical analogs that mimic metabolites that irreversibly bind to the essential enzyme, or other control methods, can be developed to target the genes and their associated reactions. There has been success for using GEM essential reaction predictions to aid in the development of drugs targeting foodborne pathogens, specifically *Vibrio vulnificus* [30].

2.6 Conclusion

This study describes the creation and qualitative experimental validation of the first six GEMs for *L. monocytogenes*, and also examines the feasibility of using semi-automated, online genome-scale metabolic reconstruction tools for GEM creation rather than the slower, manual GEM reconstructions. The GEMs used in this study reflect the expected conservation of the genome of *L. monocytogenes* by spanning similar numbers of genes, and containing similar numbers of metabolites and metabolic reactions. The numbers of genes, metabolites, and reactions are also on the expected scale for non-*E. coli* GEMs.

Experimentally validating the GEMs was accomplished using Biolog^{TM} PM plates. These experiments allowed the comparison of the nutrient utilization predictions made by the models to experimental data. Nutrient utilization comparisons were made for sole sources of carbon, nitrogen, phosphorous, and sulfur. The models were then curated to better agree with the experimental results for carbon utilization, since that is the method most commonly used to validate GEMs. It was found that the final carbon source utilization agreement of the newly created models was comparable to the agreement seen by previously published models. Additionally, the nitrogen and phosphorus agreements were lower than previously published models, but the sulfur

agreement was higher than previously published models. This indicates that unknown reactions and the corresponding genes may have yet to be identified and characterized for *L. monocytogenes* strains. These results show that GEMs constructed by KBase are comparable to manually constructed GEMs, provided they are validated through comparison to experimental qualitative data.

Additionally, the results of this work can be used in similar ways to how other validated GEMs have been used, such as identifying potential essential reactions. The potential essential reactions identified by this study provide promising directions for future studies improving the safety of foods commonly contaminated by *L. monocytogenes*, including cheese, deli meats, salads, and fruit. The models could also be used as starting points for future iterations of GEMs for *L. monocytogenes*. As exemplified by *E. coli* GEMs, making the most comprehensive and accurate GEM is an ongoing, iterative process, and the GEMs created in this study have room for improvement for future studies, particularly in identification of transporters and reactions for pathways involved in nitrogen and phosphorous metabolism.

Finally, a by-product of the experimental validation process was the generation of large quantities of *L. monocytogenes* metabolic data, of which there is little available. This data could provide insight into metabolic differences between strains and serovars, and could potentially elucidate phenotypic strategies for the successful proliferation of *L. monocytogenes* in specific environments, such as foods.

3 Quantitative Analysis of Genome-Scale Metabolic Models for Six *L. monocytogenes* Strains

3.1 Abstract

The objectives of this study were to compare *in silico* predictions of six newly validated *L. monocytogenes* GEMs to experimental results of growth rate and growth yield, and to adapt these GEMs to previously published, low-temperature growth data for *L. monocytogenes*. The GEMs for this study were created and validated as described previously in this work. Identification of the chemically defined growth medium to be used for growth rate and yield determination was accomplished by microwell growth assays. Experimental data for growth rate and growth yield were collected during batch growth experiments using aerobic spargers. Dry cell weight, optical density, and viable cell concentration measurements were collected, resulting in conversion and scaling factors that enabled refinement of *in silico* predictions.

The medium identified as most suited for batch growth experiments was Modified Welshimer's Broth supplemented with 3% Brain Heart Infusion broth. With the introduction of scaling factors, the *L. monocytogenes* GEMs were able to generate *in silico* predictions for growth rate and growth yield that were strongly and significantly ($p < 0.0013$ and $p < 0.0015$) correlated with experimental results. The GEMs also showed promise as starting points for future studies investigating low-temperature *in silico* growth predictions for *L. monocytogenes*.

3.2 Introduction

In addition to their food safety potential, GEMs can be a powerful laboratory tool for studying the biological complexity of organisms at a systems level. The usefulness of computational models of microorganisms increases through validation with experimental data. GEMs are commonly qualitatively validated through nutrient utilization data for carbon, nitrogen, phosphorous, and sulfur to iteratively improve the accuracy of *in silico* growth predictions. To increase the accuracy of the *in silico* predictions beyond the improvements made through qualitative validation, it is necessary to quantitatively validate the GEMs with additional experimental data.

Quantitative validation is primarily accomplished by performing aerobic batch growth experiments that yield growth rates (h^{-1}) and biomass yields (gDCW/g glucose) as previously described [25]. These experimental values can be compared to *in silico* predictions, and allow for adjustments to the model which improve its predictive power and accuracy for quantitative growth measurements. The improved GEMs can then be used to predict how *L. monocytogenes* will behave under a wide variety of different conditions prior to performing rigorous laboratory techniques. Additionally, this data can be used to generate numerical values that can convert between experimentally determined data, such as viable cell counts, optical density, and biomass. These conversion factors can be used to utilize existing *L. monocytogenes* growth data for new *in silico* inquiry using GEMs.

3.3 Materials and Methods

3.3.1 Growth Assays

The strains of *L. monocytogenes* used for quantitative validation are the same six strains used for qualitative validation (Table 2). To ensure the successful growth of these strains of *L. monocytogenes* in the selected minimal media, growth curves were generated using 96 well plates and the BioTek® Epoch 2 Microplate automated spectrophotometer (BioTek Instruments, Inc., Winooski, VT). Three chemically defined media were tested. The first was an improved minimal medium (IMM) described by Phan-Thanh *et al*. [58]. The second was Modified Welshimer's Broth (MWB), as described by Premaratne *et al*. [8]. The final, and most successful, medium tested was MWB supplemented with 3% BHI broth. When generating a growth curve, four isolated colonies were inoculated into separate test tubes filled with 9 mL of the chosen chemically defined media. These tubes were then incubated at 37°C for 24 hours with shaking at 250 rpm. After incubation, three of the samples for each strain (18 total) were transferred to sterile 15 mL conical tubes, and the OD_{600} was adjusted to 0.040 ± 0.010 using fresh media. After the OD_{600} adjustment, 200 µL of each sample was placed into separate wells of a 96 well plate, along with triplicate wells of fresh, uninoculated media as the negative growth control. The 96 well plate was then placed in the plate reader, where OD_{600} measurements were

taken every 10 minutes for 48 hours at 37°C. Every six readings were then averaged to generate a growth curve with hourly time points. Inconsistent growth was observed with the first two chemically defined media tested (IMM and MWB), but growth was supported by MWB supplemented with 3% BHI.

3.3.2 Batch Growth Experiments

In order to grow *L. monocytogenes* in sufficient quantities to perform dry cell weight measurements, batch spargers, as described by Sutton *et al*. [59], were used (Figure 13). The day prior to the experiment, MWB supplemented with 3% BHI broth (MWB3) was prepared and pipetted in 15 mL aliquots into nine sterile test tubes. These tubes were then each inoculated with an isolated colony of the *L. monocytogenes* strain to be tested. The tubes were incubated overnight at 37°C with shaking at 250 rpm.

MWB3 was prepared fresh prior to the start of each experiment and 600 mL aliquots were placed into three separate, sterile 1 L glass sparger bottles. Additionally, 50 mL were prepared for use as spectrophotometer blanks. A 100 µL sample of the overnight culture was taken and diluted so that the OD_{600} was 0.040 ± 0.010. Then, the remaining overnight culture was used to inoculate the three batches of 600 mL MWB3 such that the dilution was equivalent to the dilution required to obtain the desired OD_{600} in the 100 µL sample. A sample from each of the batch spargers was then taken, and its OD_{600} measured. Subsequent OD_{600} measurements were taken hourly for each of the batches for the duration of the run.

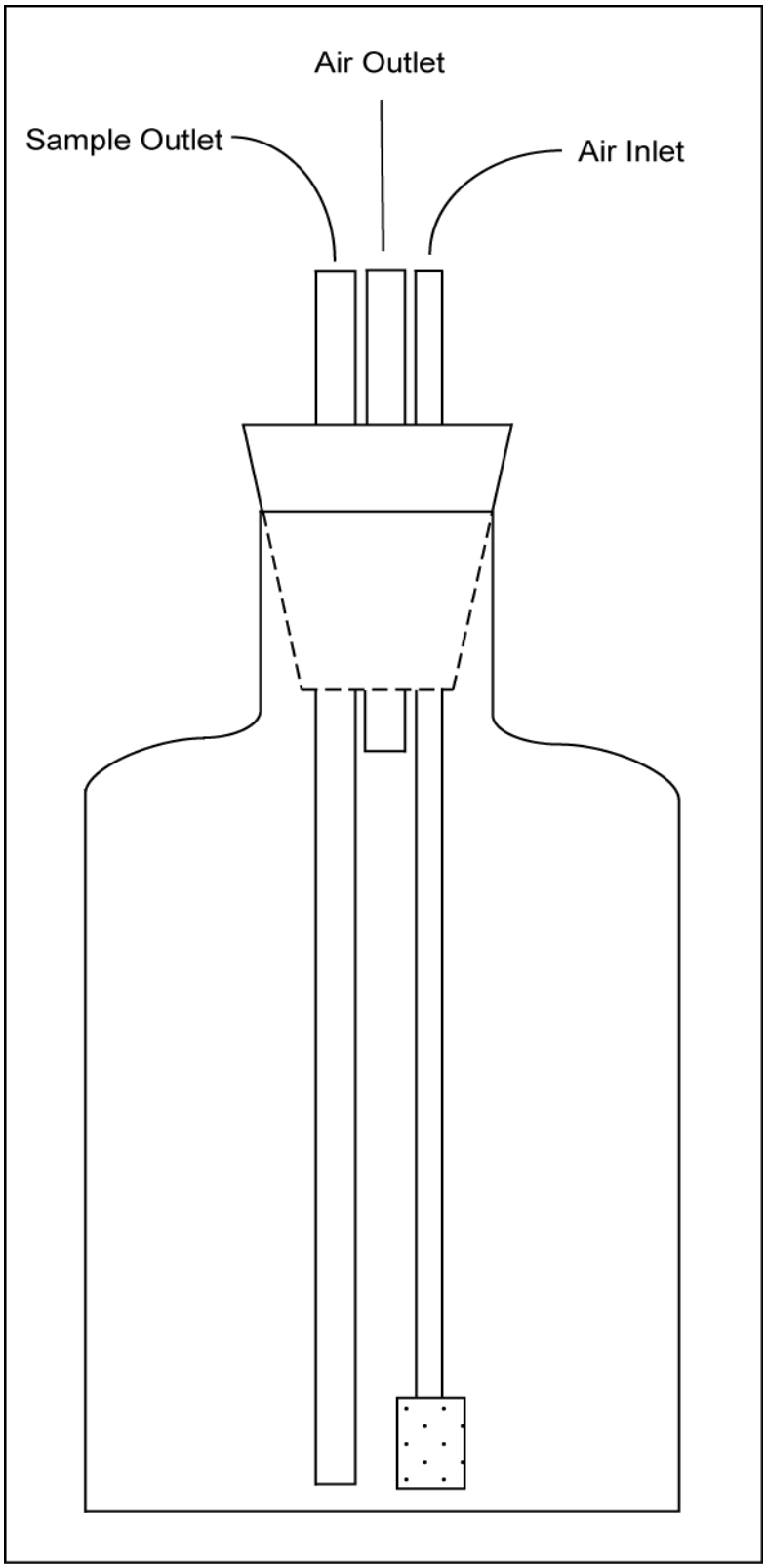

Figure 13: Diagram of the batch sparger apparatus used in this study

3.3.3 Dry Cell Weight Measurements

Dry cell weight measurements were performed in conjunction with the aerobic batch growth experiments. 12 (3 batches x 4 time points) Whatman™ glass microfiber filters (GE Healthcare Bio-Sciences, Pittsburgh, PA) were labeled and placed in an 80°C oven and left overnight to dry. One hour prior to beginning dry cell weight experiments, the dry filter weight was measured and recorded as described by Baumler *et al*. [25]. The first time point for the dry cell weight experiment was when the OD_{600} of the batch growth spargers reached 0.100.

When the OD_{600} reached 0.100, approximately 60 mL was extracted from each batch. 50 mL of each sample was vacuum filtered through its corresponding pre-weighed filter paper. After the culture was filtered, the paper was washed with approximately 5 mL of sterile DI water, left to dry for two minutes, and washed again with approximately 5 mL of sterile DI water. When all three of the filter papers for the time point were finished, they were placed in an 80°C oven to dry for a minimum of 24 hours. This process was repeated hourly for a total of four time points. 24 hours after the last time point, the filters were re-weighed, and the dry cell weight was determined from the difference between the pre-weighed filter mass, and the post-filtration mass.

3.3.4 Viable Cell Counts

At the second dry cell weight time point, a CFU/mL count was determined for each of the three batches. Duplicate samples from each batch were taken and serially diluted to 10^{-9}. The six highest dilutions ($10^{-4} – 10^{-9}$) were then plated on BHI agar plates, incubated at 37°C for 24 hours, and enumerated.

3.3.5 *In Silico* Quantitative Adjustment

Following the completion of the batch growth dry cell weight experiments, the experimental data was used to adjust the models and improve the accuracy of the *in silico* predictions for biomass corresponding to growth rate. First, the data was used to generate conversion factors between biomass and optical density and between biomass and viable cell counts. Subsequent analysis of the data generated experimental growth rates (h^{-1}),

which were compared to *in silico* predicted growth rates to generate a scaling factor as previously described for the *E. coli* K12 GEMs [60]. This scaling factor was then used as a multiplier of biomass for FBA and dynamic FBA to improve the accuracy of *in silico* growth predictions. Initial *in silico* growth rates were determined by adjusting *in silico* glucose uptake rates until the duration of the batch growth matched experimental values. These glucose uptake rates were then held constant when the scaling factor was introduced, allowing the scaling factor to increase the accuracy of predictions of biomass used to generate values for growth rate and growth yield.

3.3.6 Phylogenetic Analysis

A maximum likelihood phylogenetic tree was constructed for the six strains of *L. monocytogenes* utilized in this study, and the non-pathogenic species *L. welshimeri* was used as an outgroup. The alignment used was previously generated in Mauve for the qualitative validation. Phylogenetic analysis and tree construction was performed with MEGA7 using default parameters [61].

3.4 Results

3.4.1 Growth Curves

In order to utilize batch growth experimental data to quantitatively validate the GEMs created in this study, the media used to perform the batch growth experiments must be chemically defined so that the nutrient amounts can be accurately used for *in silico* growth simulations. Therefore, growth curves were generated to test several chemically defined media to determine which media would support growth of the six strains of *L. monocytogenes* for subsequent batch growth assays. IMM was the first medium tested, and growth was inconsistent across strains (Figure 14). Strain ScottA was particularly problematic, but growth was only observed in strains J2-064 and JO161, with late growth in strain F2365. The second medium tested was MWB. Similar to IMM, growth was inconsistent (Figure 15). Again, growth was not supported for strain ScottA, and strain F2365 was also not supported. However, growth was supported for the four remaining strains. Finally, MWB supplemented with 3% BHI was tested. This

medium supported growth of all six strains (Figure 16), and was; therefore, used for subsequent batch growth experiments.

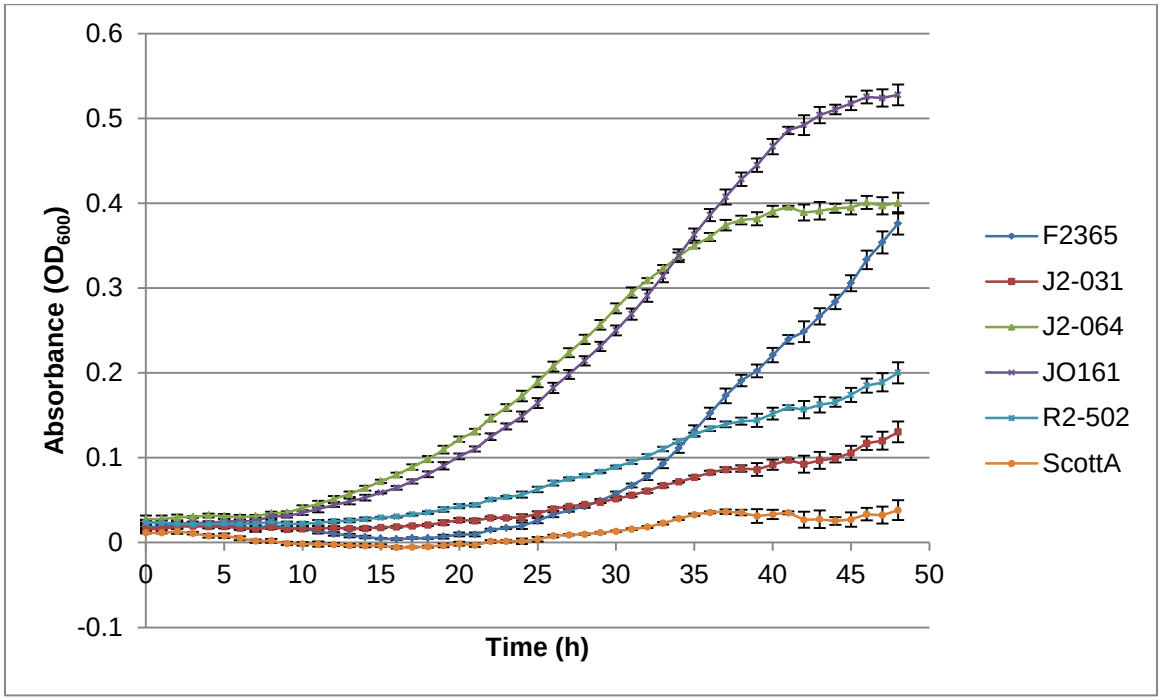

Figure 14: Microwell growth assay of six *L. monocytogenes* strains in IMM at 37°C including standard deviations

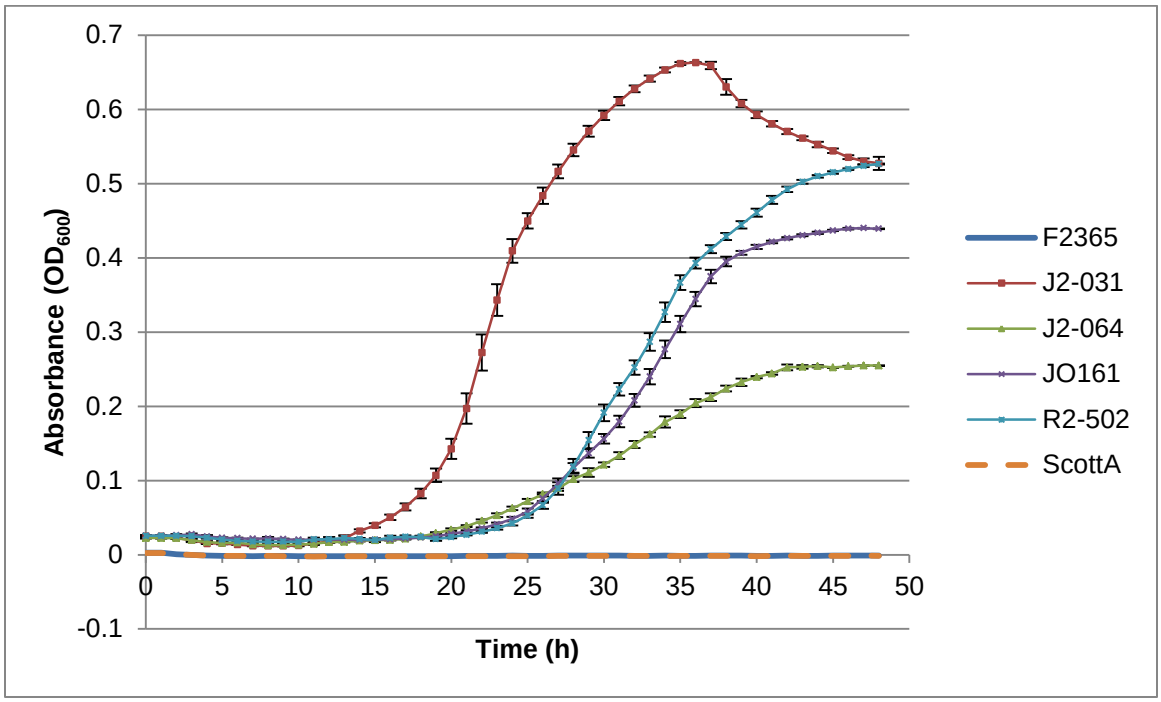

Figure 15: Microwell growth assay of six *L. monocytogenes* strains in MWB at 37°C including standard deviations

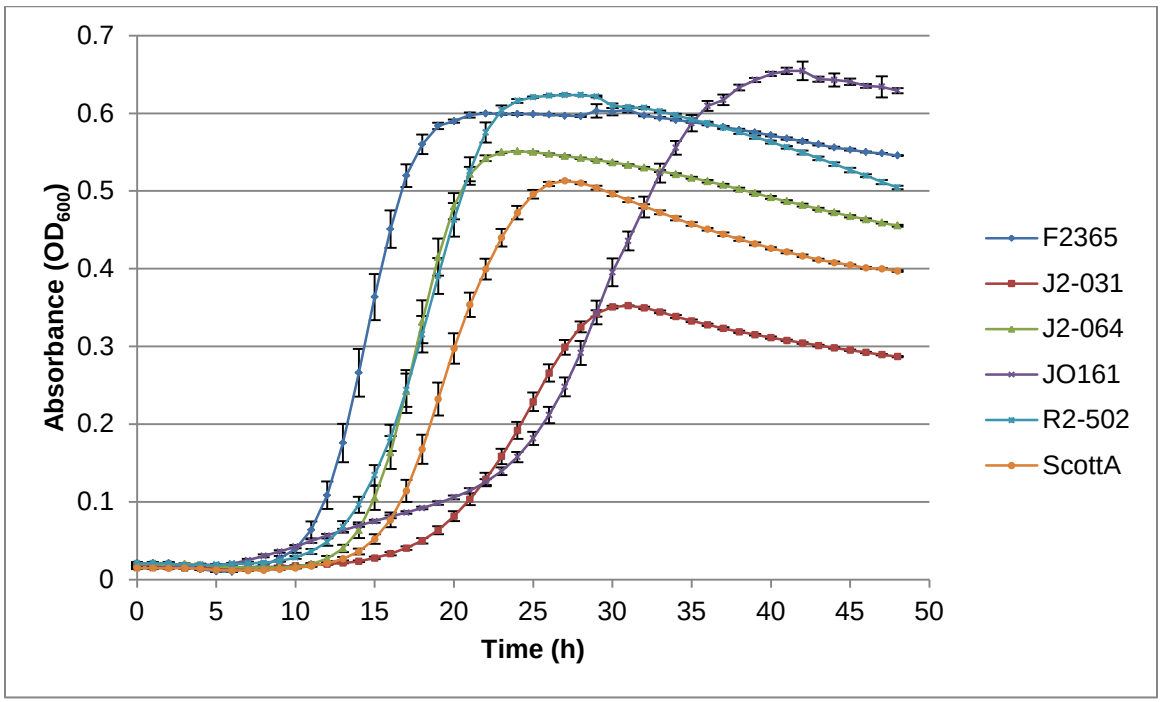

Figure 16: Microwell growth assay of six *L. monocytogenes* strains in MWB supplemented with 3% BHI at 37°C including standard deviations

3.4.2 Conversion Factors

The data generated by the batch growth experiments included data for viable cell counts, optical density, and biomass. Since this data was obtained for each of the six strains, it was possible to calculate conversion factors to convert one type of data to another. Calculated conversion factors were determined to compare and convert viable cell counts to biomass and optical density to biomass, which are the two most important conversions for subsequent model validation (Table 33).

Table 33: Experimentally determined conversion factors

Strain	Viable Cells (CFU/mL) to Biomass (gDCW/L) ± SD[b]	OD_{600} to Biomass (gDCW/L) ± SD
J2-031	---[a]	0.127 ± 0.044
JO161	$2.8 \times 10^{-10} \pm 3.3 \times 10^{-11}$	0.378 ± 0.217
J2-064	$4.4 \times 10^{-11} \pm 1.7 \times 10^{-11}$	0.0938 ± 0.0479
R2-502	$1.3 \times 10^{-10} \pm 6.4 \times 10^{-11}$	0.227 ± 0.087
F2365	$4.6 \times 10^{-11} \pm 1.2 \times 10^{-11}$	0.0819 ± 0.0391
ScottA	$4.3 \times 10^{-11} \pm 2.1 \times 10^{-11}$	0.103 ± 0.076

[a] Viable cell count was not taken for strain J2-031
[b] Standard Deviation

3.4.3 Batch Growth and Qualitative *In Silico* Refinement

Experimental growth rates (h^{-1}) were calculated using linear regression on the four, hourly dry cell weight measurements taken for each strain. The initial *in silico* growth rate predictions were obtained by first using the optical density to biomass conversion factors (Table 33) to approximate the initial biomass at the start of each of the batch inoculum. The duration for completion of batch growth *in silico* was then adjusted to reflect the amount of time determined experimentally. The correlation between the experimental growth rates and *in silico* growth rate predictions was neither strong, nor significant (Pearson correlation test statistic yields $p < 0.32$). Therefore, the scalar differences between the experimental growth rates and the initial predictions were used to multiply the predicted biomass values in order to better reflect experimental growth rates. After introduction of the scalar factors determined for each strain, the correlation between the experimental growth rates and *in silico* predictions was much stronger and significant

($p < 0.013$). Similarly, the correlation between the experimental growth yields and the initial *in silico* predictions were neither strong, nor significant ($p < 0.36$). However, after using the scalar factors to adjust the models, the correlation became much stronger and significant ($p < 0.0015$).

Table 34 shows the experimental growth rates observed, the initial model predictions before the scalar was introduced, and the model predictions after the scalar was introduced. Additionally, growth yields (g/g glucose) were determined experimentally and predicted by the model both before and after the scalar factors were introduced (Table 35).

A graphical of representation of the growth characteristics obtained experimentally and predicted *in silico* was obtained (Figure 17). The figure omits the data generated by the model before the scalar factors were introduced, due to the stronger correlation between the experimental results and the predictions after the scalar factors were introduced. Figure 18 shows a phylogenetic tree of the six strains of *L. monocytogenes* compared to a non-pathogenic species, *L. welshimeri*. Included in the tree are the experimental growth statistics for each strain, as well as the *in silico* predictions of growth characteristics after the scalar factors were introduced (also seen in Tables 34 and 35).

Table 34: Comparison of experimental and *in silico* growth rates (h^{-1})

	Experimental ± SD	*In Silico* Initial	*In Silico* with Scalar Factor
J2-031	0.00768 ± 0.00765	0.06052	0.00140
JO161	0.02391 ± 0.01636	0.05617	0.01555
J2-064	0.00542 ± 0.00638	0.07180	0.00064
R2-502	0.02175 ± 0.01562	0.07820	0.00693
F2365	0.00307 ± 0.00189	0.01841	0.00096
ScottA	0.00830 ± 0.01641	0.03872	0.00313

Table 35: Comparison of experimental and *in silico* growth yields (g/g glucose)

Strain	Experimental ± SD	*In silico* Initial	*In silico* with Scalar Factor
J2-031	0.00718 ± 0.00282	0.02690	0.00314
JO161	0.01701 ± 0.00857	0.02497	0.01221
J2-064	0.00587 ± 0.00287	0.03191	0.00228
R2-502	0.01344 ± 0.00636	0.03475	0.00767
F2365	0.00477 ± 0.00171	0.00818	0.00170
ScottA	0.00578 ± 0.00744	0.01721	0.00434

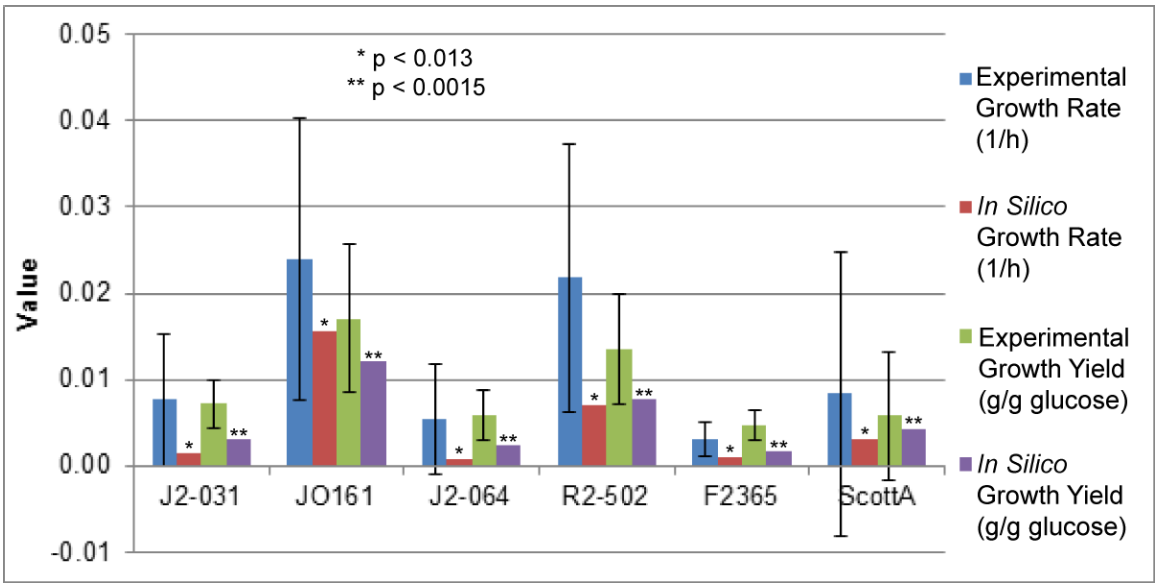

Figure 17: Comparison of experimental and *in silico* growth rates and growth yields including experimental standard deviations

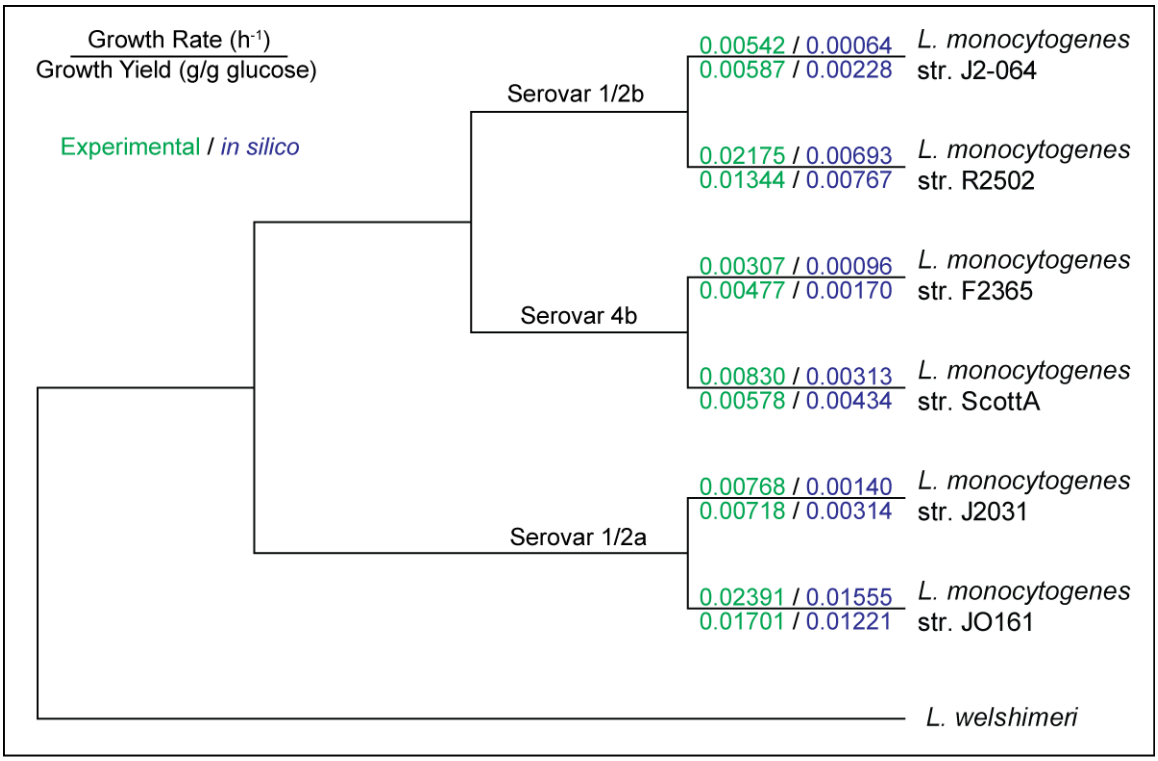

Figure 18: Maximum likelihood phylogeny of *L. monocytogenes* strains used in this study and their associated growth rates and growth yields

3.4.4 *In Silico* Growth Predictions in Food Matrices at Various Temperatures

Also of interest was whether the GEMs created in this study could be used in conjunction with previously published *L. monocytogenes* growth data. Particularly, we sought to determine if GEMs could be used to predict growth at lower temperatures in simulations of food matrices. This was of interest because GEM simulations are currently used for a single, optimum growth temperature. We sought to improve the practicality of GEMs by adjusting them to generate growth predictions at sub-optimal temperatures.

Experimental data of viable cell counts was available for two different media (beef frankfurters and sliced turkey breast) at four different temperatures (4°C, 8°C, and 12°C) for four of the strains used in this study (J2-064, R2-502, F2365, and ScottA) [62]. This data was converted to biomass using the conversion factors previously described in this work (Table 33). The converted values were then used to determine experimental growth rates, and compared to *in silico* growth rate predictions in the same media using

the models with previously determined scalar factors. The resultant scalar factors are presented in Tables 36 and 37 for growth in beef frankfurter and sliced turkey breast slurries, respectively. For comparison, the scalar factors introduced to the GEMs during the batch growth experiments are also presented (Tables 36 and 37). These scalar factors were then multiplied by the previously described *in silico* growth rates to obtain strain-specific, low temperature *in silico* growth rates. These new, low-temperature growth rates were then compared to the experimental growth data, starting from the same initial biomass (Figures 19 – 24). The correlation between the new, low-temperature growth rates and the experimental growth rates was neither strong, nor significant (Pearson correlation test statistic yields $p > 0.10$) for any of growth conditions simulated, with the exception of turkey breast slurries at 8°C ($p < 0.01$).

Table 36: Scalar factors between experimental and *in silico* growth rates for beef frankfurters compared to scalars already introduced

	4°C Franks	8°C Franks	12°C Franks	37°C Minimal Media
J2-064	1.31×10^{-4}	2.71×10^{-5}	2.73×10^{-6}	13.2
R2-502	1.71×10^{4}	1.93	0.0153	3.6
F2365	1.25×10^{3}	3.04×10^{-3}	2.62×10^{-4}	5.99
ScottA	26.5	0.318	9.03×10^{-5}	4.67

Table 37: Scalar factors between experimental and *in silico* growth rates for sliced turkey breast compared to scalars already introduced

	4°C Turkey	8°C Turkey	12°C Turkey	37°C Minimal Media
J2-064	1.07×10^{-5}	4.30×10^{-5}	5.56×10^{-7}	13.2
R2-502	0.0316	5.68×10^{-3}	3.82×10^{-5}	3.6
F2365	0.0129	2.45×10^{-6}	8.82×10^{-7}	5.99
ScottA	5.90×10^{-4}	1.30×10^{-3}	6.12×10^{-6}	4.67

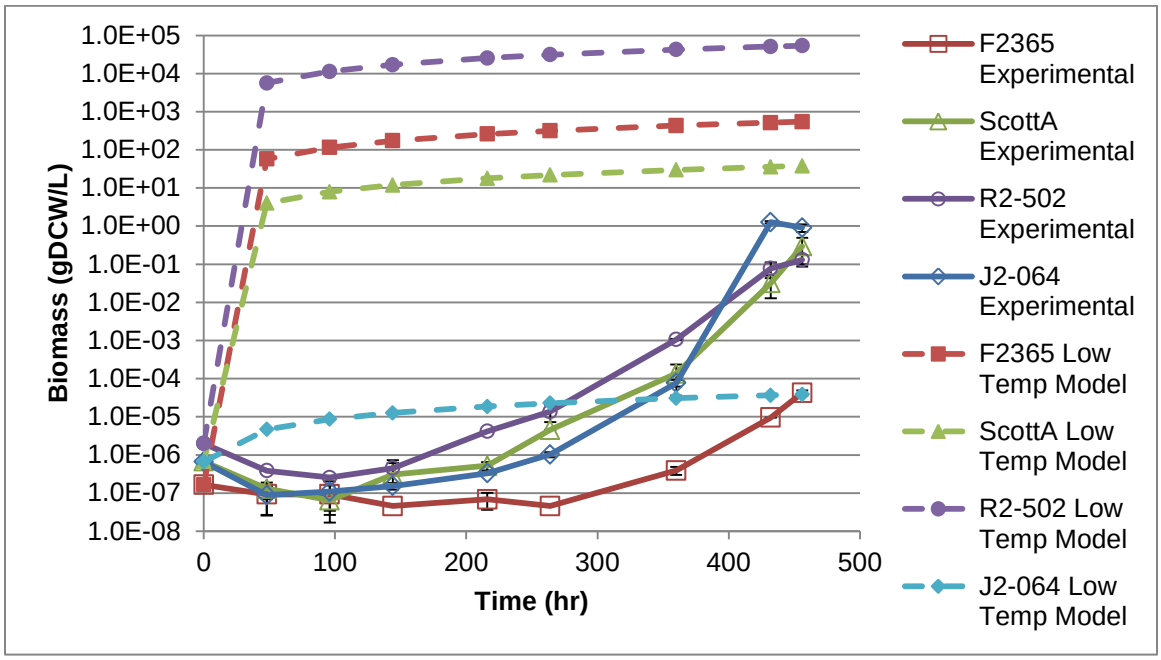

Figure 19: Comparison of *in silico* predicted growth rate and experimental growth data for four strains of *L. monocytogenes* in beef frank slurries at 4°C including experimental standard deviations

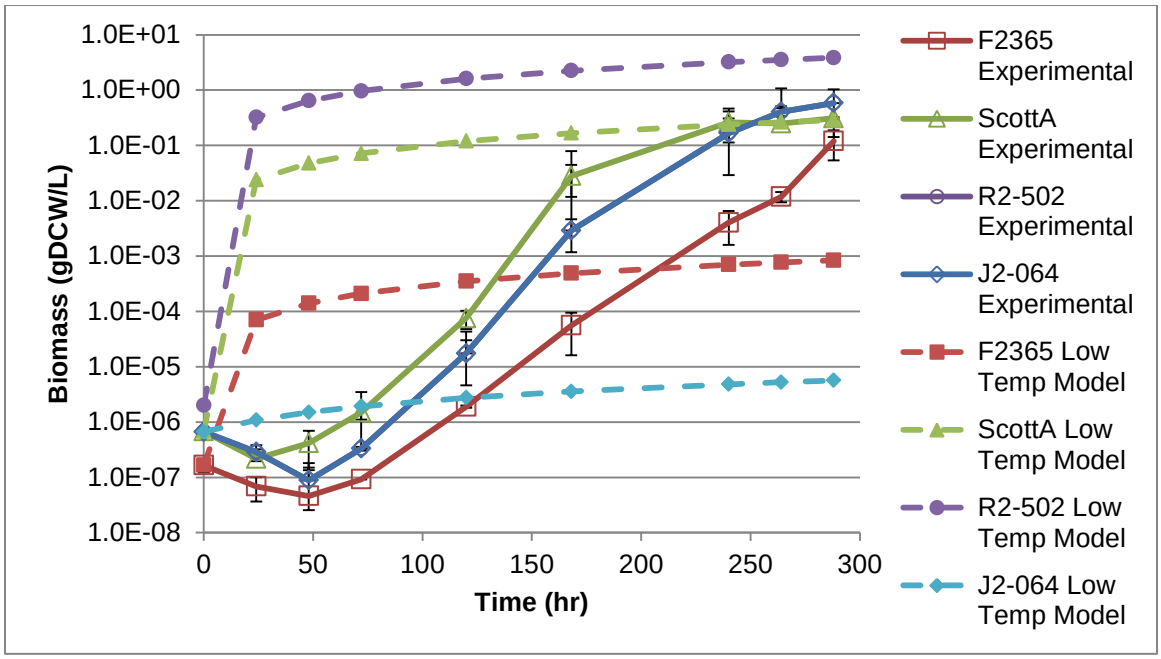

Figure 20: Comparison of *in silico* predicted growth rate and experimental growth data for four strains of *L. monocytogenes* in beef frank slurries at 8°C including experimental standard deviations

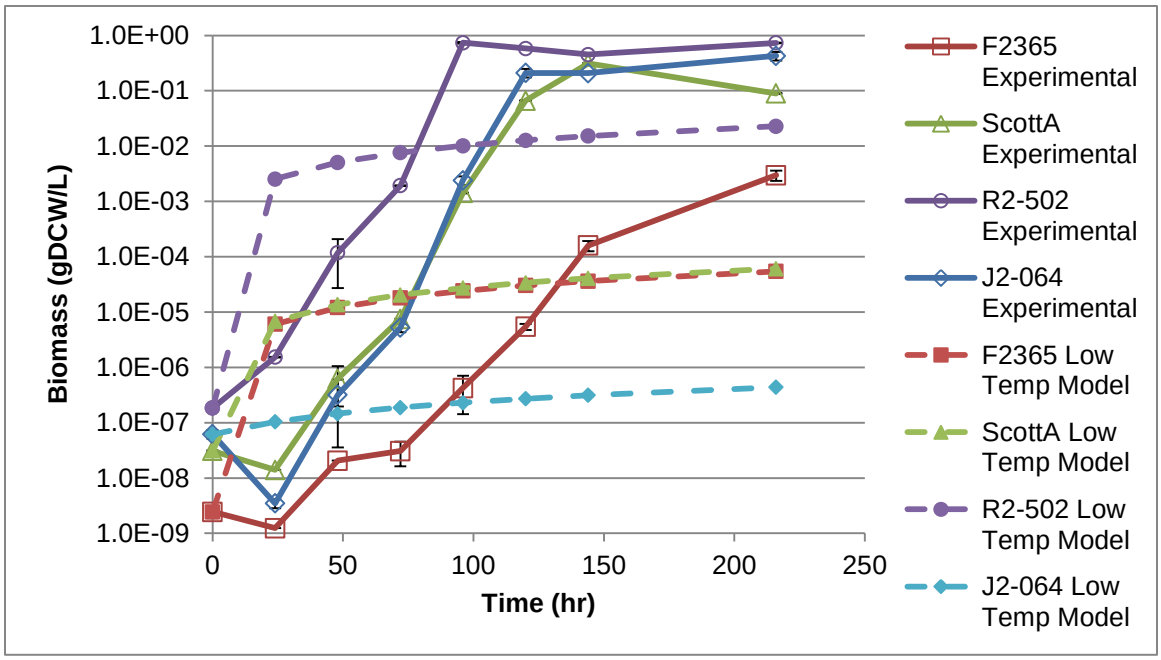

Figure 21: Comparison of *in silico* predicted growth rate and experimental growth data for four strains of *L. monocytogenes* in beef frank slurries at 12°C including experimental standard deviations

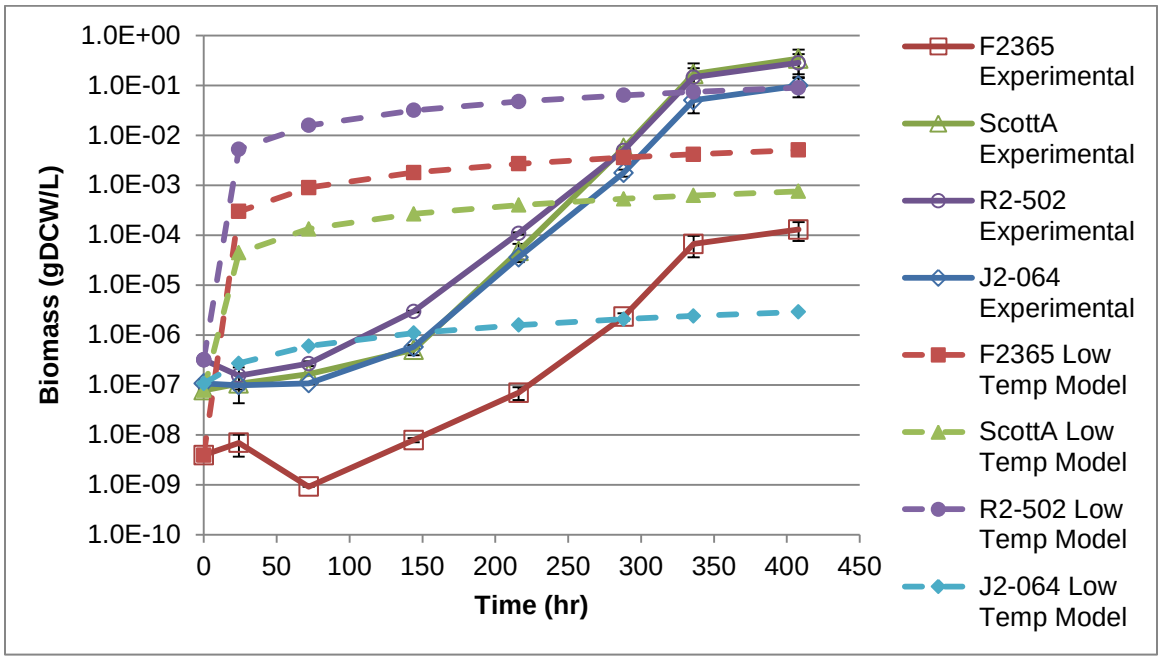

Figure 22: Comparison of *in silico* predicted growth rate and experimental growth data for four strains of *L. monocytogenes* in sliced turkey breast slurries at 4°C including experimental standard deviations

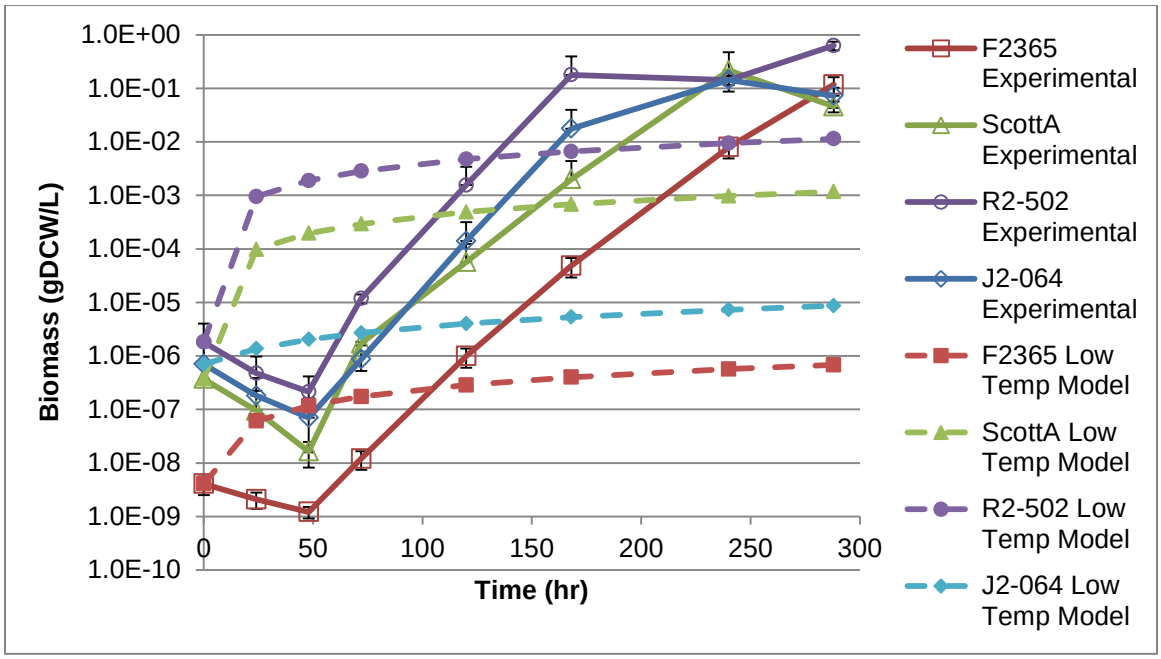

Figure 23: Comparison of *in silico* predicted growth rate and experimental growth data for four strains of *L. monocytogenes* in sliced turkey breast slurries at 8°C including experimental standard deviations

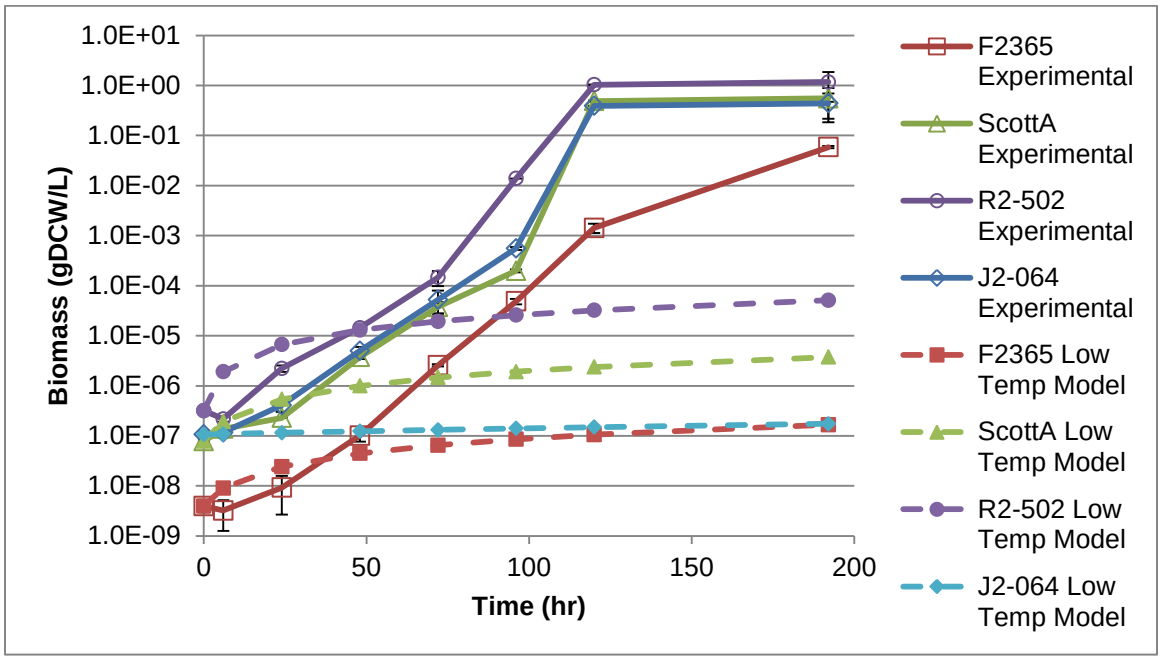

Figure 24: Comparison of *in silico* predicted growth rate and experimental growth data for four strains of *L. monocytogenes* in sliced turkey breast slurries at 12°C including experimental standard deviations

3.5 Discussion

3.5.1 Growth Assays

The importance of running qualitative batch growth experiments on chemically defined media cannot be understated. Without a precise, known nutrient composition, the environmental conditions used for the growth simulations will not be accurate, which undermines the accuracy with which the GEMs can make accurate predictions. Difficulties arose in achieving consistent growth on both IMM and MWB (Figures 14 and 15). Multiple trials were run for each media, but little consistency across strains and the inability to reproduce the work of Phan-Thanh *et al*. [58] and Premaratne *et al*. [8] resulted in the decision to supplement MWB media with 3% BHI. This is likely due to minute requirements of vitamins or cofactors absent in the previously described minimal media recipes. While this slightly compromises the chemical composition of the media, it resulted in consistent overnight growth in the media, which made the qualitative batch growth experiments feasible.

3.5.2 Conversion Factors

Some degree of variation in the conversion factors was observed (Table 33). One potential reason for the uncertainty is the relatively few measurements taken of dry cell weight (4) and viable cell concentration (1). While these measurements, taken in triplicate, allow for the generation of necessary growth rates and growth yields, the values for determination of conversion factors may be improved by increasing the number of time points used. However, the values used here to determine *in silico* growth rates and growth yields had a strong correlation to experimental values and were determined to be statistically significant. Therefore, these values provide new tools for the field of systems biology and for future studies performing *in silico* work with *L. monocytogenes* strains.

When comparing the conversion factors generated in this study to those previously published [25], it can be seen that the optical density to viable cell concentration factors are on similar scales (average = 0.168 and 0.418 for *L. monocytogenes* and *E. coli*, respectively). This was expected, since it is a comparison of two bacteria. Conversely, the conversion factors for viable cells to biomass differ by several orders of magnitude (average = 1.09×10^{-10} and 3.27×10^{-7} for *L. monocytogenes* and *E. coli*, respectively). The degree to which the viable cells to biomass conversion factors differ is not surprising, since *E. coli* is Gram negative and *L. monocytogenes* is Gram positive and the additional outer membrane and periplasm found in Gram negative bacteria may account for some of these differences.

3.5.3 Batch Growth and Growth Rate and Yield Comparison

Batch growth experiments were conducted in order to make several comparisons between six *L. monocytogenes* strains representing the three serovars (1/2a, 1/2b, and 4b) used in this study. The first comparison made was of experimental growth rates and growth yields between the six strains representing the three serovars. It was determined that there was no statistically significant (student's t-test statistic yields $p > 0.16$) difference in growth rate between any of the strains or any of the serovars. A trend was observed in which serovar 4b had a lower growth rate (average = 0.00569 h^{-1}) than either

serovar 1/2a (0.01580 h^{-1}) or 1/2b (0.01359 h^{-1}), but the difference was not significant ($p > 0.42$). Similarly, there was no significant difference observed ($p > 0.13$) in growth yields between strains or serovars. As with growth rates, there was a trend that indicated serovar 4b had lower growth yields (average = 0.00528 g/g glucose) than either serovar 1/2a (0.01210 gDCW/g glucose) or 1/2b (0.00966 gDCW/g glucose), but again, the difference was not significant ($p > 0.40$).

The second comparison was of the initial *in silico* predictions of growth rates and growth yields between serovars. As was seen with the experimental values, there was no significant ($p > 0.06$) difference between any of the serovars for either growth rate or growth yield. Again reflecting the experimental values, an observed trend indicating serovar 4b had a lower growth rate (average = 0.02857 h^{-1}) and growth yield (average = 0.01270 gDCW/g glucose) than either serovar 1/2a (0.05835 h^{-1} and 0.02594 gDCW/g glucose) or 1/2b (0.07500 h^{-1} and 0.03333 gDCW/g glucose). However, neither of these trends were significant ($p > 0.06$).

Next, we sought to compare the experimental growth rates and growth yields and the initial *in silico* predictions of these values. This was in order to see how close the initial *in silico* predictions were to the experimental results. This comparison revealed that the initial predictions were considerably inaccurate. Initial growth rate predictions were all at least one standard deviation above the experimental average, and for several of the strains, the predictions were much more than one standard deviation above. Similarly, all of the initial growth yield predictions were at least one standard deviation above the experimental average, except for strain JO161, which was just inside one standard deviation. These results, while good in that all strains were consistently over-estimated, indicated that steps needed to be taken to improve accuracy of quantitative *in silico* growth predictions, namely the introduction of scalar factors similar to the work with *E. coli* K12 GEMs [60].

Next, we sought to determine differences through comparison of the *in silico* predictions for growth rate and growth yield after the introduction of the scalar factors. Similar to the experimental and initial model comparisons, there was no significant ($p > 0.48$) difference between serovars. The trends indicating that serovar 4b had the lowest

growth rate (average = 0.00205 h^{-1} vs. 0.00848 and 0.00379 h^{-1}) and growth yield (average = 0.00302 gDCW/g glucose vs. 0.00768 and 0.00498 gDCW/g glucose) were less pronounced after the introduction of the scaling factors. Therefore, it was not unexpected that the trends, like those for experimental and initial model comparisons, were not significant ($p > 0.48$).

To improve quantitative *in silico* predictions, we compared the experimental and *in silico* values for growth rate and growth yield after the introduction of scalar factors into the FBA algorithm. As shown, the determination and implementation of the scalar factor greatly improved the accuracy of the *in silico* predictions (Tables 34 and 35, Figure 17). With the use of the scaling factors to multiply the biomass value, all of the *in silico* prediction values for growth yields correlated with the experimentally determined values, and all fell within one standard deviation from the average. Similarly, all but one of the *in silico* prediction values for growth rates correlated with the experimentally determined values, and fell within one standard deviation from the average. The one exception was strain F2365, which was just outside of one standard deviation from the experimental average. These results, combined with the strong, significant (Pearson correlation test statistic yields $p < 0.013$ and $p < 0.0015$) correlation between the experimental values, indicate that using the scalar factors was a valid method for improving the predictive accuracy of the GEMs. This is similar to the work conducted with various iterations of *E. coli* K12 GEMs, which has been useful for bioengineering strains for production of commodities and examining the evolutionary relationships of numerous pathogenic and nonpathogenic *E. coli* strains [25, 49, 57]

To view the work presented here in an evolutionary context, a maximum likelihood phylogenetic tree was constructed for the six strains of *L. monocytogenes* utilized in this study, and the non-pathogenic species *L. welshimeri* as an outgroup for tree construction (Figure 18). Included are the experimental growth rates and growth yields as well as the *in silico* predicted values after the introduction of the scalar factors. The phylogeny agrees with what was expected based on previous studies [9], and the similarities of growth characteristics of these six *L. monocytogenes* strains indicates conservation through evolutionary time. This indicates that other *L. monocytogenes*

strains would likely yield similar metabolic properties since this is likely an ancestral phenotype.

3.5.4 Utilization of Previous Data

The utilization of previous, low temperature growth data for *L. monocytogenes* was of particular interest for a number of reasons. First, it is widely known that *L. monocytogenes* is capable of growing at low temperatures. This makes it a major concern for a food industry that relies on refrigeration to keep food safe. Additionally, the previous growth data used slurries of two ready-to-eat meat products, which are products that are high-risk for *L. monocytogenes* contamination.

Using the conversion factors previously described in this thesis, previous *L. monocytogenes* growth data was converted from viable cell concentration to biomass. This biomass data was used to generate experimental growth rates for *L. monocytogenes* growth in beef frankfurter and sliced turkey breast slurries at 4, 8, and 12°C. The variable temperatures represent a typical range of consumer refrigeration temperatures. These experimental growth rates were used to calculate strain specific scalar factors for use by GEMs simulating growth at different temperatures (Tables 36 and 37). These additional scalar factors were then implemented into the FBA algorithm for low temperature growth simulations. Current GEMs are validated and optimized for use at a single growth temperature, so the ability to scale biomass predictions with relation to temperature would improve the functionality of the models as predictive tools for scientific inquiry and for the field of food safety microbiology.

Comparing the low-temperature growth rates to experimental growth data yielded interesting results. The time at which the low-temperature growth rate accurately reflected the experimental data was variable with temperature and product, but there were some observable trends. For both beef frank and sliced turkey breast slurry simulations, an increase in temperature resulted in a decrease in time until the predictions were most accurate. For example, the beef frank slurry low-temperature *in silico* predictions were most accurate at 97.4%, 64.6%, and 38.9% of the experimental runtimes for 4°C, 8°C, and 12°C, respectively. Similarly, the sliced turkey breast slurry predictions were most

accurate at 70.6%, 41.7%, and 12.5% of the experimental runtimes for 4°C, 8°C, and 12°C, respectively.

There are a number of possible reasons for the lack of a strong, significant correlation between the low-temperature *in silico* predictions of growth rate and the experimental data. One reason is that the *in silico* predictions cannot account for the initial decrease in biomass production that was seen in a number of the experimental results. This is because the *in silico* predictions assume that *L. monocytogenes* will grow at the maximum possible growth rate right from the beginning, which is not always seen experimentally. Perhaps more likely, though, is the fact that the GEMs created in this study would require additional constraints to account for decreased reaction kinetics at lower temperatures. The *L. monocytogenes* GEMs created in this study, utilize traditional FBA, which assumes optimal growth temperature. The optimal growth temperature is likely near the optimum temperature for the enzymatic metabolic reaction kinetics. However, by decreasing the temperature, a deviation from this optimum reaction temperature would occur, resulting in slowed reaction kinetics. This deviation can be accounted for using a Thermodynamics-based Flux Balance Analysis (TFBA) approach, which would add additional constraints to the GEMs, to account for thermodynamic variation [63-65]. The fact that one of the GEMs already shows a correlation to the experimental data, and the observed trend of most accurate time varying with temperature, indicate that the GEMs created in this study provide a good starting point for future studies using a TFBA approach to generate accurate low-temperature *in silico* growth predictions.

3.6 Conclusion

Quantitatively validating the GEMs as laboratory tools was done using batch growth experiments. These experiments yielded the first *L. monocytogenes* data on optical density, viable cell concentration, and biomass. This data was used to calculate conversion factors between the measurement types and to generate growth rates and growth yields for each of the six strains of *L. monocytogenes*. The growth rate data, when compared to *in silico* predictions by the GEMs, resulted in scalar factors used to

adjust the models to agree with experimental data. The scalar factor adjustments resulted in growth rate and growth yield predictions that were much more accurate, and were significantly correlated with the experimental results. These results indicate that GEMs generated by semi-automated tools provided by KBase, followed by manual curation and experimental validation, can also become successful predictive tools for use in the laboratory.

Another goal of the quantitative validation process was to use the conversion factors generated by the batch growth experiments to convert previously published experimental data to a format usable by the GEMs. By doing this, the *L. monocytogenes* GEMs could then be used with *in silico* methods from the COBRA toolbox at different conditions, such as lower temperatures relevant to the food industry. Unfortunately, the lack of a strong, significant correlation between *in silico* predictions and experimental data indicates that a simple scalar conversion factor is not sufficient to account for different temperatures. However, the success of the model predictions at optimal growth temperature, the significant correlation in turkey breast slurry at 8°C, and the trend that the accuracy of *in silico* predictions is dependent on time and temperature indicate that the models generated in this study would be good starting points for future studies using TFBA to predict low-temperature growth *in silico*.

4 Summary

Overall, this study shows that the semi-automated genome-scale metabolic reconstruction pipelines provided by the ModelSeed framework used by KBase are acceptable tools for the creation of draft versions of GEMs. With manual curation and experimental validation, these GEMs can be adjusted to successful laboratory and food safety tools. The GEMs generated from this study agree well with qualitative experimental results and their predictive accuracy is comparable to previously published GEMs. Additionally, the GEMs created in this study show strong, significant correlations between *in silico* predictions and experimental results of growth rates growth yields at optimum growth temperature. Therefore, this set of *L. monocytogenes* GEMs can be used in similar fashion to other published GEMs and provide ideal starting points

for future iterations for continued expansion of the complexity of *L. monocytogenes* GEMs. This set of *L. monocytogenes* GEMs also can serve as a template for future reconstructions of GEMs of other *Listeria* strains or species, or other microorganisms.

References

1. Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson MA, Roy SL, Jones JL, Griffin PM: **Foodborne illness acquired in the United States--major pathogens**. *Emerging infectious diseases* 2011, **17**(1):7-15.
2. Murray EGD, Webb RA, Swann MBR: **A disease of rabbits characterised by a large mononuclear leucocytosis, caused by a hitherto undescribed bacillus Bacterium monocytogenes (n. sp.)**. *The Journal of Pathology and Bacteriology* 1926, **29**(4):407-439.
3. Seeliger HP: **Listeriosis--history and actual developments**. *Infection* 1988, **16 Suppl 2**:S80-84.
4. Dworkin M, Falkow S: **The prokaryotes : a handbook on the biology of bacteria**, vol. 4, 3rd edn. New York ; London: Springer; 2006.
5. Weller D, Andrus A, Wiedmann M, den Bakker HC: **Listeria booriae sp. nov. and Listeria newyorkensis sp. nov., from food processing environments in the USA**. *International journal of systematic and evolutionary microbiology* 2015, **65**(Pt 1):286-292.
6. Boone DR, Castenholz RW, Garrity GM: **Bergey's manual of systematic bacteriology**, 2nd edn. New York: Springer; 2001.
7. Welshimer HJ: **Vitamin Requirements of Listeria Monocytogenes**. *Journal of bacteriology* 1963, **85**:1156-1159.
8. Premaratne RJ, Lin WJ, Johnson EA: **Development of an improved chemically defined minimal medium for Listeria monocytogenes**. *Applied and environmental microbiology* 1991, **57**(10):3046-3048.
9. Orsi RH, den Bakker HC, Wiedmann M: **Listeria monocytogenes lineages: Genomics, evolution, ecology, and phenotypic characteristics**. *International journal of medical microbiology : IJMM* 2011, **301**(2):79-96.
10. Doumith M, Cazalet C, Simoes N, Frangeul L, Jacquet C, Kunst F, Martin P, Cossart P, Glaser P, Buchrieser C: **New aspects regarding evolution and virulence of Listeria monocytogenes revealed by comparative genomics and DNA arrays**. *Infection and immunity* 2004, **72**(2):1072-1083.
11. Call DR, Borucki MK, Besser TE: **Mixed-genome microarrays reveal multiple serotype and lineage-specific differences among strains of Listeria monocytogenes**. *Journal of clinical microbiology* 2003, **41**(2):632-639.
12. Datta AR, Laksanalamai P, Solomotis M: **Recent developments in molecular sub-typing of Listeria monocytogenes**. *Food additives & contaminants Part A, Chemistry, analysis, control, exposure & risk assessment* 2013, **30**(8):1437-1445.
13. Kathariou S: **Listeria monocytogenes virulence and pathogenicity, a food safety perspective**. *Journal of food protection* 2002, **65**(11):1811-1829.
14. Ferreira V, Wiedmann M, Teixeira P, Stasiewicz MJ: **Listeria monocytogenes persistence in food-associated environments: epidemiology, strain characteristics, and implications for public health**. *Journal of food protection* 2014, **77**(1):150-170.
15. Ramaswamy V, Cresence VM, Rejitha JS, Lekshmi MU, Dharsana KS, Prasad SP, Vijila HM: **Listeria--review of epidemiology and pathogenesis**. *Journal of*

microbiology, immunology, and infection = Wei mian yu gan ran za zhi 2007, **40**(1):4-13.
16. Fairchild TM, Foegeding PM: **A proposed nonpathogenic biological indicator for thermal inactivation of Listeria monocytogenes**. *Applied and environmental microbiology* 1993, **59**(4):1247-1250.
17. Kamat AS, Nair PM: **Identification of Listeria innocua as a biological indicator for inactivation of L-monocytogenes by some meat processing treatments**. *Food Sci Technol-Leb* 1996, **29**(8):714-720.
18. Liu SW, Puri VM, Demirci A: **Evaluation of Listeria innocua as a suitable indicator for replacing Listeria monocytogenes during ripening of Camembert cheese**. *Int J Food Sci Tech* 2009, **44**(1):29-35.
19. Buchrieser C: **Biodiversity of the species Listeria monocytogenes and the genus Listeria**. *Microbes and infection / Institut Pasteur* 2007, **9**(10):1147-1155.
20. National Center for Biotechnology Information. **NCBI Genome Database.** [http://www.ncbi.nlm.nih.gov/genome]. Accessed 1 Aug 2016.
21. Oberhardt MA, Palsson BO, Papin JA: **Applications of genome-scale metabolic reconstructions**. *Molecular systems biology* 2009, **5**:320.
22. Bordbar A, Monk JM, King ZA, Palsson BO: **Constraint-based models predict metabolic and associated cellular functions**. *Nature reviews Genetics* 2014, **15**(2):107-120.
23. O'Brien EJ, Monk JM, Palsson BO: **Using Genome-scale Models to Predict Biological Capabilities**. *Cell* 2015, **161**(5):971-987.
24. McCloskey D, Palsson BO, Feist AM: **Basic and applied uses of genome-scale metabolic network reconstructions of Escherichia coli**. *Molecular systems biology* 2013, **9**:661.
25. Baumler DJ, Peplinski RG, Reed JL, Glasner JD, Perna NT: **The evolution of metabolic networks of E. coli**. *BMC systems biology* 2011, **5**:182.
26. Monk JM, Charusanti P, Aziz RK, Lerman JA, Premyodhin N, Orth JD, Feist AM, Palsson BO: **Genome-scale metabolic reconstructions of multiple Escherichia coli strains highlight strain-specific adaptations to nutritional environments**. *Proceedings of the National Academy of Sciences of the United States of America* 2013, **110**(50):20338-20343.
27. Raghunathan A, Reed J, Shin S, Palsson B, Daefler S: **Constraint-based analysis of metabolic capacity of Salmonella typhimurium during host-pathogen interaction**. *BMC systems biology* 2009, **3**:38.
28. Becker SA, Palsson BO: **Genome-scale reconstruction of the metabolic network in Staphylococcus aureus N315: an initial draft to the two-dimensional annotation**. *BMC microbiology* 2005, **5**:8.
29. Lee DS, Burd H, Liu J, Almaas E, Wiest O, Barabasi AL, Oltvai ZN, Kapatral V: **Comparative genome-scale metabolic reconstruction and flux balance analysis of multiple Staphylococcus aureus genomes identify novel antimicrobial drug targets**. *Journal of bacteriology* 2009, **191**(12):4015-4024.

30. Kim HU, Kim SY, Jeong H, Kim TY, Kim JJ, Choy HE, Yi KY, Rhee JH, Lee SY: **Integrative genome-scale metabolic analysis of Vibrio vulnificus for drug targeting and discovery**. *Molecular systems biology* 2011, **7**:460.
31. Hartman HB, Fell DA, Rossell S, Jensen PR, Woodward MJ, Thorndahl L, Jelsbak L, Olsen JE, Raghunathan A, Daefler S *et al*: **Identification of potential drug targets in Salmonella enterica sv. Typhimurium using metabolic modelling and experimental validation**. *Microbiology* 2014, **160**(Pt 6):1252-1266.
32. Freitag NE, Port GC, Miner MD: **Listeria monocytogenes - from saprophyte to intracellular pathogen**. *Nature reviews Microbiology* 2009, **7**(9):623-628.
33. Schauer K, Geginat G, Liang C, Goebel W, Dandekar T, Fuchs TM: **Deciphering the intracellular metabolism of Listeria monocytogenes by mutant screening and modelling**. *BMC genomics* 2010, **11**:573.
34. Bochner BR: **Global phenotypic characterization of bacteria**. *FEMS microbiology reviews* 2009, **33**(1):191-205.
35. Darling AC, Mau B, Blattner FR, Perna NT: **Mauve: multiple alignment of conserved genomic sequence with rearrangements**. *Genome research* 2004, **14**(7):1394-1403.
36. Darling AE, Mau B, Perna NT: **progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement**. *PloS one* 2010, **5**(6):e11147.
37. Nelson KE, Fouts DE, Mongodin EF, Ravel J, DeBoy RT, Kolonay JF, Rasko DA, Angiuoli SV, Gill SR, Paulsen IT *et al*: **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 research* 2004, **32**(8):2386-2395.
38. Evans MR, Swaminathan B, Graves LM, Altermann E, Klaenhammer TR, Fink RC, Kernodle S, Kathariou S: **Genetic markers unique to Listeria monocytogenes serotype 4b differentiate epidemic clone II (hot dog outbreak strains) from other lineages**. *Applied and environmental microbiology* 2004, **70**(4):2383-2390.
39. Ludeke CH, Kong N, Weimer BC, Fischer M, Jones JL: **Complete Genome Sequences of a Clinical Isolate and an Environmental Isolate of Vibrio parahaemolyticus**. *Genome announcements* 2015, **3**(2).
40. Briers Y, Klumpp J, Schuppler M, Loessner MJ: **Genome sequence of Listeria monocytogenes Scott A, a clinical isolate from a food-borne listeriosis outbreak**. *Journal of bacteriology* 2011, **193**(16):4284-4285.
41. Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL: **High-throughput generation, optimization and analysis of genome-scale metabolic models**. *Nature biotechnology* 2010, **28**(9):977-982.
42. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M *et al*: **The RAST Server: rapid annotations using subsystems technology**. *BMC genomics* 2008, **9**:75.

43. Becker SA, Feist AM, Mo ML, Hannum G, Palsson BO, Herrgard MJ: **Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox**. *Nature protocols* 2007, **2**(3):727-738.
44. Orth JD, Thiele I, Palsson BO: **What is flux balance analysis?** *Nature biotechnology* 2010, **28**(3):245-248.
45. Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, Palsson BO: **A comprehensive genome-scale reconstruction of Escherichia coli metabolism--2011**. *Molecular systems biology* 2011, **7**:535.
46. Henry CS, Zinner JF, Cohoon MP, Stevens RL: **iBsu1103: a new genome-scale metabolic model of Bacillus subtilis based on SEED annotations**. *Genome biology* 2009, **10**(6):R69.
47. Oh YK, Palsson BO, Park SM, Schilling CH, Mahadevan R: **Genome-scale reconstruction of metabolic network in Bacillus subtilis based on high-throughput phenotyping and gene essentiality data**. *The Journal of biological chemistry* 2007, **282**(39):28791-28799.
48. Price ND, Reed JL, Palsson BO: **Genome-scale models of microbial cells: evaluating the consequences of constraints**. *Nature reviews Microbiology* 2004, **2**(11):886-897.
49. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BO: **A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information**. *Molecular systems biology* 2007, **3**:121.
50. Archer CT, Kim JF, Jeong H, Park JH, Vickers CE, Lee SY, Nielsen LK: **The genome sequence of E. coli W (ATCC 9637): comparative genome analysis and an improved genome-scale reconstruction of E. coli**. *BMC genomics* 2011, **12**:9.
51. AbuOun M, Suthers PF, Jones GI, Carter BR, Saunders MP, Maranas CD, Woodward MJ, Anjum MF: **Genome scale reconstruction of a Salmonella metabolic model: comparison of similarity and differences with a commensal Escherichia coli strain**. *The Journal of biological chemistry* 2009, **284**(43):29480-29488.
52. Charusanti P, Chauhan S, McAteer K, Lerman JA, Hyduke DR, Motin VL, Ansong C, Adkins JN, Palsson BO: **An experimentally-supported genome-scale metabolic network reconstruction for Yersinia pestis CO92**. *BMC systems biology* 2011, **5**:163.
53. Liao YC, Huang TW, Chen FC, Charusanti P, Hong JS, Chang HY, Tsai SF, Palsson BO, Hsiung CA: **An experimentally validated genome-scale metabolic reconstruction of Klebsiella pneumoniae MGH 78578, iYL1228**. *Journal of bacteriology* 2011, **193**(7):1710-1717.
54. Oberhardt MA, Puchalka J, Fryer KE, Martins dos Santos VA, Papin JA: **Genome-scale metabolic network analysis of the opportunistic pathogen Pseudomonas aeruginosa PAO1**. *Journal of bacteriology* 2008, **190**(8):2790-2803.

55. Puchalka J, Oberhardt MA, Godinho M, Bielecka A, Regenhardt D, Timmis KN, Papin JA, Martins dos Santos VA: **Genome-scale reconstruction and analysis of the Pseudomonas putida KT2440 metabolic network facilitates applications in biotechnology**. *PLoS computational biology* 2008, **4**(10):e1000210.
56. Pinchuk GE, Hill EA, Geydebrekht OV, De Ingeniis J, Zhang X, Osterman A, Scott JH, Reed SB, Romine MF, Konopka AE *et al*: **Constraint-based model of Shewanella oneidensis MR-1 metabolism: a tool for data analysis and hypothesis generation**. *PLoS computational biology* 2010, **6**(6):e1000822.
57. Reed JL, Vo TD, Schilling CH, Palsson BO: **An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR)**. *Genome biology* 2003, **4**(9):R54.
58. Phan-Thanh L, Gormon T: **A chemically defined minimal medium for the optimal culture of Listeria**. *Int J Food Microbiol* 1997, **35**(1):91-95.
59. Sutton VR, Kiley PJ: **Techniques for studying the oxygen-sensitive transcription factor FNR from Escherichia coli**. *Methods in enzymology* 2003, **370**:300-312.
60. Varma A, Palsson BO: **Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110**. *Applied and environmental microbiology* 1994, **60**(10):3724-3731.
61. Kumar S, Stecher G, Tamura K: **MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets**. *Molecular biology and evolution* 2016, **33**(7):1870-1874.
62. Pal A, Labuza TP, Diez-Gonzalez F: **Comparison of primary predictive models to study the growth of Listeria monocytogenes at low temperatures in liquid cultures and selection of fastest growing ribotypes in meat and turkey product slurries**. *Food microbiology* 2008, **25**(3):460-470.
63. Jankowski MD, Henry CS, Broadbelt LJ, Hatzimanikatis V: **Group contribution method for thermodynamic analysis of complex metabolic networks**. *Biophysical journal* 2008, **95**(3):1487-1499.
64. Fleming RM, Thiele I, Nasheuer HP: **Quantitative assignment of reaction directionality in constraint-based models of metabolism: application to Escherichia coli**. *Biophysical chemistry* 2009, **145**(2-3):47-56.
65. Henry CS, Broadbelt LJ, Hatzimanikatis V: **Thermodynamics-based metabolic flux analysis**. *Biophysical journal* 2007, **92**(5):1792-1805.

Appendix A: PLOS One Article

PLOS | ONE

RESEARCH ARTICLE

Predicting Essential Metabolic Genome Content of Niche-Specific Enterobacterial Human Pathogens during Simulation of Host Environments

Tong Ding[1☯], Kyle A. Case[1☯], Morrine A. Omolo[1☯], Holly A. Reiland[1☯], Zachary P. Metz[1☯], Xinyu Diao[1☯], David J. Baumler[1,2,3*]

1 Department of Food Science and Nutrition, University of Minnesota-Twin Cities, St. Paul, Minnesota, United States of America, **2** Microbial and Plant Genomics Institute, University of Minnesota-Twin Cities, St. Paul, Minnesota, United States of America, **3** Biotechnology Institute, University of Minnesota-Twin Cities, St. Paul, Minnesota, United States of America

☯ These authors contributed equally to this work.
* dbaumler@umn.edu

OPEN ACCESS

Citation: Ding T, Case KA, Omolo MA, Reiland HA, Metz ZP, Diao X, et al. (2016) Predicting Essential Metabolic Genome Content of Niche-Specific Enterobacterial Human Pathogens during Simulation of Host Environments. PLoS ONE 11(2): e0149423. doi:10.1371/journal.pone.0149423

Editor: Yousef Abu Kwaik, University of Louisville, UNITED STATES

Received: November 4, 2015

Accepted: January 31, 2016

Published: February 17, 2016

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The authors received no specific funding for this work.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

Microorganisms have evolved to occupy certain environmental niches, and the metabolic genes essential for growth in these locations are retained in the genomes. Many microorganisms inhabit niches located in the human body, sometimes causing disease, and may retain genes essential for growth in locations such as the bloodstream and urinary tract, or growth during intracellular invasion of the hosts' macrophage cells. Strains of *Escherichia coli* (*E. coli*) and *Salmonella* spp. are thought to have evolved over 100 million years from a common ancestor, and now cause disease in specific niches within humans. Here we have used a genome scale metabolic model representing the pangenome of *E. coli* which contains all metabolic reactions encoded by genes from 16 *E. coli* genomes, and have simulated environmental conditions found in the human bloodstream, urinary tract, and macrophage to determine essential metabolic genes needed for growth in each location. We compared the predicted essential genes for three *E. coli* strains and one *Salmonella* strain that cause disease in each host environment, and determined that essential gene retention could be accurately predicted using this approach. This project demonstrated that simulating human body environments such as the bloodstream can successfully lead to accurate computational predictions of essential/important genes.

Introduction

Computational modeling has been widely used as an efficient approach in microbiology, which introduces mathematical components including variables, parameters, and equations in network constructions to reflect the behavior of organisms. Numerous types of networks have been constructed including signaling, regulatory, and metabolic pathways for organisms

ranging from microorganisms, such as *E. coli*, to multi-cellular eukaryotic organisms. By constructing genome-scale metabolic models (GEMs), the nature of an organism can be explored through computational analysis of its genome content. The *E. coli* K-12 strain MG1655 has had extensive computational metabolic networks generated for it so far, and its existing models are quite advanced that contain >2,000 reactions, >1,000 genes, and >1,000 metabolites [1–6]. These genome-scale models have been used for many studies that have guided the engineering of strains for increasing valuable end-products, promoting enzyme discovery, providing insight into the genome evolution of other enterobacteria [7,8], and leading to a new understanding of the connectivity, or coupling, of all the metabolic reactions and corresponding genes within the cell.

Currently, numerous *E. coli* metabolic networks have been constructed for commensal, enterohemorrhagic, and extra intestinal pathogenic strains [1,4]. Unlike studies using *E. coli* metabolic models, a *Salmonella* Typhimurium LT2 metabolic model was used to examine metabolic reactions and the corresponding essential genes that are necessary for cell viability during the infection process under simulated conditions inside the host [9]. The evolutionary process that leads to genome changes is based on the theory of natural selection, which states that in a given environmental niche, there is constant pressure to retain genes that are important for growth and survival in that particular condition. When the availability of nutrients in a host-cell environment can be used to further define the mathematical constraints for the metabolic model mimicking host-cell nutrient environment, a technique termed flux balance analysis (FBA) was used that identified 417 reactions used by *S. typhimurium* LT2 during human infection [9].

To systematically explore genes predicted as essential and important for cell growth in a given environment, we used an approach that focused on three main components: 1) generating a metabolic network and corresponding metabolic model representing the metabolic capabilities of the *E. coli* pangenome which contains the union of all genes that encode metabolic reactions from 16 genomes of *E. coli*, 2) using flux balance analysis to systematically test growth predictions in three simulated host environments of all single gene mutants, and 3) comparing the essential/important gene predictions (i.e. those that promote growth and would likely have been retained over time) with sequenced enterobacterial genomes to determine if these genes were retained or lost in modern day strains.

In this work, we have developed new methods using constraint-based optimization and metabolic model construction to identify genes important for growth/survival in environments simulating three locations within the human body and have compared the predictions with actual evolutionary outcomes of sequenced genomes of enterobacterial pathogens, such as extraintestinal *E. coli*, that cause human disease in locations other than the intestinal tract. Extraintestinal *E. coli* infections may result in serious illness and even death, and globally 130–175 million cases of urinary tract infections are caused by Extraintestinal *E. coli* [10]. The urinary tract is also the most common route for *E. coli* causing bloodstream infections, which cause more than 40,000 deaths from septicemia each year worldwide [10]. Therefore, an understanding of the genes that are essential for the growth of these pathogens to survive in certain human body niches is of great interest to aid efforts on developing new control strategies and therapeutics.

Computational modeling allows us to conduct experiments of disease-causing bacteria where actual testing in humans is not an option. These are the three main objectives that were investigated: i) Can different locations in the human body be modeled using constraint-based linear programming? ii) Are there different predictions of essential/important genes for growth in simulated conditions representing three human body locations? iii) Do these gene predictions correlate with the genome content of modern-day enterobacterial pathogens that actually

cause disease in each of the three locations? Overall, this study illustrated that mathematical constraints can be used with metabolic models to simulate the nutrient conditions the pathogen encounters during the infection process, and the genes predicted using FBA with the metabolic model simulating conditions during infection correlate with transcriptional gene-expression data obtained for conditions representing host-pathogen interactions. The central hypothesis is that the essential and important genes for bacterial growth in certain environments should be mostly remained over time in the genome of strains that cause disease in the corresponding human body locations, whereas the loss of those essential and important genes should not cause dire consequence for strains that invade different human locations.

Results and Discussion

Computational simulation of different niches in the human body

For the three simulated conditions, analytical data were used to add constraints that dictate metabolite availabilities respectively under three simulated conditions, the human macrophage cell [9], the bloodstream [11], and the urinary tract [12]. During the macrophage invasion, the pathogens can be engulfed and chained inside the pathogen-containing vacuoles that may restrict nutrients for cell growth. There is very little information on the nutrient compositions of those vacuoles under different macrophage activation states. Considering the pathogens may achieve nutrients from cytoplasm by modifying the membrane of vacuoles, existing literature values on the nutrient composition of the macrophage cytoplasm can be used to mimic the environment inside a macrophage for pathogen growth.

For the three simulated niches examined in human body, there were 15 available metabolites used as constraints shared in common for all three host niches, whereas 51 metabolites varied depending on the environment, indicating that differences in human body locations lead to different metabolite compositions available to the microrganisms (Table 1).

Predictions of essential/important genes for cell growth in three simulated human body locations

When FBA analysis for single reaction deletions and their corresponding genes was conducted in the three simulated environments, the results varied in the total number of predicted essential and important reactions and associated genes for each condition (Table 2). Following each gene deletion, if the rate of biomass production was calculated as a value of zero (no growth prediction) or a reduction of >1% of the wild type biomass production, the genes were considered to be essential or important, respectively. There were 38 reactions predicted to be commonly essential for all three simulated human body locations, as the absence of them led to no cellular growth (Fig 1). Besides, 38 reactions were predicted as essential that were not shared in common for those conditions (Fig 1). There was only one reaction predicted to be important that resulted in a decrease of predicted biomass for all three simulated host locations, whereas 121 reactions were predicted as important that led to a predicted biomass reduction in one or two simulated conditions (S1 Data). For all of these essential and important reactions the genes correspond to, the reactions were identified to report the number of essential or important genes' lost (S2 Data).

Comparison of essential/important gene predictions based on the genomes of real disease-causing enterobacterial pathogens in each of the three host niches

Once the essential and important genes were identified, they were compared with the sequenced genomes of enterobacterial pathogens that invade the macrophage cell, infect the

Table 1. Nutrients used to simulate three host environmental conditions.

Metabolites	Macrophage	Blood	Urine
2-Oxoglutarate	-	+	-
Acetoacetate	-	+	-
Adenine	-	-	+
Adenosine	-	+	-
Allantoin	+	+	+
Arabinose	+	-	-
Butyrate	-	+	+
Carnitine	+	-	-
Citrate	-	+	+
Cytosine	+	-	-
Deoxycytidine	+	-	-
Ethanolamine	+	-	+
Formate	-	-	+
Fructose	+	-	-
Fucose	+	-	-
Fumarate	-	+	-
Galactarate	+	-	-
Galactonate	+	-	-
Glucarate	+	-	-
Gluconate	+	-	-
Glucosamine	-	+	-
Glucose	+	+	+
Glucuronate	+	+	+
Guanine	-	-	+
Hypoxanthine	+	-	-
Inosine	+	-	-
D-lactate	-	+	+
L-lactate	-	+	+
L-Malate	-	+	-
D-Malate	-	+	-
Maltose	+	-	-
Mannitol	+	-	-
Mannose	+	-	-
Melibiose	+	-	-
Myo-Inositol	-	+	+
N-Acetyl-D-glucosamine	+	-	-
N-Acetylneuraminate	+	-	-
Nicotinate	-	+	-
Pantothenate	+	-	-
Propane-1,2-diol	+	-	-
Putrescine	+	-	-
Pyruvate	-	+	+
Rhamnose	+	-	-
Ribose	+	-	-
Sorbitol	+	-	-
Spermidine	+	-	-
Succinate	-	+	-

(*Continued*)

Table 1. (*Continued*)

Metabolites	Macrophage	Blood	Urine
Taurine	-	-	+
Thiamin	+	+	-
Uracil	+	-	-
Uridine	+	-	-

Present / Not Present = + / -

doi:10.1371/journal.pone.0149423.t001

bloodstream, or cause disease in the urinary tract. Three genomes (*E. coli* UTI89, *E. coli* 53638, and *Salmonella* LT2) were used for essential and important gene comparison, and the genome of *E. coli* O157:H7 was used as a control because of the pathogen's capability to cause disease in the human intestine. *E. coli* UTI89 is able to infect the urinary tract or the bloodstream in human body, causing disease outside the intestinal track. Both *E. coli* 53638 and *Salmonella* LT2 can cause disease by invasion of a host cell (Table 3).

The central hypothesis is that the pathogens that actually cause disease in a given host location should have lost the fewest number of essential and important genes predicted for that conditions simulated *in silico* (macrophage, bloodstream, or urinary tract). In contrast, the pathogenic *E. coli* O157:H7 that causes disease in the intestinal tract would most likely have lost the most number of essential and important genes predicted for each of the three host niches. The host niche condition was not simulated for the control in this project. As shown in Table 4, when compared to the genomes of these organisms, the number of lost essential and important genes in each strain varied. When the numbers of both lost predicted essential and important genes out of the total number are summarized (Table 5), it is clear that some of the predictions match the real evolutionary outcomes of the genome content of these organisms, whereas the simulation of the urinary tract did not match the evolutionary outcomes of these strains, and this discrepancy is addressed in the conclusions section.

Conclusions

This study investigated *in silico* metabolic modeling and prediction of genes required for growth and survival in three human body locations. Based on the numerous differences of metabolites present in three different human body niches, this study illustrates that multiple environmental niches in a human can be simulated to study microbial metabolism by using constraint-based linear programming and computational model. Simulation of these three conditions led to different predictions of essential and important genes/reactions, which match the real evolutionary outcomes when compared to the control genome of the intestinal pathogen enterohemorrhagic *E. coli* O157:H7 strain EDL933, a strain that causes disease in the intestine and was predicted to have lost the most of the essential or important genes in the three other host niches. In the case of intracellular invasion, although the strain isolated from a

Table 2. Total number of reactions and corresponding genes predicted as essential and important for growth in three simulated human body locations.

Host niche	Essential reactions	Important reactions	Essential genes	Important genes
Macrophage	195	146	290	146
Bloodstream	193	65	288	182
Urinary tract	203	52	304	151

doi:10.1371/journal.pone.0149423.t002

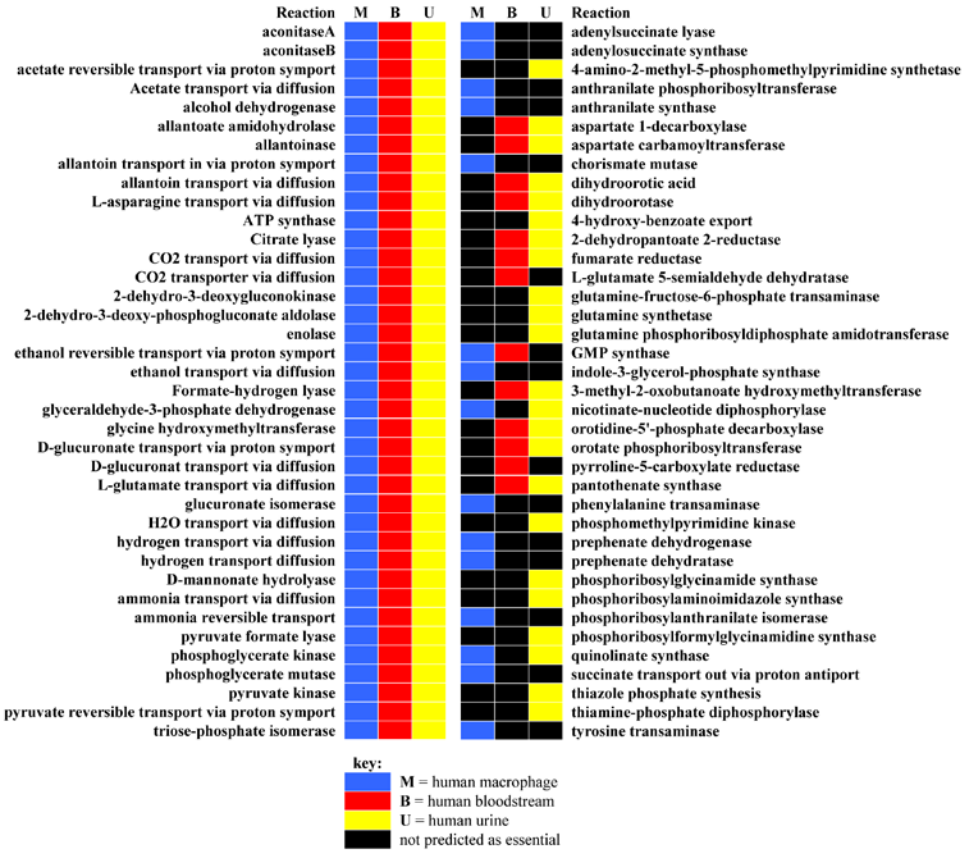

Fig 1. Essential reactions predicted for three simulated host environmental conditions. There are 38 reactions predicted to be commonly essential for all three simulated human body locations, whereas 38 essential reactions predicted that are differed for simulations of the human bloodstream, urinary tract, and macrophage.

doi:10.1371/journal.pone.0149423.g001

Table 3. *E. coli* and *Salmonella* genomes used in this study.

Host niche	Enterobacterial human pathogenic strains	Genome of strain that causes disease
Bloodstream	Extraintestinal pathogenic *E. coli*	*E. coli* UTI89
Macrophage	*Salmonella* spp., Enteroinvasive *E. coli*	*E. coli* 53628, *Salmonella* LT2
Urinary tract	Urinary tract pathogenic *E. coli*	*E. coli* UTI89
Intestinal tract (control)	Enterohemorrhagic *E. coli*	*E. coli* EDL933

doi:10.1371/journal.pone.0149423.t003

Table 4. Total number of predicted essential and important genes lost out of total predicted for each strain.

Host Niche	Genes Lost/Total Predicted	*E. coli* 53638	*E. coli* UTI89	*Salmonella* LT2	*E. coli* O157:H7
Macrophage	essential genes	2/290	4/290	3/290	13/290
Macrophage	important genes	20/366	17/366	22/366	58/366
Bloodstream	essential genes	2/288	1/288	3/288	12/288
Bloodstream	important genes	14/182	12/182	18/182	26/182
Urinary tract	essential genes	2/304	6/304	4/304	12/304
Urinary tract	important genes	9/151	11/151	15/151	19/151

doi:10.1371/journal.pone.0149423.t004

urinary tract infection has the fewest essential/important genes lost, the two genomes of strains that actually cause disease through this route had very similar low numbers of lost essential/important genes. In the case of the simulations for the human bloodstream and urinary tract, *E. coli* UTI89 is the strain that actually causes disease in these locations, and had the least amount of necessary/important genes lost, which agreed with the evolutionary outcome. This project demonstrated that human body environments such as the bloodstream can successfully lead to accurate predictions of essential/important genes using optimization and constraint-based metabolic techniques. The discrepancies from the predictions for the urinary tract may indicate that more information is required for additional constraints to more accurately simulate this environment, or that the *E. coli* strains that have been characterized as causing disease in only one niche in the human body may also be capable of causing disease in numerous locations in the human body. Overall, this project was a success and lays a foundation towards future work to model metabolism of pathogenic microbes in different locations inside a human host. Since the actual infection study of these organisms in human is not a possibility, computer modeling of related disease processes becomes an emerging approach and field that is likely to grow immensely. By addressing these research ideas revealed by this project using optimization and constraint-based linear programming, the field of microbial system biology can be furthered to efficiently examine genome evolution.

Materials and Methods

Pangenome Metabolic Network Reconstruction

The metabolic model representing the *E. coli* pangenome (iEco1712_pan) used in this work was previously reconstructed based on the gene to protein to reaction (GPR) information of 16 *E. coli* genomes obtained from the ASAP database [1]. Draft and complete genomes have been continually updated using new publicly accessible genomes in the ASAP database since its inception [13]. There currently are 39 genomes among more than 150 enterobacteria genomes in the ASAP database that belong to *E. coli*, of which 16 are completely finished and were used

Table 5. Total number of predicted essential and important genes lost out of total predicted for each strain.

Host niche	*E. coli* 53638	*E. coli* UTI89	*Salmonella* LT2	*E. coli* EDL933 (control)
Macrophage	22/656[c]	21/656[c]	25/656[c]	71/656[a]
Bloodstream	16/470[a]	13/470[a]	21/470[a]	38/470[a]
Urinary tract	11/455[b]	17/455[b]	19/455[b]	31/455[a]

[a]Evolutionary outcome agrees with *in silico* predictions for genome content

[b]Evolutionary outcome disagrees with *in silico* predictions for genome content

[c]Evolutionary outcome is within standard deviation with *in silico* predictions for genome content

doi:10.1371/journal.pone.0149423.t005

in the construction of the metabolic model of *E. coli* pangenome (iEco1712_pan) [1]. The reconstructed network contains metabolic enzymes present in a union of 76,080 Open Reading Frames (ORFs) that map 17,647 Clusters of Orthologous Groups (COGs), with each ORF being assigned to an COG in the ASAP database, and all of the information for model composition, GPR associations for the *E. coli* pangenome (iEco1712_pan) reconstruction used in this work are available as supplemental information along with the sbml file for the iEco1712_pan GEM [1].

Flux Balance Analysis

Flux balance analysis (FBA) has been commonly applied for mathematical analysis of GEMs, which can predict reactions-related fluxes in a metabolic network [14]. By constraining fluxes with steady-state mass balances, reaction directionality, and metabolite availability, a range of possible flux values can be generated in FBA. An objective function then can be used to identify flux distributions that maximize (or minimize) the objective function with those constraints. Biomass production, a commonly used objective function for FBA performance and for a proxy of growth, was adapted in this study [15]. FBA was conducted using the software package GAMS in this study, in which the *E. coli* pangenome metabolic network is described as a stoichiometric matrix ($\mathbf{S_{i,j}}$) with rows (i∈I) representing the metabolites and columns (j∈J) indicating reactions that correspond to genes (g∈G). In a steady-state, the mass balance equation can be described as below, with v being the flux vector. Additional constraints are showed as lower and upper limits for the values of fluxes through reactions in a network.

$$\begin{aligned} &Max\ Vbiomass \\ &\text{s.t. } \mathbf{S}_{ij} \bullet \mathbf{v} = 0 \\ &v_{j,lb} < v_j < v_{j,up} \end{aligned} \tag{1}$$

The matrix built for *E. coli* pangenome GEM contains 1,726 metabolites (I) and 2,324 reactions (J) that associate with 1,712 genes (G). Three different niches located in the human body (macrophage, blood, and urinary tract) were simulated to set constraints for FBA in this study, with possible metabolite compositions being identified through literature review that determined analytical compositions of nutrients present in each bodily location. The simulated condition for macrophage contains 32 metabolites, the bloodstream environment contains 19 metabolites, while there are 14 metabolites that belong to the urinary tract niche (Table 1).

Gene Essentiality

Unlike virulence factor genes [16], essential genes are those required to maintain critical cellular functions under specific environments, while important genes are not irreplaceable but still necessary for robust bacterial growth under those conditions. To determine the essentiality of genes expressed under different environmental pressures (macrophage cell, bloodstream, and urine tract), genes were removed one-by-one in networks and the resulting changes in biomass production rate can be estimated to reveal the impact of gene loss (a proxy for fitness). Following each gene deletion, if the calculated value of biomass production rate was zero, meaning no predicted intracellular growth, the gene would be considered essential. Important genes were predicated based on >1% reduction of the wild type biomass production rate. A graphic description on identifying essential genes and corresponding metabolic reactions using GEMs constructing and computational predictions is showed in Fig 2.

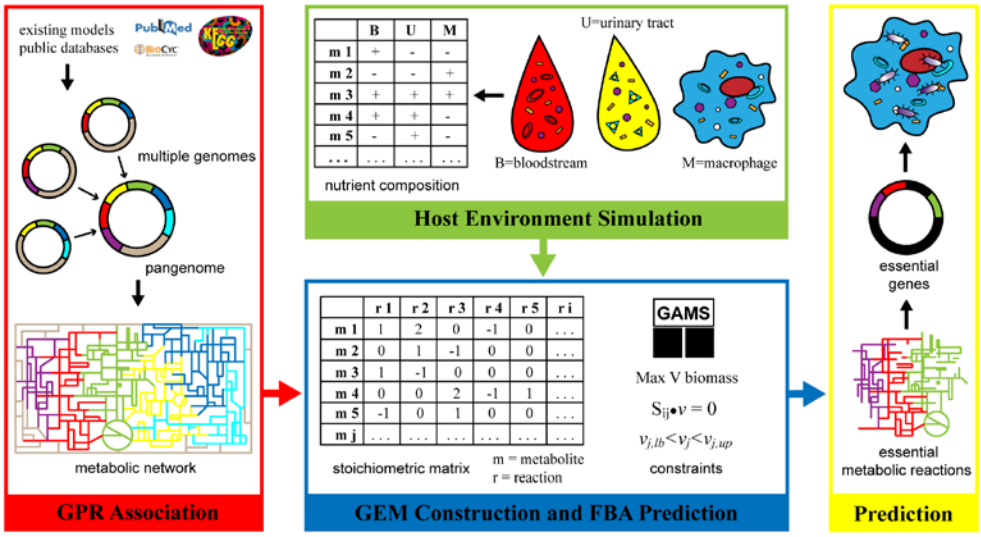

Fig 2. Essential gene identification using GEMs predictions under simulated environment. The GEM constructed upon pangenome incorporated from *E. coli* genomes can be used to generate predations with simulated nutrient conditions to identify essential genes along with corresponding essential metabolic reactions under multiple human body niches.

doi:10.1371/journal.pone.0149423.g002

Supporting Information

S1 Data. Reactions predicted as important for all three simulated environmental conditions and that differed during simulation of the human bloodstream, urinary tract, and macrophage.
(XLSX)

S2 Data. Reactions corresponding to essential gene predictions for the *E. coli* pangenome GEM. This file contains three tables, the first contains all predicted essential reactions during simulation of human macrophage, the second contains all human bloodstream predicted essential reactions, and the third contains predicted essential reactions during simulation of the human urinary tract.
(XLSX)

Acknowledgments

We would also like to thank Dr(s). William R. Harcombe, Guy Plunkett III, Bob Mau, and Eric Cabot for insightful discussions regarding gene essentiality within the genomes of members of the family *Enterobacteriaceae*. This work was partially funded by the Department of Food Science and Nutrition and the College of Food, Agricultural and Natural Resource Sciences at the University of Minnesota-Twin Cities (DJB), and partially by the Global Food Ventures Graduate Student Fellowship from the University of Minnesota-Twin Cities (ZPM), and the Schlumberger Faculty for the Future Graduate Student Fellowship (MAO).

Author Contributions

Conceived and designed the experiments: DB. Performed the experiments: TD KC MO HR ZM XD. Analyzed the data: TD KC MO HR ZM XD DB. Wrote the paper: TD KC MO HR ZM XD DB.

References

1. Baumler D, Peplinski R, Reed J, Glasner J, Perna N (2011) The evolution of metabolic networks of E. coli. BMC Systems Biology 5: 1–21.
2. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, et al. (2007) A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol 3: 121. PMID: 17593909
3. Feist AM, Palsson BO (2008) The growing scope of applications of genome-scale metabolic reconstructions using Escherichia coli. Nat Biotechnol 26: 659–667. doi: 10.1038/nbt1401 PMID: 18536691
4. Monk JM, Charusanti P, Aziz RK, Lerman JA, Premyodhin N, Orth JD, et al. (2013) Genome-scale metabolic reconstructions of multiple Escherichia coli strains highlight strain-specific adaptations to nutritional environments. Proceedings of the National Academy of Sciences 110: 20338–20343.
5. Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, et al. (2011) A comprehensive genome-scale reconstruction of Escherichia coli metabolism—2011. Mol Syst Biol 7: 535. doi: 10.1038/msb.2011.65 PMID: 21988831
6. Reed JL, Vo TD, Schilling CH, Palsson BO (2003) An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR). Genome Biology 4: R54–R54. PMID: 12952533
7. Baumler DJ, Ma B, Reed JL, Perna NT (2013) Inferring ancient metabolism using ancestral core metabolic models of enterobacteria. BMC Syst Biol 7: 46. doi: 10.1186/1752-0509-7-46 PMID: 23758866
8. Pal C, Papp B, Lercher MJ, Csermely P, Oliver SG, Hurst LD, et al. (2006) Chance and necessity in the evolution of minimal metabolic networks. Nature 440: 667–670. PMID: 16572170
9. Raghunathan A, Reed J, Shin S, Palsson B, Daefler S (2009) Constraint-based analysis of metabolic capacity of Salmonella typhimurium during host-pathogen interaction. BMC Syst Biol 3: 38. doi: 10.1186/1752-0509-3-38 PMID: 19356237
10. Russo TA, Johnson JR (2003) Medical and economic impact of extraintestinal infections due to Escherichia coli: focus on an increasingly important endemic problem. Microbes Infect 5: 449–456. PMID: 12738001
11. Keitel HG, Berman H, Jones H, Maclachlan E (1955) The Chemical Composition of Normal Human Red Blood Cells, including Variability among Centrifuged Cells. Blood 10: 370–376. PMID: 14363319
12. Putnam DF (1971) Composition and concentrative properties of human urine. McDonnell Douglas Astronautics Company. #CR-1802. 1–112 p.
13. Glasner JD, Rusch M, Liss P, Plunkett G, Cabot EL, Darling A, et al. (2006) ASAP: a resource for annotating, curating, comparing, and disseminating genomic data. Nucleic Acids Research 34: D41–D45. PMID: 16381899
14. Orth JD, Thiele I, Palsson BØ (2010) What is flux balance analysis? Nature biotechnology 28: 245–248. doi: 10.1038/nbt.1614 PMID: 20212490
15. Feist AM, Palsson BO (2010) The biomass objective function. Curr Opin Microbiol 13: 344–349. doi: 10.1016/j.mib.2010.03.003 PMID: 20430689
16. Reiland HA, Omolo MA, Johnson TJ, Baumler DJ. (2014) A survey of Escherichia coli O157:H7 virulence factors: The first 25 years and 13 genomes. Advances in Microbiology. 4:7.