ABSTRACT

Title of Dissertation: MACHINE LEARNING AND GENOMICS FOR IMPROVED FOOD SAFETY AND RISK ASSESSMENT OF *SALMONELLA ENTERICA* IN CHICKEN

Edmund Ofosu Benefo, Doctor of Philosophy, 2024

Dissertation directed by: Professor Abani K. Pradhan,
Department of Nutrition and Food Science

Salmonella enterica is a leading cause of foodborne illnesses worldwide and is commonly associated with poultry. *Salmonella* has many closely related serovars, yet these serovars exhibit significant variability in many characteristics including host range, virulence, growth behavior, stress response, and antimicrobial resistance. In the past, this intricate and dynamic population heterogeneity of *Salmonella* severely hampered control efforts, but, today, this has improved through the sequencing of *Salmonella* genomes. Whole genome sequencing (WGS) provides a better understanding of the evolutionary and ecological adaptations that underlie the survival of *Salmonella* against antimicrobials, oxidative agents, non-optimal temperatures, and other stressors in the environment and their hosts. Coupling machine learning with WGS expands on these advantages by enabling the identification of genetic patterns that may not be immediately apparent. The overall goal of this research was to explore how machine learning and genomics can be integrated to improve food safety. First, a machine learning model was developed to identify stress response genes in *Salmonella* isolated from different poultry processing stages. It was found that

beyond genes encoding for cold and heat shock proteins, other genes involved in lipopolysaccharide biosynthesis, DNA repair and replication, and biofilm formation are involved in *Salmonella*'s overall stress response mechanism. Additionally, a machine learning model was developed to predict antimicrobial resistance (AMR) phenotypes in *Salmonella* isolates using WGS data. The model predictions were comparable to existing bioinformatic methods for AMR prediction and identified AMR genes that are typically not the resistance determinants public health agencies focus on. Expanding this approach for AMR surveillance could lead to the discovery of novel AMR genes. Lastly, a quantitative microbial risk assessment for *Salmonella* in chicken that incorporated *Salmonella* heterogeneity in growth and virulence was developed. The findings revealed that variations in virulence have a greater impact on the risk of salmonellosis than variations in growth rate. Overall, this research contributes to efforts to enhance food safety measures and reduce chicken-associated *Salmonella* illnesses.

MACHINE LEARNING AND GENOMICS FOR IMPROVED FOOD SAFETY AND RISK ASSESSMENT OF *SALMONELLA ENTERICA* IN CHICKEN

by

Edmund Ofosu Benefo

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2024

Advisory Committee:
Dr. Abani K. Pradhan, Chair
Dr. Ryan Blaustein
Dr. Seong-Ho Lee
Dr. Jianghong Meng
Dr. Nidhi Rawat

© Copyright by
Edmund Ofosu Benefo
2024

Dedication

I dedicate this dissertation to my family who have been my support system all my life. I am a product of their love, care, generosity, and guidance. My father, R. B. Baning, was my first teacher, mentor, and motivator. My stepmother, J. Darko, was my friend and nurturer. Both are long gone but never forgotten. My mother, M. Ofori, and siblings Kudjo, Frank, Mike, Prince, Sarah, and Richard have been a constant source of support throughout this journey. My wife, Dromo, for always being there for me, providing unwavering support and motivation, and even downloading some of my ComBase data for me.

Acknowledgments

I am immensely grateful to my advisor, Dr. Abani Pradhan. I still remember my email to him asking about opportunities in his research group and his kind response. Five years on, I have greatly benefited from his mentorship, guidance, resources, support, and many life lessons. I am grateful to thank Dr. Ryan Blaustein, Dr. Seong-Ho Lee, Dr. Jianghong Meng, and Dr. Nidhi Rawat for their expertise and contributions in shaping my research. I would also like to thank my colleagues, past and present, in the Pradhan Lab, Dr. Surabhi Rani, Dr. Shraddha Karanth, Dr. Collins Tanui, Taryn Horr, Aishwarya Rao, Shuyi Feng, Padmini Ramachandran, Ujjwol Subedi, Thabang Msimango, Samiksha Bhattarai, and Gracielle Pereira, for their support and help in many ways. I am thankful to Dr. Debasmita Patra for the opportunities she provided for me to work on projects with her, as well as the many doses of encouragement. I am thankful to my BSc. and MPhil. advisor, Dr. Isaac W. Ofosu, who got me interested in food safety and has kept motivating me since then. I am thankful to Kwesi Sam and the many friends I have shared memorable moments with throughout my Ph.D. journey.

Table of Contents

Dedication ii
Acknowledgments iii
Table of Contents iv
List of Tables vii
List of Figures viii
List of Abbreviations ix

Chapter 1: Introduction 1
1.1 *Salmonella enterica* in chicken: contamination pathways, prevalence, and outbreaks 1
1.2 *Salmonella* population heterogeneity 3
1.3 *Salmonella* antimicrobial resistance and stress response and mechanisms 5
1.4 Genomics and food safety 14
1.5. Machine learning applications in food safety 16
1.6 Food safety risk assessment: past, present, and future 18
1.7 Project overview 21

Chapter 2: A machine learning approach to identifying *Salmonella* stress response genes in isolates from poultry processing 23
2.1 Abstract 23
2.2 Introduction 24
2.3 Materials and Methods 25
2.3.1 Isolates collection 25
2.3.2 Data labeling 26
2.3.3 Bioinformatics analysis 26
2.3.4 Machine learning 28
2.3.4.1. Data exploration 28
2.3.4.2 Model building and evaluation 30
2.4 Results and Discussion 30
2.4.1 *Salmonella* strains 30
2.4.2 Machine learning modeling 31

2.4.3 Important *Salmonella* genes associated with stress response 34
2.5 Conclusion 40

Chapter 3: **Genome-based machine learning for predicting antimicrobial resistance in *Salmonella* isolated from chicken** 41
3.1 Abstract 41
3.2 Introduction 42
3.3 Materials and Methods 45
3.3.1 Isolates collection 45
3.3.2 Bioinformatics analysis and AMR gene identification 45
3.3.3 Machine learning 46
3.3.3.1 Data labeling and exploration 47
3.3.3.2 Model building and prediction 49
3.4 Results and Discussion 50
3.4.1 Strain characteristics 50
3.4.2 Machine learning modeling 50
3.4.3 AMR phenotype prediction and gene identification 52
3.5 Conclusion 57

Chapter 4: Exploring the impact of population heterogeneity on risk estimates for *Salmonella enterica* in chicken 58
4.1 Abstract 58
4.2 Introduction 59
4.3 Materials and Methods 62
4.3.1 Study overview 62
4.3.2 *Salmonella* growth data 63
4.3.3 Finite mixture modeling 64
4.3.4 Secondary growth modeling 67
4.3.5 Baseline QMRA model 68
4.3.5.1 Farm 68
4.3.5.2 Processing 76
4.3.5.3 Retail storage 78
4.3.5.4 Transportation to home 79

4.3.5.5 Home storage 80
4.3.5.6 Cooking 81
4.3.5.7 Dose-response modeling and risk characterization 82
4.3.6 Scenario models 83
4.3.7 Model simulation and sensitivity analysis 85
4.4 Results and Discussion 85
4.4.1 Finite mixture modeling of *Salmonella* growth data 85
4.4.2 *Salmonella* levels across the chicken farm-to-fork chain 90
4.4.3 Risk estimates and sensitivity analysis 92
4.5 Conclusions 98

Chapter 5: Summary and future studies 100

References 103

List of Tables

Table 2.1 Model performance and optimum tuning parameters for the six machine learning algorithms used in this study……...……...……...…..33

Table 2.2 The twenty most important genes associated with stress responses in *Salmonella* isolates from poultry processing……..……35

Table 3.1 Number of resistant and susceptible isolates used for machine learning modeling…….48

Table 3.2 Machine learning model performance across eight antimicrobials studied.……...…51

Table 3.3 Comparison of predictions from machine learning modeling and the ResFinder database……..54

Table 3.4 Important genes used by machine learning models for classification……..……...……56

Table 4.1 Summary of variables and distributions used in the baseline QMRA model.....70

Table 4.2 Conditions for the baseline and scenario QMRA models…….....…….85

Table 4.3 Mixing proportions, mean, variance, and BIC values of various components identified from computer-assisted mixture analysis (finite mixture modeling) ……....…….87

Table 4.4 Probability of illness per serving and number of illnesses per year due to the consumption of *Salmonella*-contaminated chicken…….……..94

List of Figures

Fig. 2.1 Typical stages of chicken processing. Part A of the cycle indicates the stages where *Salmonella* encounters pre-chill stresses including heat, oxidative, and acid stresses and Part B indicates the stages where *Salmonella* encounters cold, desiccation, and starvation stresses…….27

Fig. 2.2 Receiver operating characteristic (ROC) curves for random forest (RF), cost-sensitive learning (CSL), and neural network (NNET) models…….32

Fig. 2.3 Receiver operating characteristic (ROC) curves for support vector machine radial kernel (SVMR), support vector machine linear kernel (SVML), and logit boost (LB) models……..32

Fig. 3.1 Machine learning workflow for AMR phenotype prediction…….47

Fig. 3.2 Machine learning predictions of AMR prevalence in 200 *Salmonella* isolates……..53

Fig. 4.1 Histogram and density plot of the growth rates of *Salmonella* isolates……..86

Fig. 4.2 Density plot showing the mixture components of Model C……...89

Fig. 4.3 *Salmonella* growth rate during countertop thawing……...90

Fig. 4.4 Simulated *Salmonella* levels at the end of each stage in the baseline QMRA model……..91

Fig. 4.5 Tornado plot showing the model input variables with the largest impact on the number of illnesses per year……98

List of Abbreviations

ADS	Arginine Decarboxylase System
ALE	Assembly Likelihood Evaluation
AMR	Antimicrobial Resistance
ANOVA	Analysis of Variance
ASP	Acid Shock Protein
AUROC	Area Under the Receiver-Operating Characteristic
BIC	Bayesian Information Criterion
BLAST	Basic Local Alignment Search Tool
BV-BRC	Bacterial and Viral Bioinformatics Resource Center
CARD	Comprehensive Antibiotic Resistance Database
CDC	Centers for Disease Control and Prevention
CSL	Cost-sensitive Learning
FAO	Food And Agriculture Organization
FDA	Food and Drug Administration
FSIS	Food Safety Inspection Service
HSP	Heat Shock Protein
IDBA	Iterative De Bruijn Graph De Novo Assembler
IFSAC	Interagency Food Safety Analytics Collaboration
KPI	Key Performance Indicator
LB	Logit Boost
LHR	Locus of Heat Resistance
NARMS	National Antimicrobial Resistance Monitoring System
NNET	Neural Network
PATRIC	Pathosystems Resource Integration Center
QMRA	Quantitative Microbial Risk Assessment
RF	Random Forest
RNAP	RNA Polymerase
ROSE	Random Over-Sampling Examples
SPI	*Salmonella* Pathogenicity Island
SVML	Support Vector Machine Linear Kernel
SVMR	Support Vector Machine Radial Kernel
USDA	U.S. Department of Agriculture
VEM	Variational Expectation-Maximization
WGS	Whole Genome Sequencing
WHO	World Health Organization
WOAH	World Organization for Animal Health
XGB	Extreme Gradient Boosting

Chapter 1: Introduction

1.1 *Salmonella enterica* in chicken: contamination pathways, prevalence, and outbreaks

Chicken can become infected with *Salmonella enterica* through many sources such as breeder flocks, litter, pests, and feces. The hatchery, litter, and feces have been reported as the most important factors contributing to *Salmonella* prevalence in chickens (Wang et al., 2023). In hatcheries, *Salmonella* contamination of breeder eggs can occur transovarially (vertical transmission) when birds have an infected reproductive system, thus *Salmonella* contamination of eggs occurs internally prior to eggshell formation and laying (Shaji et al., 2023). Contamination may also occur via horizontal transmission where *Salmonella* penetrates the eggshells when eggs are being laid, during incubation, or hatching (Raut et al., 2023; Shaji et al., 2023). In either form of transmission, the hatched chicks could become long-term carriers of *Salmonella* (Gast et al., 2024). Once *Salmonella* colonizes the gastrointestinal tract of chickens, it can be shed through feces. Other birds in the same enclosure can then be infected through the fecal-oral route. Chickens are usually reared on litter, a mixture of wood shavings, straw, and other materials, which serve as bedding. *Salmonella* survival in litter is largely influenced by the moisture content of the litter. Contaminated litter is a significant predictor of *Salmonella* presence in birds, especially in newly hatched chicks as they are more susceptible to colonization (Wang et al., 2023). Additionally, chicken feed and water have also been implicated in *Salmonella* contamination of chickens. Insects such as cockroaches and litter beetles are known vectors of *Salmonella*, and chickens can become infected by eating these insects (Leffer et al., 2010). Rodents have also been reported to be a significant source of *Salmonella* through their feces which chickens may consume or when they contaminate chicken feed (Meerburg & Kijlstra, 2007).

The prevalence of *Salmonella* in chickens is, expectedly, higher on farms. A recent meta-analysis by Wang et al. (2023) on broiler production systems in the U.S. estimated *Salmonella* prevalence of 12.5%, 25.4%, 48.8%, 1.2%, and 6.5% in feces, litter, hatchery, water, and chicks, respectively. During chicken processing, due to the inactivation effect of heat and processing aids, the prevalence of *Salmonella* generally reduces. Beginning with a prevalence of 60% during the bleeding of chickens, Rivera-Pérez et al. (2014) observed a decrease to 10% during the chilling stage. A similar trend has been reported by Golden and Mishra (2020), who estimated a prevalence of 42.49% at the rehang stage (before evisceration) and 14.3% at the chilling stage. Routine sampling reports from the Food Safety Inspection Service (FSIS) of the U.S. Department of Agriculture (USDA) from 2019 to 2023 indicate that 128,702 chicken samples were taken from processing establishments and 11,302 samples tested positive for *Salmonella*, thus a prevalence of 8.78% (USDA FSIS, 2024a). A similar review of U.S. Food and Drug Administration (FDA) National Antimicrobial Resistance Monitoring System (NARMS) data shows that 30,130 retail chicken samples were analyzed between 2017 and 2021, with 2,946 samples testing positive for *Salmonella*, representing a prevalence of 9.78% (FDA, 2024).

Salmonella outbreaks are a function of many factors including the prevalence and concentration in chicken, the infectious dose of *Salmonella*, the immunity status of consumers, and preparation and consumption practices, among others. The Interagency Food Safety Analytics Collaboration (IFSAC) attributes 18.6% of *Salmonella* illnesses in the U.S. to chicken (IFSAC 2023). Over the last decade, there have been 88 chicken-associated *Salmonella* outbreaks in the U.S., resulting in 2,442 illnesses, 359 hospitalizations, and 4 deaths (CDC, 2024a). These outbreaks were widespread across the U.S., occurring in all but 3 states, North Dakota, South Dakota, and Mississippi.

1.2 *Salmonella* population heterogeneity

The *Salmonella* genus has two species, *enterica* and *bongori* (Eng et al., 2015). The *S. enterica* species is further divided into six subspecies namely, *enterica* (I), *salamae* (II), *arizonae* (IIIa), *diarizonae* (IIIb), *houtenae* (IV), and *indica* (VI), based on genome similarity and biochemical properties (Foley et al., 2013). *S. enterica* subsp. *enterica* is largely found in mammals while the other subspecies as well as *S. bongori* are found in cold-blooded animals and the environment (Eng et al., 2015). Beyond the subspecies, *Salmonella* is further classified into serovars based on the presence of three surface antigens, somatic O antigens, flagellar H antigens, and K capsular antigens (Eng et al., 2015; Ryan et al., 2017). Over 2600 serovars exhibiting considerable heterogeneity have been identified, with more than half of them belonging to *S. enterica* subsp. *enterica* (Eng et al., 2015; Foley et al., 2013; Ryan et al., 2017). Despite all serovars sharing the same core genome, significant variations exist in their pathogenicity, host range, and other physiological profiles (Cheng et al., 2019).

While some serovars such as *S.* Enteritidis, *S*. Heidelberg, *S*. Typhimurium, and *S*. Newport, have a wide host range and infect several host species, others such as *S*. Choleraesuis, *S*. Gallinarum, *S*. Abortusovis, and *S*. Dublin have limited host ranges (Foley et al., 2013). A review of FSIS quarterly sampling reports indicates a difference in host preference for *Salmonella* serovars (USDA FSIS, 2024a). While serovars Kentucky, Infantis, Enteritidis, and Typhimurium tend to be predominantly found in chicken samples, serovars Montevideo, Muenchen, Anatum, and Dublin dominate beef samples. Even for turkeys, which are evolutionarily related to chickens, a different set of serovars, Schwarzengrund, Reading, Agona, and Hadar, are more commonly isolated from turkeys. Host preference is influenced by several factors including variations in the host environment (attachment site, pH, nutrient availability, temperature), host immune response

to different serovars, host native commensal microbiota, and most importantly, the genetics of the serovar itself (Foley et al., 2013).

Salmonella serovars exhibit differences in virulence as well. A vast majority of the more than 2600 serovars are not known to cause human infections. Serovars Infantis, Enteritidis, and Typhimurium are mostly associated with human illnesses in the U.S. while serovar Kentucky rarely causes human illnesses (Richards et al., 2023; USDA FSIS, 2022). The consensus on *S.* Kentucky is that it lacks many known virulence genes making it less harmful to humans. Tate et al. (2022) reported on the absence of virulence genes including *sseK2* (Type III secretion system effector), *lpfD* (fimbrial protein), and *tcf* (fimbrial protein-associated operon) in certain sequence types of *S.* Kentucky resulting in a lowered ability to invade, adhere, colonize, and survive in humans. Other serovars, such as Dublin and Choleraesuis, although responsible for only a small number of human illnesses, tend to cause invasive illnesses due to the presence of a type 6 secretion system that facilitates their invasion into human cells (Cheng et al., 2019). Many *Salmonella* virulence genes are located within the *Salmonella* pathogenicity island (SPI); as such differences in the SPIs can result in different virulence levels and disease outcomes. An example can be seen in *S*. Typhimurium and *S.* Typhi, where SPIs-7, 15, 17, and 18 are present in the latter but absent from the former (Sabbagh et al., 2010). These SPIs contain several genes that are necessary for invasion into epithelial cells, an increase of pathogen uptake by macrophages, and Vi capsule biosynthesis, and are thought to be responsible for *S.* Typhi's ability to cause a more severe illness than *S*. Typhimurium (Sabbagh et al., 2010).

Differences in stress response and growth capacity have also been reported. Bacterial heterogeneity in stress adaptation typically follows two schemes: one where a subpopulation

expresses stress response genes even before changes in the environment occur (bet-hedging) and another where the heterogeneity is caused by changes in the environment resulting in differences in gene expression by bacteria in the population (Davis & Isberg, 2016). Capita (2007) observed that *S.* Derby, *S.* Typhimurium, and *S.* Newport were more resistant to decontamination on poultry carcasses using sodium chlorite as compared to serovars Enteritidis, Infantis, Virchow, Agona, and Poona. Pye et al. (2023) also observed that *Salmonella* serovars exhibit varying growth rates when exposed to desiccation, acid, salt, heat, and cold stresses. Additionally, Díez-García et al. (2012) reported on differences in the growth kinetic parameters and biofilm-forming abilities of *Salmonella* serovars isolated from poultry. Heterogeneity in *Salmonella* has been shown to not only occur between (inter) serovars but also within (intra) the same serovar (Aspridou et al., 2019; Pye et al., 2023; Papagianeli et al., 2022). Given the diversity in *Salmonella* populations, they should be seen as a group of individuals rather than a homogenous population (Staes et al., 2019).

1.3 *Salmonella* antimicrobial resistance and stress response and mechanisms

Antimicrobial resistance mechanisms

Salmonella possesses a complex array of AMR mechanisms including the reduction of membrane permeability, modification of antimicrobial target site, removal of antimicrobial by transporters, enzymatic inactivation, and substitution of target proteins. *Salmonella* has an asymmetric lipid bilayer outer membrane that is thought to be involved in AMR (Zhou et al., 2023). The outer membrane contains lipopolysaccharides which have hydrophilic carbohydrate components. These hydrophilic components have been found to form hydrated spheres that decrease the movement and permeability of hydrophobic molecules across the outer membrane (Wu et al., 2015). Additionally, antimicrobials such as beta-lactams that target cell wall synthesis

are usually unable to cross the outer membrane due to efficient lipid packing which also reduces membrane fluidity (Chaudari et al., 2023; Zhou et al., 2023). *Salmonella* can also modify membrane permeability by downregulating the expression of outer membrane porin proteins, thus reducing the porin-mediated transport of antimicrobials into the cell (Y. Wang et al., 2022). Du et al. 2020 observed a reduction in the expression of outer membrane porin F in ciprofloxacin-resistant *Salmonella*.

Efflux pumps promote AMR by exporting antimicrobial compounds outside the cell membrane, thus reducing the concentration of antimicrobial compounds in the cell (Y. Wang et al., 2022). *Salmonella* has efflux pump systems including AcrAB, AcrEF, AcrD, MdsABC, MdtABC, EmrAB, and MdtK, with the AcrAB-TolC (resistance nodulation division family) being one of the most important for *Salmonella* AMR (Chaudari et al., 2023; Y. Wang et al., 2022; Zhou et al., 2023). The AcrAB-Tolc efflux pump is regulated by the RamA and MarA (Chaudari et al., 2023). Ricci et al. (2006) found that mutant *S.* Typhimurium strains that lacked *marA*, *ramA*, *acrB,* and *tolC* genes had a reduced resistance to ciprofloxacin.

Salmonella AMR can also be mediated by antimicrobial-inactivating enzymes. Resistance to beta-lactams, for example, is caused by beta-lactamases (encoded by the *blaTEM* gene) cleaving the amide bond in beta-lactams **(**Wright, 2005**)**. Resistance in aminoglycosides is also mediated by phosphotransferases, acyltransferases, and adenyltransferases, which modify portions of the aminoglycoside compound, thus rendering it inactive (Frye & Jackson, 2013; Zhou et al., 2023). Furthermore, *Salmonella* can modify the binding sites for antimicrobials, resulting in a reduced susceptibility to these agents. Fluoroquinolones, for instance, destroy *Salmonella* by inhibiting topoisomerase IV (*parCE*) and DNA gyrases (*gyrAB*), which are necessary for DNA replication.

Mutations in these genes have been found to result in reduced susceptibility or resistance to fluoroquinolones (Zhang et al., 2017).

Temperature stress response

Salmonella detects changes in temperature through a complex system of thermosensors including proteins, lipids, and thermoresponsive RNAs located in the *Salmonella* 5′ untranslated region (Dawoud et al., 2017). Exposure to cold temperatures results in the halting of *Salmonella* growth and adaptation for survival. *Salmonella* cold shock response occurs in a series of stages: acclimatization (10 °C to 30 °C) during which cold shock proteins (CSPs) are synthesized, then recovery, and finally a permanent adaptation to the reduced temperature (Ricke et al., 2018). During cold shock, transcriptional and translational processes limit the function of the ribosome leading to inadequate protein synthesis and folding, and finally a reduction of growth rate (Chattopadhyay, 2006). *Salmonella* then adjusts the transcription and translation functions toward the production of CSPs (encoded by *cspABEH* genes) and DNA supercoiling (Barria et al., 2013; Shah et al., 2013). These proteins are essential for DNA repair and translation processes that allow *Salmonella* to grow at cold temperatures. *Salmonella* also adapts to low temperatures by modifying fatty acid chain lengths and the unsaturated/saturated fatty acid ratio of its cell membrane to decrease membrane fluidity and protect the cell (Yoon et al., 2015).

Exposure to higher-than-optimal temperatures initiates the expression of responses in two ways: a cytoplasmic heat shock response which is controlled by the heat shock sigma factor σ^H and an extracytoplasmic heat shock response controlled by sigma factor σ^E (Spector & Kenyon, 2012). At temperatures below 30 °C, intramolecular hydrogen bonding prevents translation of the *rpoH* mRNA, while at temperatures above 42 °C, the *rpoH* mRNA structure unfolds enabling

ribosomal binding and translation (Morita et al., 1999). At high temperatures, the DnaK-DnaJ chaperone complex binds to misfolded proteins and allows binding and stability of the RNAP-σ^H complex, thus increasing heat shock proteins (HSPs) production. σ^H then directs the transcription of several small HSPs which function as proteases and molecular chaperones (Guisbert et al., 2008).

The RNAP-σ^E complex also directs the transcription of proteases, periplasmic chaperones, and other factors which aid in reducing the effects of the increased temperature on the extracytoplasmic compartment (Ades, 2008). Furthermore, σ^E has been reported to be involved in the regulation of the expression of cell envelope factors required for the prevention and repair of damage caused by high temperatures (Ades, 2008). Recently, Mercer et al. (2017) found that the heat resistance of *S. enterica* is enhanced by a locus of heat resistance (LHR). When they exposed *S.* Typhimurium and *S.* Senftenberg strains to a temperature of 60 °C for 5 min, a 7-log reduction was observed in LHR-negative strains as compared to less than 1-log reduction in LHR-positive strains.

Oxidative stress response

Salmonella is exposed to several oxidizing agents in food processing environments. Two major oxidative stress response pathways that have been studied are the OxyR and SoxRS pathways. OxyR forms a protein complex with subunits containing cysteine residues that act as oxidative stress sensors (Storz & Zheng, 2000). The exposure of *S.* Typhimurium to hydrogen peroxide triggers conformational changes in OxyR, promoting its functions as a transcriptional activator. Oxidized OxyR induces the transcription of genes that function in the protection of DNA, formation of disulfide linkages, breakdown of hydrogen peroxide, and reduction of oxidized

lipids (Storz & Zheng, 2000). OxyR also activates the expression of small RNAs and the *fur* gene. The *fur* gene produces Fur proteins which prevent the formation of hydroxy radicals through the Fenton reaction by limiting iron uptake (Storz & Zheng, 2000). The small RNAs (RyhB-1 and RyhB-2) modulate iron usage by down-regulating iron-containing proteins such as aconitase and fumarase; this reduces the availability of iron to partake in the Fenton reaction, thus reducing the formation of hydroxy radicals (Calderón et al., 2014). To maintain a redox balance, OxyR is reduced by the OxyR-regulated GrxA using glutathione as the reductant. Reduced OxyR then represses the transcription of *oxyR* (Calderón et al., 2014).

The SoxRS pathway is used when oxidative stress is caused by a redox-cycling compound. SoxRS forms a protein dimer with each unit containing an iron-sulfur center which is oxidized in the presence of redox cycling compounds (Demple et al., 1999). The oxidation of SoxR activates the transcription of *soxS* which induces the expression of several genes (Wu & Weiss, 1991). Some of the genes include *sodA* which catalyzes the degradation of superoxide; *nfsA* which prevents superoxide formation; *nfo* for DNA damage repair, *tolC, micF,* and *acrAB* for efflux of redox cycling compounds from the cell; *fldA, fldB,* and *fpr* for maintenance of the reduced iron-sulfur centers; and *zwf* for an increase of the reducing power within the cell (Storz & Zheng, 2000). Storz and Zheng (2000) also described other regulons (*katE, xthA,* and *sodC* encoding hydroperoxidase II, exonuclease III, and periplasmic superoxide dismutase respectively) that play a role in alleviating oxidative stress.

Acid stress response

Álvarez-Ordóñez et al. (2010) described an arginine decarboxylase (ADS) and lysine decarboxylase system (LDS) that acts as a major player in *S.* Typhimurium intracellular pH control. The LDS is made up of a CadA (lysine decarboxylase enzyme), CadB (lysine-cadaverine antiporter), and CadC (a regulator of the *cadBA* operon). CadC senses a reduction in the extracellular pH and activates the transcription of *cadBA*. CadA then catalyzes the conversion of intracellular lysine to cadaverine using a proton, thus increasing the pH of the cell. Cadaverine is then transported out of the cell and lysine into the cell using the CadB antiporter. In addition, CadC has been reported to act as a global regulator linked to the EnvZ/OmpR two-component regulatory system which regulates the expression of proteins during exposure of *S.* Typhimurium to low pH (Lee et al., 2007). This highlights the importance of the LDS in protecting *Salmonella* at low pH conditions.

The ADS follows a similar mechanism as the LDS described above, however, using a different set of enzymes and operons. Arginine is converted to agmatine by AdiA (arginine decarboxylase) using a proton leading to an increase in pH. Agmatine is then transported out of the cell in exchange for arginine via an AdiC antiporter (Kieboom & Abee, 2006). Kieboom and Abee (2006) reported an upregulation of *adiY, adiA,* and *adiC* in acid-stressed *S.* Typhimurium whereas mutant cells with disrupted genes showed a decreased acid resistance. They suggested that the arginine decarboxylase system is only induced under anaerobic conditions, however, a later study by Álvarez-Ordóñez et al. (2010) found that it can also be induced under aerobic conditions.

The acid stress response of *Salmonella* is enhanced by the induction of acid shock proteins (ASPs). Foster (2001) identified four ASPs that are involved in *S.* Typhimurium acid tolerance response: an iron regulatory protein Fur, a two-component signal transduction system PhoPQ, alternative sigma factor σ^S, and OmpR. The alternative sigma factor σ^S has been shown to be involved in acid tolerance response by regulating the expression of more than 10 ASPs which protect the cell from acid stress (Foster, 2001). The Fur protein, though usually associated with iron uptake, has been shown to control some ASPs involved in *Salmonella* acid stress response by conferring protection against organic acids (Hall & Foster, 1996).

Membrane composition modification has been studied as one of the responses to a low-pH environment. It has been reported that when *S.* Typhimurium is exposed to low pH, it undergoes a membrane modification by reducing its saturated fatty acid to unsaturated ratio while increasing methylenoctadecenoic and octadecenoic acid ratio (Álvarez-Ordóñez et al., 2008). These modifications result in decreased membrane fluidity and increased tolerance to acid stress.

Desiccation and osmotic stress response

Though *Salmonella* requires a high water activity (about 0.99) for growth, several studies have reported on the presence of *Salmonella* in low-water activity foods; this has been attributed to the induction of responses to low-water activity environments. When *Salmonella* is exposed to desiccation or osmotic stress, there is a loss of water, shrinkage of the cell, and an increase in the concentration of intracellular metabolites (Csonka, 1989). Continued exposure to this environment results in plasmolysis and inhibition of several cellular processes. To maintain cell turgor pressure, *Salmonella* induces the production of solutes and osmoprotectants including glycine betaine, proline, potassium ions, and glutamate (Csonka, 1989). According to Csonka and Epstein (1996),

there appears to be a hierarchy in the utilization and efficiency of osmoprotectants: betaine is preferred over proline-betaine, and proline-betaine over proline.

The alternative sigma factor σ^S regulates the transcription of the *otsAB* operon, which in turn induces an increase in the synthesis of trehalose, an osmoprotectant (Csonka & Epstein, 1996; Kempf & Bremer, 1998). Fong and Wang (2016) reported an upregulation of *otsAB* following incubation of *Salmonella* in peanut oil ($a_w = 0.52$) for 6 days. The increase in trehalose synthesis is accompanied by a decrease in the concentration of K^+ and glutamate. When osmoprotectants are present externally, *Salmonella* transports them into the cell using transporters such as ProP (proline, glycine-betaine), OsmU (betaine), and ProU (glycine-betaine) (Frossard et al., 2012). Finn et al. (2013) observed an up-regulation of the *osmU* and *proU* genes in *S.* Typhimurium during desiccation; mutants lacking these genes showed a reduced survival rate.

Another response to low water activity conditions is the formation of filaments which increase desiccation resistance. Mattick et al. (2000) investigated the long-term survival and morphology changes in *Salmonella* under low water activity. They observed filament formation which was about 200 µm long in *S.* Typhimurium when the water activity was between 0.93 and 0.98. Mattick et al. (2000) argued that in response to osmotic stress, *Salmonella* inhibits cell division proteins as a means of energy conservation and instead forms filaments that allow it to increase its biomass without increasing numbers. It has been suggested that biofilm formation plays a role in the desiccation survival of *Salmonella.* Biofilms increased the survival of *S.* Tennessee stored in dried milk powder ($a_w = 0.3$) for up to 30 days (Aviles, et al., 2013). The production of curli fimbriae and cellulose, which are involved in biofilm formation, has been associated with increased long-term desiccation survival (Vestby et al., 2009).

Starvation stress response

When *Salmonella* is starved of its preferred carbon source, glucose, it undergoes genetic and physiologic changes. Kenyon et al. (2002) showed that starved *S.* Typhimurium cells are metabolically and structurally distinct from non-starved cells. During the initial stages of carbon starvation, *S.* Typhimurium counteracts starvation by up-regulating alternative carbon sources utilization and transport systems. As starvation continues, metabolic changes result in a smaller, more resilient, and physiologically efficient cell. The stress response mechanism involves the production of high-affinity transport systems for scavenging extracellular nutrients, enzymes for the metabolism of alternative carbon sources and modification of membrane components, and proteins that cause chromosome condensation to protect it from damage (Kenyon et al., 2002, 2007).

As extracellular glucose levels drop, there is a reduction in glucose uptake by *S.* Typhimurium; in response to this, cyclic 3′,5′-adenosine monophosphate (cAMP) levels increase via the activation and upregulation of adenylate cyclase (*cyaA* gene) and CRP (*crp* gene) (Görke & Stülke, 2008). Following these initial processes, putative transport and catabolic utilization systems such as carbohydrate phosphotransferase systems, periplasmic-binding protein/ABC transporter systems, ethanolamine, and propanediol are induced whether alternative carbon sources are present or not (Tsoy et al., 2009). Tsoy et al. (2009) suggested that this is done to enable rapid uptake of alternative carbon sources if they become available.

Genes such as *fadA, fadB, fadD, fadE,* and *fadI* have also been reported to be involved in the utilization of alternative carbon sources by degrading fatty acids. It has been observed that up-regulation of these genes is correlated with a decrease in cell size and long-chain monounsaturated

fatty acids content of the cell (Spector et al., 1999). This suggests the use of cell membrane fatty acids as alternative carbon sources, a form of gluconeogenesis using fatty acids like humans would do.

As a response to starvation, *Salmonella* can reduce energy utilization and undergo anaerobic respiration even when grown in aerobic conditions (Gennis & Stewart, 1996). *Salmonella* may also modify its envelope structures in response to starvation. Additionally, glucose starvation induces the *yohC-pbpG* operon which encodes for a periplasmic DD-endopeptidase and penicillin-binding protein (PBP) 7. PBP 7 produces a peptidoglycan that inhibits starvation-induced autolysis (Kenyon et al., 2007).

1.4 Genomics and food safety

Whole genome sequencing (WGS) is a technology that involves the determination of the complete sequence of an organism's genome, offering a comprehensive view of its genetic makeup. WGS provides insights into genetic variations, mutations, and structural changes that occur within an organism. WGS has been helpful in the rapid and accurate identification of foodborne pathogens, investigating antimicrobial resistance, and studying virulence properties and other genome-based attributes of pathogens (Allard et al., 2016; E. Brown et al., 2021; Cocolin et al., 2018; McDermott et al., 2016). Advancements in WGS technologies have resulted in WGS becoming more affordable and it is increasingly gaining popularity for the epidemiological surveillance of foodborne diseases. For example, the U.S. Food and Drug Administration (FDA) and Centers for Disease Control and Prevention (CDC) routinely employ WGS technology in pathogen detection and foodborne disease outbreak investigations, and this has increased the

precision and effectiveness of their surveillance responsibilities (Allard et al., 2016; E. Brown et al., 2019, 2021; Stevens et al., 2022). For every 1,000 WGS isolates uploaded to the National Center for Biotechnology Information (NCBI) Pathogen Detection database, the number of illnesses per WGS pathogen is reduced by six each year (B. Brown et al., 2021). As of October 2024, the Pathogen Detection database had more than two million isolates, with over 676,000 of them being *Salmonella* isolates (NCBI, 2024). Beyond the significant public health effects achieved by the use of WGS, the economic value associated with the reduction of illness was estimated to be almost $500 million in 2019 and was expected to increase into billions of dollars (B. Brown et al., 2021).

WGS provides insights into intra- and intra-species variations in virulence, pathogenicity, nutrient requirements, growth and inactivation rates, and other physiological processes. Furthermore, WGS technology enables a better understanding of the evolutionary and ecological adaptations that underlie the survival of foodborne pathogens against antimicrobials, oxidative agents, and other stressors in the environment and their hosts (E. Brown et al., 2021; Cocolin et al., 2018; Cook & Nightingale, 2018). This is crucial to food safety as genetic adaptations in a pathogen could reduce the effectiveness of preventive control measures. Using WGS, Hoffman et al. (2016) identified a novel arsenic resistance operon in an outbreak strain of S. Bareilly that allowed *S*. Bareilly to persist in tuna. Campioni et al. (2018, 2021) applied WGS to analyze changes in epithelial cell invasion, macrophage phagocytosis, and acid and oxidative stress survival abilities of *S*. Enteritidis isolates collected over 48 years. Genomic changes in foodborne pathogens can be vital to managing foodborne disease outbreaks, designing improved preventive controls for food processing facilities, and developing antibiotics. Further examples of WGS usage for food safety applications include the determination of antimicrobial resistance (AMR)

determinants of *Salmonella* isolates from chicken (Alzahrani et al., 2023; Tate et al 2017), characterization of virulence profiles and clonal relationships of *Salmonella* isolates from chicken (Ben Hassena et al., 2021), and characterization of phages for *Salmonella* biocontrol (Khan et al., 2024; Lu et al., 2020). WGS techniques tend to produce large amounts of complex and highly dimensional data which are more suited for, and efficiently analyzed, through machine learning.

1.5. Machine learning applications in food safety

Machine learning is a subfield of artificial intelligence that uses algorithms to interpret and learn from input data and use this learning to complete specific tasks (Benefo et al., 2022; X. Wang et al., 2022). It differs from traditional statistical methods in that it is more data-driven than rule-based and focuses on identifying patterns in data, drawing inferences from these patterns, and using them to make predictions (Rajula et al., 2020).

Some of the commonly used algorithms include random forest (RF), cost-sensitive learning (CSL), support vector machine linear (SVML) and radial kernels (SVMR), neural network (NNET), and logit boost (LB). The RF algorithm is a collection of randomized decision trees. It randomly partitions data into several decision trees based on the values of predictor variables. The predictions from these individual trees are then aggregated (to create a 'forest') and averaged, resulting in higher accuracies than any one individual tree (Biau & Scornet, 2016; Breiman, 2001). RF is widely recognized for its accuracy and capacity to handle small sample sizes. It has been found to have excellent performance when used on data such as WGS data where the number of predictors far exceeds the number of observations (Biau & Scornet, 2016; Breiman, 2001).

The LB algorithm uses an ensemble of weak models (slightly better than random models) where each new model step focuses on lowering the residual error for previously built models (Friedman et al., 2000). The LB algorithm performs classification tasks by constructing a generalized additive model and applying a logistic regression loss function via stepwise optimization (Sun et al., 2014).

The support vector machine algorithm performs classification tasks by generating and optimizing hyperplanes that can separate data into different classes (Hearst, 1998). The SVM algorithm performs classification tasks by generating hyperplanes that can separate data into different classes. The algorithm focuses on identifying the optimal hyperplane that maximizes the distance (margin) between data points (support vectors) and the plane such that any new data can be correctly classified (Hearst, 1998). Neural networks attempt to simulate the information-processing functions of human neurons. The NNET architecture comprises layers of nodes (input, hidden, layers, and output). The nodes are interconnected and have an associated weight and threshold value. When input data is passed on to a node, computations occurring at the node determine whether the data is passed on to the next layer. Data is only passed on when the computed value is greater than the threshold value of the node (Gershenson, 2003; Hardesty, 2017). CSL is particularly useful in dealing with imbalanced datasets (Thai-Nghe et al., 2010) as is usually the case when working with WGS data. Most algorithms usually only focus on minimizing the error rate (fraction of incorrect predictions) and assume all misclassification errors are equal. CSL, however, accounts for the costs of misclassification, computation, data acquisition, and active learning when training data (Ling & Sheng, 2011).

Supervised machine learning, where the algorithm learns from labeled input data, perhaps due to its utility and lower level of complexity, has been more widely used in food safety and risk-

assessment research than other machine learning categories. Du and Guo (2022) proposed a framework for a machine learning-based foodborne disease surveillance system, using data from traditional case reporting, molecular tracing networks, social media, and search engine queries to provide real-time disease outbreak prediction. Recently, Karanth et al. (2022) explored the use of WGS and supervised machine learning to classify *Salmonella* isolates based on the severity of the disease they were likely to cause. Their machine learning approach involved training the WGS data of *Salmonella* isolates using random forest, support vector machine, adaptive boosting, and logistic regression with ridge-penalty algorithms. The models performed well and were able to classify isolates as causing extraintestinal or gastrointestinal diseases with an area under the receiver operating characteristic (AUROC) curve and outcome-prediction accuracy of over 0.83 and 0.74, respectively. Logistic regression is a simple machine learning algorithm and performs poorly when used to classify highly dimensional data such as WGS data, but Karanth et al. (2022) showed that applying a ridge penalty and elastic net regularization can increase the algorithm's utility. A similar study by (Tanui et al., 2022a) developed a semi-supervised machine learning model to identify genotypic markers associated with different *Salmonella* disease endpoints (systemic infection, bacteremia, or gastroenteritis) in humans using WGS data. WGS data has been analyzed using machine learning algorithms for food source of attribution of *Listeria monocytogenes* (Tanui et al., 2022b), prediction of *Listeria monocytogenes* population behavior (Hiura et al., 2021), and virulence potential (Gmeiner et al., 2024).

1.6 Food safety risk assessment: past, present, and future

Food safety and risk assessment form the basis of ensuring the quality, integrity, and safety of food supply chains. In an era where global food supply chains are complex and vast,

safeguarding public health from potential hazards associated with foodborne illnesses is paramount. In the past, food safety systems relied on hazard-based approaches. In this approach, the presence of a hazard in food served as a basis for risk management actions (Barlow et al., 2015). The challenge with this approach is that not all hazards may constitute a significant risk meriting management. Beginning in the 1980s, the growth of the predictive microbiology field led to the development of models that enabled the characterization of pathogen growth, survival, and death (Wu & Rodricks, 2020). This led to a shift towards risk-based approaches, which are in use today. Today, there is a better understanding of the complex interactions between various factors that influence food safety. Risk assessment goes beyond the hazard-based approaches by integrating pathogen characteristics, exposure levels of pathogens, and host immune responses to provide information on risks and are then used to develop food safety policies. Risk assessment together with risk management and risk communication fit within the broader risk analysis framework.

Risk assessment is a "scientifically based process consisting of the following steps: hazard identification, hazard characterization, exposure assessment, and risk characterization" (Codex Alimentarius Commission, 2015). In quantitative microbial risk assessment (QMRA), the hazard identification step typically involves data-gathering activities to establish an association between the presence of a pathogen in food and the occurrence of illnesses due to the consumption of that food. Clinical, microbiological, and epidemiological data can provide information on the growth or decline of pathogens, pathogenesis, disease severity, susceptible populations, hospitalization rates, and morbidity rates (Haas et al., 2014; Lammerding & Fazil, 2000; Lammerding & Paoli, 1997). The data gathered here provides a risk profile of the pathogen under study.

The hazard characterization (dose-response) step describes the nature of the adverse health effects associated with the identified pathogen and the likelihood that they may occur; these depend on the immune status of the host, characteristics of the food that may modulate the immune status, virulence characteristics of the pathogen, and the number of cells ingested (Buchanan et al., 2000; Lammerding, 1997; Lammerding & Paoli, 1997; Schroeder et al., 2007). At the end of the dose-response step, a mathematical equation is established to relate the number of pathogen cells that are ingested (dose) to the occurrence of an illness (response). Some commonly used models are the exponential, Weibull-Gamma, beta-binomial, simple threshold, negative binomial, variable threshold, and beta-Poisson models (Haas et al., 2014). The beta-Poisson model, for instance, is based on a Poisson distribution of the number of pathogens in each dose. It assumes each pathogen has the potential to cause an infection and has a non-constant survival and infection probability, thus the use of a beta distribution to describe the survival probability (Haas et al., 2014).

The exposure assessment step relies on data on the prevalence of pathogens or their initial concentrations in food, processing conditions that may result in pathogen growth, inactivation, or cross-contamination, and the amount of food consumed to estimate the number of pathogen cells that will be ingested (Lammerding & Fazil, 2000; Pérez-Rodríguez et al. 2008; 2011). Finally, in the risk characterization step, data from the three preceding steps is combined to give an overall risk estimate i.e., the likelihood of adverse health effects in a population due to the pathogen that was ingested (Buchanan et al., 2000; Lammerding, 1997; Lammerding & Paoli, 1997; Schroeder et al., 2007). The risk characterization step also allows for sensitivity analysis to determine the impact of various variables on the risk estimate.

The current QMRA approach has been used in various studies to analyze food product pathways, estimate risks, design intervention strategies, and inform food safety management decisions (Collineau et al., 2020a, 2020b; Pang et al., 2017; Pradhan et al., 2019; Whiting & Buchanan, 2008). Despite their usefulness, QMRA only focuses on foodborne pathogens at the species level using changes in their numbers and does not directly incorporate any changes that may occur at the gene level. Ongoing studies will usher us into the future of QMRA, which has been termed, 'next-gen QMRA' (Cocolin et al., 2018; Rantsiou et al., 2018). Next-gen QMRA focuses on the integration of genomic, transcriptomic, proteomic, and metabolomic data into QMRA models to capture gene-level differences between closely related pathogens (Cocolin et al., 2018; Fritsch et al., 2018; Godínez-Oviedo et al., 2022; Njage et al., 2019, 2020). This omics data can be used in the hazard identification and characterization steps to refine AMR and virulence profiles, adjust dose-response models, and predict clinical outcomes (Cocolin et al., 2018; Collineau et al., 2019). In the exposure assessment phase, it can be used in adjusting the growth and survival capacity of pathogens based on prevailing conditions in the farm-to-fork chain (den Besten et al., 2018; Rantsiou et al., 2018; Collineau et al., 2019).

1.7 Project overview

The prevalence of *Salmonella* in chicken, coupled with the high consumption of chicken, has made this food-pathogen pair an important food safety concern for regulators, poultry processing industries, as well as consumers. Over 18% of the *Salmonella* illnesses that occur annually have been attributed to chicken (IFSAC, 2023). Despite a significant reduction in the prevalence of *Salmonella* in chicken, the number of illnesses has not reduced at the same rate (USDA FSIS, 2024b). This could be because some of the *Salmonella* serovars present in chicken

are not clinically important i.e., do not cause human diseases (USDA FSIS, 2024b). Also, genetic changes in *Salmonella* may have made it more resistant to antimicrobials and food processing controls. Understanding the ecological and evolutionary adaptations that fuel the persistence of foodborne pathogens like *Salmonella* heavily relies on genomic data. The utilization of WGS has yielded remarkable insights into *Salmonella*'s epidemiology, biology, evolution, and population dynamics (Allard et al., 2018). With the increasing adoption of WGS technology, machine learning presents a viable method to analyze WGS data. By harnessing the advantages of WGS data and machine learning's pattern recognition and predictive capabilities, we can develop or fine-tune models for risk assessment, food source attribution, prediction of pathogen stress response, *in silico* prediction of novel genes as well as targeted intervention strategies. The objectives for this research were:

1. To develop a machine learning model to identify *Salmonella enterica* stress response genes in isolates from poultry processing isolates.

2. To develop a genome-based machine learning model for predicting antimicrobial resistance phenotypes in *Salmonella enterica* isolated from chicken.

3. To evaluate the impact of population heterogeneity on risk estimates for *Salmonella enterica* in chicken.

Chapter 2: A machine learning approach to identifying *Salmonella* stress response genes in isolates from poultry processing

2.1 Abstract

This study explored the potential of machine learning to identify significant genes associated with *Salmonella* stress response during poultry processing using WGS data. The *Salmonella* isolates (n = 177) used in this study were obtained from various chicken sources (skin before chiller, chicken carcass before chiller, frozen chicken, and post-chill chicken carcass). Six machine learning algorithms (random forest, neural network, cost-sensitive learning, logit boost, and support vector machine linear and radial kernels) were trained on *Salmonella* WGS data, and model fit was assessed using standard evaluation metrics such as the AUROC curve and confusion matrix statistics. All models achieved high performances based on the AUROC, with logit boost showing the best performance with an AUROC curve of 0.904, sensitivity of 0.889, and specificity of 0.920. The significant genes identified included *ybtX*, which encodes a Yersiniabactin-associated zinc transporter, and the transferase-encoding genes *yccK* and *thiS*. Additionally, genes coding for cold (*cspA*, *cspD*, and *cspE*) and heat shock (*rpoH* and *rpoE*) responses were identified. Other significant genes included those involved in lipopolysaccharide biosynthesis (*irp1*, *waaD, rfc,* and *rfbX*), DNA repair and replication (*traI*), biofilm formation (*ccdA* and *fyuA*), and cellular metabolism (*irtA*).

2.2 Introduction

In the U.S., poultry processing facilities are always striving to produce safe products by adhering to regulatory requirements aimed at limiting the spread of pathogens during processing (Dogan et al., 2022; González et al., 2019; NCC, 2022). Despite these efforts, *Salmonella* is almost always found in all stages of poultry processing regardless of food safety measures that are in place (Nagel et al., 2013; NCC, 2022). *Salmonella*'s genetic diversity and exceptional capacity for adaptation are the two main factors that contribute to its success as a foodborne pathogen. *Salmonella* can endure adverse conditions through a complex and fine-tuned regulatory network that allows it to rapidly adjust its metabolism, transitioning from a growth to a stress-resistance phase in a short time, or even expressing stress resistance and virulence factors at the same time, if necessary (Guillén et al., 2021).

During poultry processing, *Salmonella* is exposed to stressors including low and high temperatures, disinfectants (oxidizing agents), and low nutrients at various stages of the cycle. Scalding of slaughtered chicken, for example, not only loosens feathers for easier removal but also inactivates and reduces bacterial numbers on the chicken carcass. *Salmonella*e can survive scalding through an (extra)cytoplasmic heat shock response characterized by the activation of several regulons that code for heat shock proteins that repair heat-induced misfolded proteins (Guisbert et al., 2008). The mechanisms by which *Salmonella* cells acclimate to cold have recently been reviewed by Ricke et al. (2018) and include modifications to the DNA supercoiling process, changes in envelope composition, and inducing the synthesis of several proteins. It is well known that the development of stress tolerance and resistance often comes at a fitness cost for bacterial cells and that the induction of stress responses to particular substances or conditions could promote

cross-protection for others or cause variations in *Salmonella* pathogenicity (Dawoud et al., 2017; Guillén et al., 2021). For instance, the NusA protein, which is synthesized under cold temperature conditions, has been found to influence the expression of the transcriptional activator of *Salmonella* pathogenicity island-1, *hilA* (Immerseel et al., 2008). Furthermore, virulence-related genes, *spvR, hilA, gyrA*, and *sefA* were found to be more highly expressed in *S.* Enteritidis cells cultured at elevated temperatures (Yang et al., 2014).

In this study, we aimed to use machine learning to identify genetic markers associated with stress response in *Salmonella* isolates collected from chicken sources during processing. It is expected that findings from this study could guide the development of improved preventive controls during poultry processing and WGS-based quantitative QMRA.

2.3 Materials and Methods

2.3.1 Isolates collection

The *Salmonella* isolates used in this study were sampled from previously sequenced isolates publicly available in the National Center for Biotechnology Information (NCBI) Pathogen Detection database (https://www.ncbi.nlm.nih.gov/pathogens/). One challenge associated with the use of data from this database is that sometimes there are inconsistencies in tags (labels) given to the isolation sources of pathogens. Since data is submitted to NCBI from all over the world, there is no harmonized naming convention for the metadata associated with genomic sequences. To reduce the limitations that would be associated with this study, the major inclusion criterion for isolates was that their isolation sources provided an unambiguous knowledge of their stage in a typical chicken processing cycle. For example, *Salmonella* isolates tagged with isolation sources

such as ‘chicken meat, ‘chicken carcass’, ‘chicken’, whole chicken, and ‘raw chicken’ were excluded from our dataset since these tags could describe chicken at any stage in the processing cycle. Another criterion was the availability of genome assemblies. Based on these criteria, a total of 177 *Salmonella* isolates obtained from chicken at various stages of processing were selected. The isolation sources of the selected isolates included frozen chicken, post-chill chicken carcass, skin before chiller, and chicken carcass before chiller. All available isolates in the Pathogen Detection database classified under the above-listed sources were selected.

2.3.2 Data labeling

The typical chicken processing cycle was divided into two parts, A and B, for a binary classification problem, based on the likely stress conditions present across the cycle. In Part A of the cycle, *Salmonella* isolates are predominantly exposed to heat, oxidative, and acid stresses. In Part B of the cycle, the predominant stress conditions are cold, desiccation, and starvation. *Salmonella* labeled as belonging to Part A included isolates from ‘skin before chiller’ (60) and chicken carcass before chiller (10). Those in Part B were frozen chicken (97) and post-chill chicken carcasses (10) **(Figure 2.1)**.

2.3.3 Bioinformatics analysis

The genome assemblies of all 177 isolates were readily available and downloaded from the NCBI Pathogen Detection database. The individual assemblies were then submitted to the Evolutionary Genealogy of Genes: Non-supervised Orthologous Groups (eggNOG)-mapper v2 webserver (http://eggnog-mapper.embl.de/) for annotation. eggNOG-mapper v2 is a web tool used

for the functional annotation of huge sets of sequences. It is based on fast orthology assignments using precomputed clusters and phylogenies from an associated database (eggNOG v5.0) (Cantalapiedra et al., 2021; Huerta-Cepas et al., 2019). The use of orthology predictions for functional annotations allows for higher precision than traditional homology-based approaches like basic local alignment search tool (BLAST) searches, as it minimizes the transfer of annotations from close paralogs (duplicate genes that have a higher chance of being involved in functional divergence) (Cantalapiedra et al., 2021; Huerta-Cepas et al., 2017). The genome annotation was performed using the Prodigal gene prediction method with minimum hit e-value and minimum hit bit score set at 0.001 and 60, respectively.

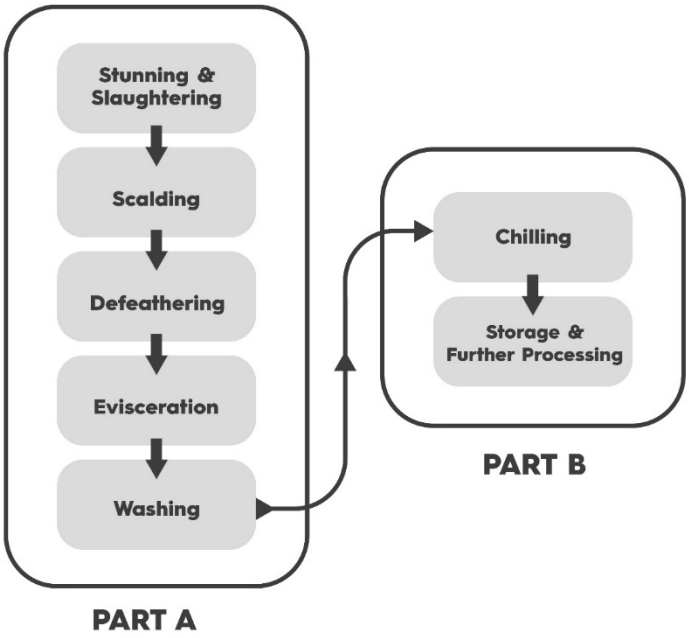


Figure 2.1. Typical stages of chicken processing. Part A of the cycle indicates the stages where *Salmonella* encounters pre-chill stresses including heat, oxidative, and acid stresses and Part B indicates the stages where *Salmonella* encounters cold, desiccation, and starvation stresses.

2.3.4 Machine learning

A supervised machine learning approach was utilized in this study. Six machine learning algorithms, random forest (RF), cost-sensitive learning (CSL), support vector machine linear (SVML) and radial kernels (SVMR), neural network (NNET), and logit boost (LB), were chosen to train our dataset. These algorithms were selected because they are robust, work well with high-dimensional data, and have been used by previous studies (Karanth et al., 2022; Tanui et al., 2022a; 2022b; Njage et al. 2019) focused on WGS data.

Candidate models were developed based on the hypothesis that machine learning could be used to accurately classify *Salmonella* isolates based on stress-induced variations in WGS and identify the significant genes associated with *Salmonella* stress response. All analyses were performed using the *caret* package (Kuhn, 2008; 2019) in R (v. 4.3.0, R Core Team, 2023, Vienna, Austria).

2.3.4.1. Data exploration

Our dataset contained 177 isolates with 3894 unique genes spread across these isolates. The nearZeroVar function in the *caret* package in R was applied to identify features (genes) that had a single unique value (zero variance predictor). Features that only had a few unique values (near-zero variance predictor) were also identified since they could become zero-variance predictors during data split into cross-validation subsamples (Kuhn, 2019; Kuhn & Johnson, 2013). These features were identified by calculating two metrics, the frequency ratio and the unique values percentage. The frequency ratio is the ratio of the frequency of the most common value to the frequency of the second most common value. Typically, ideal frequency ratios are near one for

good features but could be larger for highly unbalanced datasets. The unique values percentage is the ratio of the number of unique values per feature to the total number of values in the dataset expressed as a percentage (Kuhn, 2019; Kuhn & Johnson, 2013). This metric approaches zero as the data granularity increases. In our analysis, a frequency ratio greater than 19 and unique values percentage less than 10% was employed to avoid false detection of predictors as near-zero variance as described by Kuhn (2019). These near-zero variance predictors (genes) were removed, as keeping them would increase model complexity and computational time while providing little to no benefit for model performance. Near-zero variance predictors tend to be non-informative and could result in an unstable fit or cause the model to crash, especially for algorithms that require more than one unique value per predictor.

Further exploration of this dataset showed a mild class imbalance (when the minority class is 20 – 40% of the dataset) (GoogleML, 2022). Our dataset had 70 (39.55%) *Salmonella* isolates in the minority class (isolates labeled as Part A) and 107 (60.45%) *Salmonella* isolates in the majority class (Part B). Most machine learning algorithms are designed for a close to equal number of observations in each class (Ling & Sheng, 2011). Thus, when there is a class imbalance, the algorithms tend to be biased toward the majority class. This results in models with low balanced accuracy, although overall accuracy could be high, and poor predictive performance, especially for the minority class. The random over-sampling examples (ROSE) technique was applied to address the mild imbalance in our dataset. ROSE downsamples the majority class and uses a smoothed-bootstrap approach to synthesize new data points from a conditional kernel density estimate of the minority class (Lunardon et al., 2014). The resulting data set is more balanced. ROSE does not result in information loss and has been shown to outperform traditional approaches such as upsampling and downsampling (Menardi & Torelli, 2014).

2.3.4.2 Model building and evaluation

Six machine learning algorithms, CSL, RF, SVML, SVMR, NNET, and LB, were used in training the data. The model sequence began by randomly partitioning the dataset into a training (70%) and testing (30%) set. A train control parameter was applied to each model to allow for resampling, hyperparameter tuning, and parallel processing. Resampling of the training dataset was achieved via 10-fold cross-validation. This approach generated 10 random subsets (folds) of the dataset; nine folds were used in model fitting while the remaining fold (a hold-out) was used to evaluate model performance (Kuhn, 2019). This process was repeated until all 10 folds of data had gone through cycles of hold-out and model fitting. For all six models, a random hyperparameter search was applied to select the optimal tuning parameter combinations (Bergstra & Bengio, 2012). The developed models were then evaluated against the testing set and a confusion matrix was constructed. Model performances were evaluated based on the AUROC curve, sensitivity, specificity, and Kappa value.

2.4 Results and Discussion

2.4.1 *Salmonella* strains

The dataset used in this study was *Salmonella* WGS obtained from chicken sources (frozen chicken, post-chill chicken carcass, skin before chiller, and chicken carcass before chiller). The dataset showed high serovar diversity, and comprised of serovars Infantis, Enteritidis, Typhimurium, Kentucky, Schwazengrund, -:r:1,5, Agona, and Indiana, among others, and had been isolated in Brazil, China (Hanzhong), Ecuador, and the U.S. (Maryland, Minnesota, and Pennsylvania) between 2015 and 2022.

2.4.2 Machine learning modeling

In our study, supervised machine learning was employed to identify genetic markers associated with stress response in *Salmonella* isolates collected from chicken sources during processing. During data exploration, a feature reduction function was used to remove 3454 (out of 3894) genes due to zero and near-zero variance, leaving 440 genes that were used in modeling. Furthermore, ROSE was applied to address the mild class imbalance (60.45%/39.55%) in our dataset. The resulting ROSEd data was perfectly balanced at 51%/49%. The two sets of models developed, using both ROSEd (perfectly balanced) and non-ROSEd (mild imbalance) datasets, achieved similar performance scores indicating that the mild imbalance in the non-ROSEd dataset had no impact on model performance. Thus, only results from the models based on the non-ROSEd dataset will be presented here.

All models showed high performance, with AUROC values ranging between 0.855 and 0.904 (**Figures 2.2 & 2.3**). A receiver operating characteristic (ROC) curve is used to evaluate a model's classification or discrimination ability. It graphs a model's true positive rate (sensitivity) against its false positive rate (1- specificity). An ideal AUROC curve has a value of 1.0. In **Figures 2.2 & 2.3**, the diagonal line represents an AUROC of 0.5, which is indicative of a model whose classification ability is no better than a random classifier or guess.

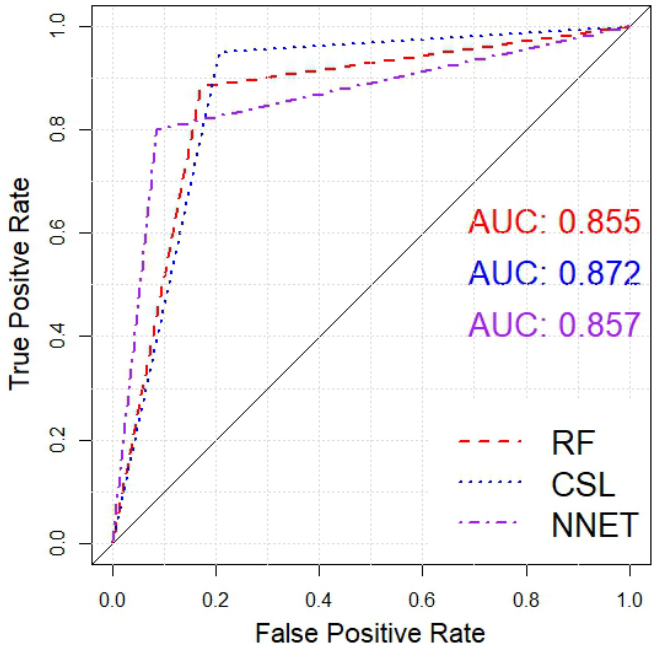


Figure 2.2. Receiver operating characteristic (ROC) curves for random forest (RF), cost-sensitive learning (CSL), and neural network (NNET) models.

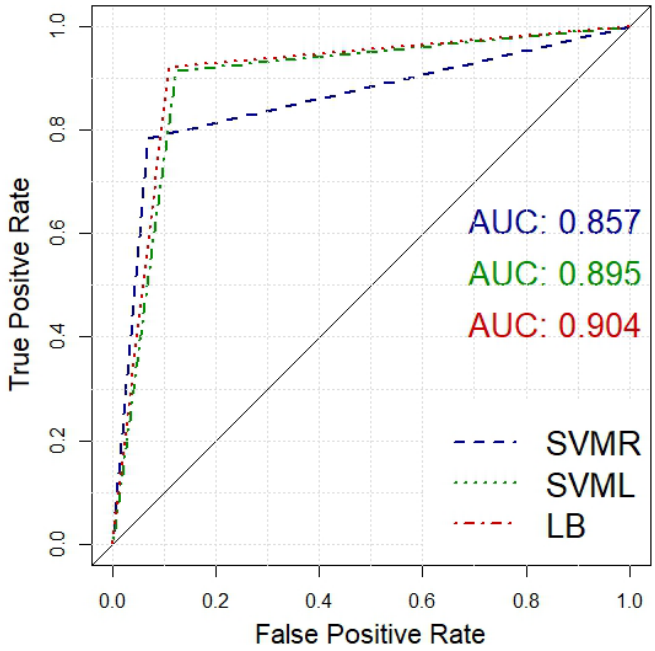


Figure 2.3. Receiver operating characteristic (ROC) curves for support vector machine radial kernel (SVMR), support vector machine linear kernel (SVML), and logit boost (LB) models.

The closer a model's ROC is to the upper left corner of the graph (**Figures 2.2 & 2.3**), the higher its accuracy or discriminative ability. Generally, AUROC values above 0.8 are often regarded as acceptable and indicative of high-performing models (Nahm, 2022). Specifically, the AUROC values for RF, CSL, SVMR, SVML, NNET, and LB were 0.855, 0.872, 0.857, 0.895, 0.857, and 0.904, respectively.

Table 2.1. Model performance and optimum tuning parameters for the six machine learning algorithms used in this study.

Model	Optimum hyperparameters	Sensitivity	Specificity	Kappa
RF	mtry = 288	0.828	0.883	0.712
CSL	trials = 20 model = tree winnow = false cost = 2	0.793	0.950	0.745
NNET	size = 1 decay = 0.1	0.914	0.800	0.712
SVMR	sigma = 0.001662214 C = 6.904842	0.931	0.783	0.713
SVML	C = 1	0.931	0.733	0.662
LB	nIter = 96	0.889	0.920	0.810

RF: random forest; CSL: cost-sensitive learning; NNET: neural network; SVMR: support vector machine radial kernel; SVML: support vector machine linear kernel; LB: logit boost.

Similarly, the models recorded a high sensitivity and specificity above 0.79 and 0.73, respectively **(Table 2.1)**. Karanth et al. (2022), Tanui et al. (2022a; 2022b), and Njage et al. (2019) have used similar machine learning approaches and algorithms to varying degrees of model performance. Our results here fall into the ranges previously reported. We also evaluated the models using Cohen's κ statistic (Cohen, 1960). Cohen's κ measures the level of agreement between observed and predicted classes and is an indicator of how much better one classifier performs than another random classifier that simply guesses based on the frequency of each class. We recorded κ values above 0.6 for all models (**Table 2.1**), indicating substantial (0.61 – 0.80)

and almost perfect (0.81 – 1.0) agreement between the observed and predicted classes based on Landis and Koch's (1977) interpretation. The LB model (AUROC = 0.904; sensitivity = 0.889; specificity = 0.920; kappa = 0.81), which achieved maximum performance after 96 boosting iterations (nIter), was selected as the best-performing model.

2.4.3 Important *Salmonella* genes associated with stress response

The LB model was then used to determine the most important genes associated with stress response in the *Salmonella* isolates used in this study. The 20 most important genes have been presented in **Table 2.2**. Protein names and functional information were obtained from the UniProt knowledgebase (UniProt Consortium, 2017; https://www.uniprot.org/). The survival of higher organisms continually depends on their ability to sense and respond to changes in their environment. Similarly, microorganisms have mechanisms for responding to unfavorable changes in their environment. To ultimately cause diseases, *Salmonella* must traverse adverse conditions in the environment, food, and the gastrointestinal tract. *Salmonella* achieves this through coordinated regulation of gene expression in response to various extracellular stress conditions. In a poultry processing cycle, *Salmonella* is exposed to heat, cold, acid, oxidative, desiccation, and starvation stresses. Response mechanisms to these stresses involve the induction of regulatory genes and transcription of their associated proteins and gene products (Ye et al., 2019). Generally, the mechanisms of these stress responses involve the induction of regulatory genes and the transcription of their associated proteins and gene products (Andino & Hanning, 2015). Several genes associated with *Salmonella* heat and cold stress response, transcriptional regulation, transport mechanisms, and cellular metabolism were identified as significant genes in the poultry processing cycle (**Table 2.2**).

Table 2.2. The twenty most important genes associated with stress responses in *Salmonella* isolates from poultry processing.

Gene	Protein name	Functions
ybtX	Yersiniabactin-associated zinc MFS transporter YbtX	Transmembrane transporter activity
amiD_2	N-acetylmuramoyl-L-alanine amidase	Hydrolase; Hydrolyzes cell wall peptidoglycans
irp1	Yersiniabactin polyketide synthase HMWP1	Lipid biosynthesis; Organic cyclic compound biosynthetic process
rpoH	RNA polymerase sigma factor RpoH	DNA-templated transcription initiation; Response to heat shock
pchR	AraC family transcriptional regulator	Sequence-specific DNA binding
rpoE	RNA polymerase sigma-E factor	DNA-templated transcription initiation; Response to temperature stimulus; Cellular stress response to acid
fyuA	Siderophore yersiniabactin receptor FyuA	Siderophore-iron transmembrane transporter activity; Signaling receptor activity
irtA	ABC transporter ATP-binding protein	ATP hydrolysis activity
cspA	Cold shock protein CspA	Cold stress response; Transcription regulation
rffH	Glucose-1-phosphate thymidylyltransferase	Extracellular polysaccharide biosynthetic process; Catalyzes the formation of dTDP-glucose from dTTP and glucose 1-phosphate, as well as its pyrophosphorolysis.
ccdA	Antitoxin CcdA	Acts as an antitoxin in the CcdA/CcdB type II toxin-antitoxin system
cspD	Cold shock-like protein CspD	Regulation of DNA-templated transcription
traI	Conjugative transfer relaxase/helicase TraI	DNA helicase activity
waaD	ADP-L-glycero-D-manno-heptose-6-epimerase	Lipopolysaccharide core region biosynthetic process
Z012_07430	Hypothetical protein	N/A
yccK	Sulfurtransferase	Transferase activity: part of a sulfur-relay system required for 2-thiolation of 5-methylaminomethyl-2-thiouridine at tRNA wobble positions.
cspE	Cold shock protein CspE	DNA binding; Negative regulator of cspA transcription
thiS	Sulfur carrier protein ThiS	Transferase activity
rfc	O-antigen polymerase	Lipopolysaccharide biosynthetic process
rfbX	Putative o-antigen transporter	O-antigen biosynthetic process

N/A – not available

Salmonella detects changes in temperature through a complex system of thermosensors including proteins, lipids, and thermoresponsive RNAs. Thermosensors are located in the *Salmonella* 5′ untranslated region and can regulate gene expression to produce temperature stress responses (Dawoud et al., 2017). Exposure to cold temperatures, like those present during the chilling stage (< 5°C), results in a stoppage of *Salmonella* growth and initiation of adaptation for survival. The *cspA, cspD,* and *cspE* genes encode cold shock proteins (CSPs) in response to cold stress. The *Salmonella* cold stress response mechanism involves various stages of acclimatization, during which CSPs are synthesized, cellular recovery, and a permanent adaptation to the reduced temperature (Ricke et al., 2018). During cold shock, transcriptional and translational processes limit the function of the ribosome, leading to inadequate protein synthesis and folding, and finally a reduction of growth rate (Chattopadhyay, 2006). *Salmonella* then adjusts the transcription and translation functions toward the production of CSPs and DNA supercoiling (Barria et al., 2013). These CSPs are essential for DNA repair and translation that allow *Salmonella* to grow at cold temperatures. *Salmonella* also adapts to low temperatures by modifying fatty acid chain lengths and the unsaturated/saturated fatty acid ratio of its cell membrane to decrease membrane fluidity and protect the cell (Yoon et al., 2015). The presence of the *irp1* gene**,** which functions in lipid biosynthesis could be indicative of fatty acid synthesis in response to cold stress.

Salmonella is exposed to higher than optimal temperatures (>50°C) during the scalding of chicken (Dogan et al., 2022). Exposure to high temperatures results in the unfolding and misfolding of proteins in the *Salmonella* cytoplasmic membrane and the outer membrane (Yura & Nakahigashi, 1999). This initiates a two-fold response: a cytoplasmic heat shock response controlled by the heat shock sigma factor H and an extracytoplasmic heat shock response controlled by sigma factor E (Spector & Kenyon, 2012). Sigma factors H and E, both part of the

RNA polymerase (RNAP) holoenzyme complex, are encoded by the *rpoH* and *rpoE* genes. Both *rpoH* and *rpoE* oversee a chaperone-mediated binding to misfolded and unfolded proteins leading to an increase in the production of small heat shock proteins that act as proteases (Guisbert et al., 2008). Additionally, there is the transcription of DNA repair enzymes. These processes aid in reducing the effects of increased temperatures on the cytoplasmic and extracytoplasmic compartments and enable *Salmonella* survival at high temperatures (Ades, 2008).

Liu et al. (2021) showed that *Salmonella* can survive on chicken carcasses by forming biofilms in response to harsh environmental conditions. Biofilms are more resistant to disinfectants and temperature stresses, which are used as pathogen control measures during poultry processing and are a huge concern for food safety (Liu et al., 2021). The *ccdA* gene codes for antitoxin CcdA in the CcdA/CcdB type II toxin-antitoxin system. CcdA acts as an antidote to CcdB under normal physiological conditions. However, under stress conditions, the toxin-antitoxin system plays an important role in *Salmonella* survival through involvement in post-segregational killing, persister cell formation, biofilm formation, and antibiotic persistence development (Singh et al., 2021). Another gene that was present, *fyuA*, encodes for the siderophore yersiniabactin receptor FyuA which is essential for biofilm formation in iron-poor environments (Hancock et al., 2008). The genes *ccdA* and *fyuA* may enhance the survival and persistence of *Salmonella* on chicken carcasses through biofilm formation.

Peracetic acid, acidified sodium chlorite, trisodium phosphate, and chlorine dioxide are some common disinfectants that are used in reducing the microbial load on chicken carcasses (Dogan et al., 2022). These disinfectants act by inducing oxidative stress in bacterial cells, but *Salmonella* possesses stress response mechanisms that allow it to survive the attack of oxidants.

We found genes, *yccK* and *thiS*, which encode enzymes involved in sulfur transfer as part of the important stress-related genes in poultry processing. The gene *yccK*, in particular, is part of a sulfur-relay system that is required for 2-thiolation of 5-methylaminomethyl-2-thiouridine (mnm_5s_2U) at tRNA wobble positions. This modification of mnm_5s_2U is important in *Salmonella'*s response to environmental stresses including oxidative stress (Z. Wang et al., 2022).

Other important genes identified in our analysis were those involved in transmembrane transport (*ybtX* and *irtA*), DNA replication and repair (*traI*), and lipopolysaccharide biosynthesis (*rfbX, rfc,* and *waaD*). The *rfc* gene encodes the O-antigen polymerase, which is required for the addition of newly synthesized O-antigen repeating units to the lipopolysaccharide core region (Kong et al., 2011). Lipopolysaccharides are a major structural component of the *Salmonella* outer membrane, contributing to the structural integrity of the bacterial cell wall and providing resistance to chemical stresses (Bertani & Ruiz, 2018; Jefferies et al., 2019). Thus, the presence of *traI, rfbX, rfc,* and *waaD* suggests *Salmonella* cells undergo DNA and outer membrane repair in response to stress conditions. Additionally, lipopolysaccharides are a major *Salmonella* virulence factor and play an important role in bacteria-host interactions by modulating host immune response.

The prevalence of *Salmonella* in chicken, coupled with the high consumption of chicken, has made *Salmonella* in chicken an important food safety concern for regulators, poultry processing industries, as well as consumers. Over 18% of the *Salmonella* illnesses that occur annually have been attributed to chicken (IFSAC, 2023). During poultry processing, stresses such as extreme temperatures and disinfectants are used to decontaminate chicken carcasses. However, these stresses do not only induce adaptive responses but can also induce the expression of virulence factors (Fang et al., 2016; Guillén et al., 2021). Thus, typical environmental stresses can have

indirect clinical importance. Etter et al. (2019) found that outbreak strains of *Salmonella* Heidelberg had enhanced heat tolerance and biofilm-forming ability and that exposure to heat stress increased the expression of multidrug efflux and virulence genes in *Salmonella* Heidelberg. The ability of one stress response to induce cross-protection against other stresses as well as virulence highlights the need to consider these important genes (**Table 2.2**), which represent a stress response profile, in designing poultry process operations, preventive controls, and risk models to minimize the risk of foodborne illness.

The use of WGS techniques to reveal the complete genomic sequence of microorganisms is increasingly becoming popular in food safety monitoring and outbreak investigations (Allard et al., 2016; E. Brown et al., 2021; Cook & Nightingale, 2018; Stevens et al., 2022). No two *Salmonella* serovars are the same; large differences exist in their virulence, epidemiology, antibiotic resistance, stress response mechanisms, growth kinetics, and other physiological functions (Lianou & Koutsoumanis, 2011; Pornsukarom et al., 2018). WGS offers opportunities to explain the intra-species variability in *Salmonella* virulence, adaptation to different stresses, and the genetic determinants of these adaptations. The successful connection of WGS data to specific phenotypes will aid in the prediction of pathogen behavior and the development of risk-based control measures (Fritsch et al., 2018; Guillier et al., 2022). Our current study provides an approach to using WGS and machine learning to identify stress genes. By leveraging genomic data and applying advanced data analytic techniques, more effective processing interventions and risk assessment models can be developed to reduce *Salmonella* infections attributable to chicken.

2.5 Conclusion

WGS has numerous advantages, from pathogen identification to providing insights into the evolutionary and ecological adaptations that underlie the survival of foodborne pathogens in the environment and their hosts. The ongoing developments in WGS will likely result in the widespread application of WGS data to address various food safety and foodborne illness issues. Due to the high-dimensional nature of WGS data, they are better analyzed through advanced data analytic techniques such as machine and deep learning. Machine learning is very useful in analyzing any data type to identify appropriate patterns and make predictions. Nonetheless, when applied to food safety problems, it can be limited by the absence of WGS metadata (e.g., serotype, isolation date and time, isolation source). Combining appropriate metadata with WGS data will greatly improve the predictability and usefulness of WGS-based machine learning models.

In this study, we trained machine learning algorithms to identify significant genes involved in *Salmonella* physiology during poultry processing. We identified genes associated with *Salmonella* heat and cold stress response, transcriptional regulation, transport mechanisms, and cellular metabolism as being important to the *Salmonella* stress response process. In our future work, these genes will form the basis for improved *Salmonella* control measures and risk models.

Chapter 3: Genome-based machine learning for predicting antimicrobial resistance in *Salmonella* isolated from chicken

3.1 Abstract

Antimicrobial resistance (AMR) continues to pose a grave threat to public health. The increase in the burden of AMR is fueled by the indiscriminate use of antimicrobial agents in agriculture. The objective of this study was to develop a genome-based machine learning model to predict AMR in *Salmonella* isolated from chicken meat. Genomic information on 205 *Salmonella* isolates from chicken meat was combined with data on the AMR phenotype of these isolates to cefoxitin, tetracycline, streptomycin, sulfisoxazole, ceftriaxone, ceftiofur, ampicillin, and amoxicillin-clavulanic acid. Four machine learning algorithms i.e., logit boost, random forest, support vector machine, and extreme gradient boosting were trained on this data to build models. The best-performing model for each antimicrobial was used to predict the AMR phenotypes of a new set of 200 *Salmonella* isolates also from chicken meat, and the predictions were compared to AMR phenotype predictions from ResFinder. The machine learning models showed high sensitivity (≥0.833), specificity (≥0.833), and balanced accuracy (≥0.866), across all the antimicrobials. The models predicted resistance prevalences ranging from 1% (ceftriaxone) to 65.5% (streptomycin). When the AMR phenotype predictions of the machine learning models were compared to predictions from ResFinder, the predictions from this study were accurate (>95%).

3.2 Introduction

Antimicrobial resistance is a global health crisis characterized by microorganisms evolving to withstand the effects of antimicrobial drugs, thus making the drugs ineffective. It is a multifaceted challenge and has far-reaching implications for food production and safety, healthcare, public health, and national economies, thus understanding its scope and consequences is essential. The consequences of AMR range from prolonged hospital stays, increased healthcare costs, and higher morbidity and mortality rates (WHO, 2023). It has been estimated that AMR effects could result in a 1.1% – 3.8% decrease in the global gross domestic product by 2050 (WBG, 2017). Another report estimated that the increase in drug-resistant infections could put 10 million human lives at risk annually by 2050, causing a cumulative loss of US$100 trillion in economic output (O'Neill, 2016). This indicates an even more dire impact of AMR on public health if no solutions are found on time.

Antimicrobial agents can kill or slow the growth of susceptible microorganisms, allowing resistant ones to multiply. AMR is essentially a natural evolutionary process. In fact, resistance to penicillin, one of the world's first discovered antibiotics, was observed even before the first administration of the drug (Thakur & Gray, 2019). Microorganisms naturally mutate, and those with genetic changes that confer resistance to antimicrobial agents survive and multiply while the susceptible ones are killed. Though AMR occurs naturally, factors such as international trade, increased globalization, increased demand for animal protein, and the slow pace of development of new antimicrobials and antimicrobial alternatives (e.g. vaccines, peptides, and bacteriophages) to replace older, less effective drugs, have been recognized as drivers of the emergence and spread of AMR microorganisms (de Mesquita Souza Saraiva et al., 2022; Thakur & Gray, 2019).

Additionally, the overuse and misuse of antimicrobials in agriculture is a major driver of AMR in *Salmonella enterica* (Agyare et al., 2019).

Antimicrobial agents play a crucial role in modern agricultural systems like poultry production, where they are used to treat and control pathogens such as *Clostridium perfringens*, *Escherichia coli*, and *Salmonella enterica* (de Mesquita Souza Saraiva et al., 2022; Yin et al., 2021). Recent reports reveal a decrease in antimicrobial use in animal production. Singer et al. (2023) reported a substantial decrease in the use of medically important in-feed antimicrobials in U.S. broiler chicken production. There was a 96% decrease in tetracycline use between 2013 and 2017, and in 2020 and 2021, there was no recorded in-feed tetracycline use. For virginiamycin, there was a 97% decrease in usage between 2013 and 2021. A similar trend was observed for several antimicrobials used in broiler chicken production in Canada (Huber et al., 2021). Furthermore, the use of antimicrobial agents in animal production decreased by 13% worldwide between 2017 and 2019 (WOAH, 2023).

Salmonella possesses a large array of AMR mechanisms including reduced membrane permeability, modification of antimicrobial target site, removal of antimicrobial by transporters, enzymatic inactivation, and substitution of target proteins (Castro-Vargas et al., 2020; Rakitin et al., 2022; Tan et al., 2022). Like other cellular processes, these mechanisms are controlled by genes. Due to recent advances in sequencing technologies, whole-genome sequencing (WGS) has gradually become a part of routine epidemiologic surveillance (Allard et al., 2016; E. Brown et al., 2021; Stevens et al., 2022). Genomic surveillance of AMR offers a more precise characterization of pathogens, than traditional culture-based (phenotypic) antimicrobial susceptibility testing methods, as it can reveal the different genetic elements that act as AMR determinants (Djordjevic

et al., 2023). Moreover, studies have shown that the presence of an antimicrobial resistance gene in a genome sequence can be a good predictor of phenotypic resistance to antimicrobial agents (Bharat et al., 2022; McDermott et al., 2016; Sia et al., 2021).

Machine learning is increasingly becoming the preferred method for analyzing large datasets generated from sequencing platforms and consequent bioinformatic analyses (Benefo et al., 2022). Recent studies have combined genomic data with other relevant metadata and analyzed them with machine learning algorithms for predicting disease severity, stress response genes, food safety risk, and food source attribution (Benefo et al., 2024; Karanth et al., 2022; Njage et al., 2020; Tanui et al., 2022b). The availability of genome sequences and AMR gene databases presents an opportunity to apply machine learning tools in AMR prediction. To date, few studies have implemented machine learning workflows for predicting AMR in foodborne pathogens (Maguire et al., 2019; Ren et al., 2022). The use of machine learning in AMR studies would be helpful in the identification of novel AMR genes and resistance patterns.

The objective of this study was to develop a genome-based machine learning model to predict the AMR phenotypes of *Salmonella* isolated from chicken meat. We focused on eight antimicrobials namely amoxicillin-clavulanic acid, ampicillin, ceftiofur, ceftriaxone, sulfisoxazole, streptomycin, tetracycline, and cefoxitin.

3.3 Materials and Methods

3.3.1 Isolates collection

All the *Salmonella* (n = 405) isolates used in this study were taken from previously sequenced isolates publicly available in the National Center for Biotechnology Information (NCBI) Pathogen Detection database (https://www.ncbi.nlm.nih.gov/pathogens/). Two hundred and five (205) of these isolates were collected from chicken meat samples in the U.S. by the National Antimicrobial Resistance Monitoring System (NARMS) and had complete antimicrobial susceptibility phenotype information (McDermott et al., 2016). Of these 205 isolates, 81 were susceptible to all eight antimicrobials mentioned earlier while 124 were resistant to at least one of the antimicrobials. The remaining 200 isolates were also collected from chicken meat samples in the U.S. by NARMS and their antimicrobial susceptibility phenotypes were unknown. All 405 isolates were from NCBI BioProject PRJNA242614 and PRJNA292661.

3.3.2 Bioinformatics analysis and AMR gene identification

Raw reads for all isolates were obtained from the Sequence Read Archive in the NCBI database and *de novo* assembled on the Bacterial and Viral Bioinformatics Resource Center (BV-BRC; https://bvbrc.org) webserver (Olson et al., 2023). The default 'Auto' assembly method in BV-BRC was selected. This approach runs BayesHammer on short reads and Canu on long reads, followed by assembly strategies including Velvet, IDBA, SPAdes, and MiniASM (Bankevich et al., 2012; Koren et al., 2017; Li, 2016; Nikolenko et al., 2013; Peng et al., 2010; Zerbino & Birney, 2008). Each assembled genome was assigned an assembly score by ARAST or ALE and the quality

was evaluated using QUAST (Clark et al., 2013; Gurevich et al., 2013). *De novo* assembled sequences were then annotated using the Rapid Annotation using Subsystems Technology toolkit (RASTtk). Annotated AMR genes were then obtained from the Pathosystems Resource Integration Center (PATRIC) and Comprehensive Antibiotic Resistance Database (CARD) databases in BV-BRC (Alcock et al., 2023; Davis et al., 2020).

To ensure a comprehensive analysis of AMR genes, similar analyses were performed on the ResFinder, NCBI AMRFinderPlus, ARG-ANNOT, and MEGARes databases using genome assemblies and raw reads of the isolates. (Bortolaia et al., 2020; Feldgarden et al., 2021; Florensa et al., 2022). For ResFinder, the raw reads were mapped against the ResFinder 4.0 database using the resfinder analysis pipeline (Bortolaia et al., 2020), where we identified the different antimicrobial resistance genes and antimicrobial resistance genes with a prediction of the expected phenotypes for a selection of bacterial species. Species-specific predictions for selected antimicrobial agents which are part of the ResFinder were implemented to only present clinically relevant phenotypes for these species and also to include mutation-mediated phenotypes. In addition to adding the phenotype feature, the PointFinder database was also used (Florensa et al., 2022). The predicted phenotype results table for each sample was further grouped using R. Downstream data analysis was carried out in RStudio using the following R packages: dplyr (v1.1.3), reshape2 (1.4.4), ggh4x (0.2.8), and stringr (v1.5.1).

3.3.3 Machine learning

We used a supervised machine learning approach in this study as described by Benefo et al. (2024), Karanth et al. (2022), and Tanui et al. (2022b). Four machine learning algorithms,

random forest (RF), logit boost (LB), support vector machine radial kernel (SVMR), and extreme gradient boosting (XGB), were used to train our dataset. These algorithms were selected because of their robustness and utility in analyzing WGS data. The *Salmonella* isolates (n = 205) with known AMR phenotype information were used in building machine learning models that were used in predicting the AMR phenotype of the remaining 200 *Salmonella* isolates **(Figure 3.1**).

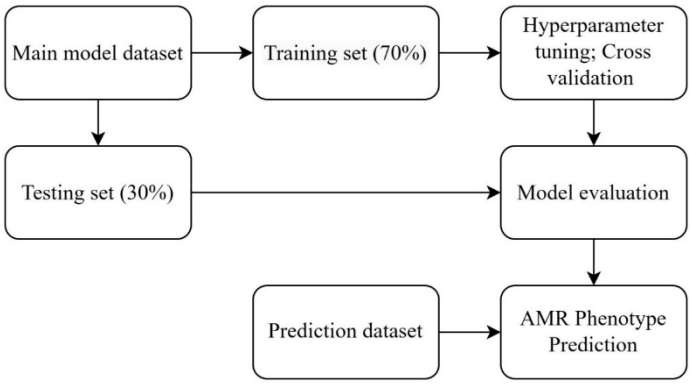


Figure 3.1. Machine learning workflow for AMR phenotype prediction.

3.3.3.1 Data labeling and exploration

Of the 205 *Salmonella* isolates with known AMR phenotypes (from broth microdilution tests), 81 were susceptible to all eight antimicrobials while 124 were resistant to at least one antimicrobial. Based on this, isolates were labeled as 'resistant' or 'susceptible', setting up a binary classification problem for the machine learning model. We have shown in **Table 3.1** the number of isolates used in building machine learning models for each of the antimicrobials.

Table 3.1. Number of resistant and susceptible isolates used for machine learning modeling.

Antimicrobial	Resistant	Susceptible
Amoxicillin-clavulanic acid	40	81
Ampicillin	46	81
Ceftiofur	40	81
Ceftriaxone	40	81
Sulfisoxazole	70	81
Streptomycin	57	81
Tetracycline	99	81
Cefoxitin	35	81

It can be seen from **Table 3.1** that all antimicrobials had an unequal number of resistant and susceptible isolates. This difference in isolate numbers resulted in a mild class imbalance (when the minority class is 20–40% of the dataset) for amoxicillin-clavulanic acid ampicillin, ceftiofur, ceftriaxone, and cefoxitin. Most machine learning algorithms are designed for a close to equal number of observations in each class. As such, when there is a class imbalance, the algorithms tend to be biased toward the majority class and could result in models with low accuracy and poor predictive performance. Therefore, we applied the random over-sampling examples (ROSE) technique to address the mild class imbalance. ROSE applies a smoothed bootstrap method to generate new observations from a conditional kernel estimate of the minority class (Lunardon et al., 2014).

Furthermore, we performed feature reduction by identifying and removing features (AMR genes) that had a single unique value (zero variance feature) or had a few unique values (near-zero variance feature). These features were excluded from the model because their inclusion would increase model complexity and computational time without substantially improving model performance. Additionally, near-zero variance features often lack informative value and could lead

to model crashes, particularly in algorithms that require multiple values for each feature (Kuhn & Johnson, 2013).

3.3.3.2 Model building and prediction

Machine learning models were developed for each antimicrobial using four algorithms, RF, CSL, SVMR, and XGB. For each antimicrobial, its dataset was randomly partitioned into a training (70%) and testing set (30%). This was followed by defining a train control parameter to allow for resampling, hyperparameter tuning, and parallel processing. Resampling of the training dataset was achieved via 10-fold cross-validation. This approach generated 10 roughly equal random folds (subsets) of the dataset; nine folds were used in model fitting, while the remaining fold (a hold-out) was used to evaluate model performance (Kuhn, 2019). This was an iterative process that continued until all 10 folds of data had gone through cycles of hold-out and model fitting. A random hyperparameter search was applied to identify the optimal tuning parameter combinations for all the models during training. The developed classification models were then evaluated against the testing set (30% of data not used in training) and a confusion matrix was constructed. Model performances were evaluated based on the balanced accuracy, sensitivity, specificity, and Kappa value. The best-performing model for each antimicrobial was then used to predict the AMR phenotypes of the 200 *Salmonella* isolates whose phenotype was initially unknown. Our machine learning predictions were then compared to WGS-based AMR predictions from the ResFinder database. All analyses were performed using the Caret package (Kuhn, 2008; 2019) in R (v. 4.3.2, R Core Team, 2023, Vienna, Austria).

3.4 Results and Discussion

3.4.1 Strain characteristics

Our dataset included *Salmonella* genomes obtained from retail chicken meat in the U.S. and included serovars such as Kentucky, Enteritidis, Heidelberg, Schwazengrund, Senftenberg, Typhimurium, Thompson, Hadar, Infantis, and Livingstone, among others. Across the 205 isolates used in model building, 40% (81) were resistant to none of the eight antimicrobials, 60% (124) were resistant to at least one antimicrobial, and 16% (33) resistant to at least seven antimicrobials. The most prevalent resistances were tetracycline (48%), sulfisoxazole (34%), and streptomycin (28%). Twenty-five percent (51) of the *Salmonella* isolates were multidrug-resistant (resistant to at least one antimicrobial in three or more antimicrobial classes).

Across all *Salmonella* isolates used in the study, 31 AMR genes were identified in the MEGARes database, 31 in the ARG-ANNOT database, and 32 in the ResFinder database. A further 54, 74, and 88 AMR genes were identified in the PATRIC, NCBI AMRFinder Plus, and CARD databases, respectively. AMR genes including *tetA*, *fosA*, *aadA7*, *aph(3'')-Ia,* and *blaCMY-4* were found in at least four databases. Since some of the genes were identified in multiple databases, our final harmonized dataset contained 285 unique genes.

3.4.2 Machine learning modeling

Machine learning algorithms are useful for identifying patterns in data, learning from these patterns, and making predictions based on the learnings. In this study, we trained supervised machine learning algorithms on *Salmonella* AMR gene data to identify the patterns of gene presence or absence across several isolates and how these genetic patterns are linked to the AMR

phenotype of isolates. The classification performance of each model during testing was evaluated based on four metrics: sensitivity, specificity, balanced accuracy, and Cohen's κ statistic. Algorithms with the best performance for each antimicrobial have been presented in **Table 3.2**. Overall, all the algorithms used were able to achieve a high performance during modeling. For all eight antimicrobials, sensitivity values greater than 0.833 were recorded indicating that during model testing, the algorithms were able to accurately predict resistance in more than 83% of the test dataset. Similarly, specificity values above 0.833 indicate the accuracy of the algorithms used in predicting susceptibility during the model testing phase.

Table 3.2. Machine learning model performance across eight antimicrobials studied.

Antimicrobial	Algorithm	Sensitivity	Specificity	Balanced accuracy	Cohen's κ
AMC	LB	0.917	0.870	0.893	0.757
AMP	LB	0.923	0.913	0.918	0.785
TIO	RF	0.833	0.958	0.896	0.734
CRO	XGB	0.846	0.913	0.880	0.761
FIS	SVMR	0.904	0.875	0.890	0.807
STR	RF	0.941	0.917	0.929	0.822
TET	SVMR	0.965	0.958	0.962	0.841
FOX	SVMR	0.900	0.833	0.866	0.711

AMC: amoxicillin-clavulanic acid; AMP: ampicillin; TIO: ceftiofur; CRO: ceftriaxone; FIS: sulfisoxazole; STR: streptomycin; TET: tetracycline; FOX: cefoxitin. LB: logit boost; RF: random forest; XGB: extreme gradient boosting; SVMR: support vector machine radial kernel.

The balanced accuracy metric calculates classification accuracy within both the true positive and true negative class and is typically preferred over the overall accuracy metric, especially when dealing with imbalanced data like our study. Ideal models would have a balanced accuracy of 1.0; however, this is not always possible in practice, and models with accuracy scores above 0.5 are deemed to have performed better than a random classifier or guessing. All the models had a balanced accuracy above 0.86 (**Table 3.2**) indicative of an acceptable classification

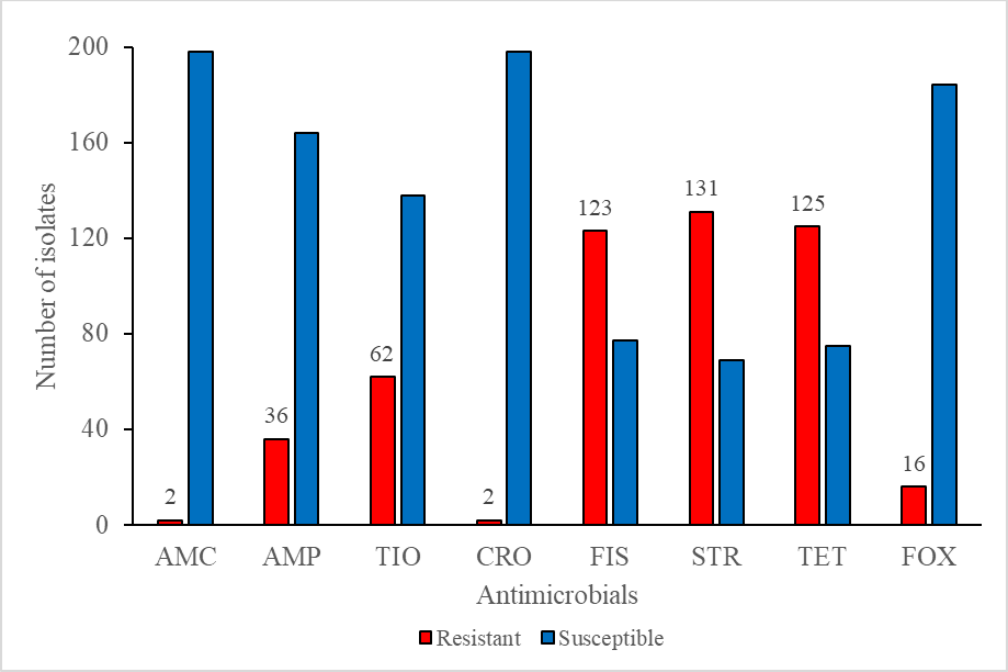
performance. Cohen's κ statistic assesses the level of concordance between observed and predicted classes and is an indicator of how much better one classifier performs than another random classifier that simply guesses based on the number of observations in each class. We obtained κ values above 0.7 for all models (**Table 3.2**). This is interpreted as a substantial (0.61–0.80) and almost perfect (0.81–1.0) agreement between the observed and predicted classes (Landis & Koch 1977).

3.4.3 AMR phenotype prediction and gene identification

Antimicrobial usage in poultry farming is widespread across the world and involves the use of drugs such as streptomycin, ceftiofur, ampicillin, and amoxicillin-clavulanic acid which have been recognized by the World Health Organization as critically important to human medicine (WHO, 2019). The inappropriate use of antimicrobials in poultry farming is a driver of AMR in *Salmonella* (Agyare et al., 2019).

The machine learning models shown in **Table 3.2** were used to predict the AMR phenotype of our 200 *Salmonella* isolates whose phenotypes were unknown. The proportion of isolates predicted to be resistant ranged from 1% in amoxicillin-clavulanic acid and ceftriaxone to 65.5% in streptomycin (**Figure 3.2**). Eighteen percent (18%), 61.5%, and 62.5% of the isolates were predicted to be resistant to ampicillin, sulfisoxazole, and tetracycline, respectively. Yin et al. (2021) reported that 61.8% of *Salmonella* isolates in chicken meat from conventional farms in the U.S. were resistant to at least one antimicrobial. In our case, our models predicted a much higher resistance level, 93% of the 200 isolates were resistant to at least one antimicrobial. Another study on *Salmonella* isolates from retail chicken in the U.S. reported a 52.2% prevalence in resistance to

both streptomycin and tetracycline (Lee et al., 2022). Lee et al. (2022) also reported a 4.4% and 21.7% prevalence in resistance to amoxicillin-clavulanic acid and sulfisoxazole, respectively.



AMC: amoxicillin-clavulanic acid; AMP: ampicillin; TIO: ceftiofur; CRO: ceftriaxone; FIS: sulfisoxazole; STR: streptomycin; TET: tetracycline; FOX: cefoxitin.

Figure 3.2. Machine learning predictions of AMR prevalence in 200 *Salmonella* isolates.

NARMS is an interagency program that monitors trends in AMR in foodborne pathogens in the U.S. Data on retail chicken from NARMS showed that the prevalence of *Salmonella* resistant to tetracycline ranged from 45.6% and 58.2% between 2016 and 2020 and was predicted to reach 66% in 2023% (FDA, 2024). The data also showed that resistance to amoxicillin-clavulanic acid and cefoxitin ranged from 2.2% to 13.3% of isolates between 2016 and 2020 and was expected to be 3.6% in 2023. Also, resistance to ceftriaxone and ampicillin was predicted to reach 12.4% in 2023 while resistance to streptomycin was 62.4% in 2020. Our machine learning prediction

(**Figure 3.2**) achieved comparable resistance prevalences for tetracycline, cefoxitin, streptomycin, and ampicillin. This shows the accuracy and reliability of our machine learning models.

Table 3.3. Comparison of predictions from machine learning modeling and the ResFinder database.

Antimicrobial	Machine learning prediction	ResFinder Prediction		Sensitivity (%)	Specificity (%)	Balanced accuracy (%)
		Resistant	Susceptible			
AMP	Resistant	28	8	96.55	95.32	95.94
	Susceptible	1	163			
TET	Resistant	122	3	96.83	95.95	96.39
	Susceptible	4	71			
FOX	Resistant	2	14	100.00	92.93	96.46
	Susceptible	0	184			

AMP: ampicillin; TET: tetracycline; FOX: cefoxitin.

The machine learning AMR phenotype predictions were compared with predictions from the ResFinder database. ResFinder acts as both a manually curated database of AMR genes and a bioinformatics pipeline (Bortolaia et al., 2020; Florensa et al., 2022). ResFinder performs phenotype predictions by identifying acquired genes and/or chromosomal mutations mediating AMR and linking these with known phenotypes from its database (Florensa et al., 2022). ResFinder predicts the AMR phenotypes of isolates as 'resistant' or 'no resistance'. 'No resistance' is interpreted as the absence of evidence of resistance in the ResFinder database, although resistance may exist from unidentified or unimplemented sources. For simplicity in comparison, we called these 'no resistance' phenotypes 'susceptible' in our study. We have only presented comparisons for ampicillin, tetracycline, and cefoxitin **(Table 3.3)** as these were the only antimicrobials from ResFinder prediction results that were used in our study. Our results in **Table 3.3** show sensitivity and specificity scores above 96% and 92%, respectively. This means for at least 92% of the time when our model predicted a *Salmonella* isolate to be susceptible to an

antimicrobial, ResFinder predicted the same. Similarly, for at least 96% of the time when the model predicted an isolated to be resistant, ResFinder predicted the same.

Genome-based machine learning methods are gaining momentum due to the growing availability of WGS data. Machine learning is particularly adept at handling the highly dimensional data typically produced from WGS studies (Karanth et al., 2022). Our results here show the applicability of machine learning to the identification of AMR phenotypes of foodborne pathogens. We have shown that *in silico* genome-based machine learning prediction of AMR genes can achieve similar results to genome-based predictions and by extension traditional antimicrobial susceptibility testing methods.

We also identified important genes that formed the basis for our machine learning modeling and prediction. The top 10 genes for each antimicrobial have been presented in **Table 3.4**. For amoxicillin-clavulanic acid, ampicillin, and ceftriaxone, *blaCMY-2* (beta-lactamase inactivation enzyme), *tet(C)* (major facilitator superfamily antibiotic efflux pump), and *sul2* (sulfonamide resistance protein) were identified as the three most important genes. Across other antimicrobials, the model identified genes including *arnA* (bifunctional polymyxin resistance protein), *aph(6)-Id* (aminoglycoside inactivation enzyme), and *vgaC* (ATP-binding cassette ribosomal protection protein).

Genes (or model features) identified as important by a machine learning algorithm do not always correlate with those that form the basis for AMR phenotype (or other physiological function) prediction by other bioinformatic resources due to differences in the methods underlying these predictions. ResFinder, for example, uses a rule-based method to detect the absence and

presence of AMR genes and resistance-conferring point mutations in an isolate to predict an AMR phenotype (Florensa et al., 2022). On the other hand, machine learning models (like in our study) use a pattern recognition approach, learning patterns and relationships from datasets and making predictions based on the learned information.

Table 3.4. Important genes used by machine learning models for classification.

AMC	AMP	TIO	CRO	FIS	STR	TET	FOX
blaCMY-2	*tet(C)*	*aph(6)-Id*	*blaCMY-2*	*tet(C)*	*aph(3'')-Ib*	*tet(C)*	*blaCMY-2*
tet(C)	*blaCMY-2*	*aph(3'')-Ib*	*tet(C)*	*sul2*	*aph(6)-Id*	*sul2)*	*sul2*
sul2	*sul2*	*tet(D)*	*sul2*	*arnA*	*tet(A)*	*arnA*	*fosA7*
ybtQ	*arnA*	*blaCMY-2*	*tet(A)*	*aac(6')-Iaa*	*tet(R)*	*fyuA*	*vgaC*
arnA	*ybtP*	*tet(A)*	*tet(R)*	*yptQ*	*tet(D)*	*sodC1*	*blaCTX-M-65*
aac(6')-Iaa	*ybtQ*	*tet(C)*	*aph(6)-Id*	*yptP*	*sodC1*	*golT*	*aadA*
aac(6')-Iy	*fyuA*	*sul2*	*tet(D)*	*aac(6')-Iy*	*adiY*	*blaCMY-2*	*adiY*
golT	*sinH*	*yptP*	*aph(3'')-Ib*	*gesC*	*vgaC*	*tet(R)*	*pmrG*
sodC1	*floR*	*sodC1*	*yptP*	*gesB*	*sul1*	*aph(6)-Id*	*gesC*
gesB	*sul1*	*aac(6')*	*fyuA*	*sinH*	*tet(C)*	*tet(A)*	*golT*

AMC: amoxicillin-clavulanic acid; AMP: ampicillin; TIO: ceftiofur; CRO: ceftriaxone; FIS: sulfisoxazole; STR: streptomycin; TET: tetracycline; FOX: cefoxitin.

Furthermore, machine learning approaches may deem certain genes as non-informative and exclude them from modeling if they occur above a certain threshold in the training data while a rule-based approach may have made predictions based on the same genes. Hence though we achieved results comparable to the ResFinder predictions, the underlying methods for these predictions differ. For example, ResFinder identified *blaCTX-M-65* as the genetic background for ampicillin resistance in most of our isolates, but this was not identified as one of the important genes for ampicillin resistance by our model. With large sample sizes and whole genomes, instead

of AMR genes only, a machine learning approach to AMR phenotype prediction could aid in the identification of genes that are typically not recognized as resistance markers but could be playing a role in a pathogen's AMR.

3.5 Conclusion

The multifaceted challenges posed by AMR demand proactive solutions. This global health crisis not only affects healthcare systems but also has profound implications for food safety, economies, and public health. Advancements in genomic technologies, particularly WGS, have revolutionized the surveillance and understanding of AMR. These technologies offer a more precise and comprehensive characterization of pathogens, aiding in the identification of antimicrobial resistance genes and mutations. Harnessing the predictive capabilities of machine learning, especially when combined with genomic data, offers a promising avenue for addressing AMR. By combining antimicrobial susceptibility phenotype and genomic data, we were able to develop a robust genome-based machine learning model for AMR prediction. This study underscores the potential of machine learning in predicting AMR in foodborne pathogens and its alignment with traditional antimicrobial susceptibility testing methods. These findings highlight the applicability and reliability of machine learning approaches in understanding AMR mechanisms and predicting resistance patterns, offering a valuable tool in combatting this global health threat. Further exploration and integration of such predictive models into surveillance systems could significantly enhance our ability to address and mitigate the growing concerns of antimicrobial resistance.

Chapter 4: Exploring the impact of population heterogeneity on risk estimates for *Salmonella enterica* in chicken

4.1 Abstract

Salmonella continues to pose a significant public health risk worldwide. Although it is well-established that *Salmonella* serovars may exhibit considerable differences in characteristics, these variations are typically not modeled in QMRA. The purpose of this study was to develop a QMRA to investigate the risk of illness associated with the consumption of *Salmonella*-contaminated chicken in the United States. This study incorporated *Salmonella* population heterogeneity by using different dose-response model parameters to account for differences in serovar virulence and finite mixture modeling to identify differences in growth rates. Across all developed models, the median number of illnesses was zero. The estimated number of illnesses at the 95th percentile ranged from 3,930 to 184,000, reflecting realistic estimates of the annual number of illnesses. In scenarios where 57% of *Salmonella* serovars were considered highly virulent, the estimated number of illnesses ranged from 173,000 to 184,000 (95th percentile). Reducing the proportion of high-virulence serovars to 50% and 0% resulted in a 27% and 97.7% reduction in illnesses, respectively. Furthermore, models that incorporated varying *Salmonella* growth rates estimated 6.6% more illnesses compared to those that did not. These findings highlight the impact of pathogen heterogeneity on risk estimates and the necessity of integrating this variability into QMRA models. Sensitivity analysis showed that at-home food preparation practices had the highest effect on the risk of chicken-associated salmonellosis. When 90% of consumers cooked chicken adequately and avoided cross-contamination, the reduction in illnesses was nearly as substantial as eliminating all high-virulence serovars. As regulators and industry

work to reduce *Salmonella* levels and the proportion of high-virulence serovars in poultry, consumer food safety education is needed to complement these efforts.

4.2 Introduction

Salmonella enterica continues to pose a major public health burden in the U.S. and around the world. The infection, salmonellosis, presents with symptoms such as nausea, diarrhea, vomiting, and abdominal pain which are typical of other bacterial gastroenteritis. Scallan et al. (2011a) reported that *Salmonella* approximately 1.03 million illnesses and 378 deaths annually in the U.S. Current estimates by the Centers for Disease Control and Prevention (CDC) put that figure at about 1.35 million illnesses and 420 deaths annually (CDC, 2024b). Foodborne illness source attribution for *Salmonella* by the Interagency Food Safety Analytics Collaboration (IFSAC) estimates that 18.6% of salmonellosis cases in the U.S. are attributable to chicken (IFSAC, 2023). Reports from the U.S. Department of Agriculture Food Safety Inspection Service (USDA FSIS) indicate that even though *Salmonella* prevalence in chicken has declined over the past decade, there has not been an accompanying decline in the number of illnesses (USDA FSIS, 2024b)

Salmonella enterica has over 2600 named serovars with considerable heterogeneity in several characteristics. For instance, a review of FSIS quarterly sampling reports indicates a difference in host preference for *Salmonella* serovars (USDA FSIS, 2024a). While serovars Kentucky, Infantis, Enteritidis, and Typhimurium, tend to be the most isolated from chicken samples, serovars Montevideo, Muenchen, Anatum, and Dublin dominate beef samples. Even for turkeys, which are evolutionarily related to chickens, a different set of serovars, Schwarzengrund, Reading, Agona, and Hadar, are more commonly isolated from turkeys. In the U.S., there are even

geographic differences in the distribution of *Salmonella* serovars; there are higher proportions of *S.* Infantis and *S.* Typhimurium in chicken samples from the Atlantic region whereas *S.* Schwarzengrund has higher proportions in the Southeast region (Siceloff et al., 2022).

Salmonella serovars exhibit differences in virulence as well. A vast majority of the more than 2600 serovars are not known to cause human infections. Serovars Infantis, Enteritidis, and Typhimurium are mostly associated with human illnesses in the U.S. while serovar Kentucky rarely causes human illnesses (Richards et al., 2023; USDA FSIS, 2022). The consensus on *S.* Kentucky is that it lacks many known virulence genes making it less harmful to humans. Tate et al. (2022) reported on the absence of virulence genes including *sseK2* (Type III secretion system effector), *lpfD* (fimbrial protein), and *tcf* (fimbrial protein-associated operon) in certain sequence types of *S.* Kentucky resulting in a lowered ability to invade, adhere, colonize, and survive in humans. Other serovars, such as Dublin and Choleraesuis, although responsible for only a small number of human illnesses, tend to cause invasive illnesses due to the presence of a Type 6 secretion system that facilitates their invasion into human cells (Cheng et al., 2019). Recent studies by Strickland et al. (2023) and Kim et al. (2024) have shown that most salmonellosis cases can be narrowed down to highly virulent serovars of *Salmonella*.

Differences in stress response and growth capacity have also been reported. Capita (2007) observed that *S.* Derby, *S.* Typhimurium, and *S.* Newport were more resistant to decontamination on poultry carcasses using sodium chlorite as compared to serovars Enteritidis, Infantis, Virchow, Agona, and Poona. Pye et al. (2023) also observed that *Salmonella* serovars exhibit varying growth rates when exposed to desiccation, acid, salt, heat, and cold stresses. Additionally, Díez-García et al. (2012) reported on differences in the growth kinetic parameters and biofilm-forming abilities

of *Salmonella* serovars isolated from poultry. Heterogeneity in *Salmonella* has been shown to not only occur between (inter) serovars but also within (intra) the same serovar (Aspridou et al., 2019; Pye et al., 2023). Given the diversity in *Salmonella* populations, they should be seen as a group of individuals rather than a homogenous population (Staes et al., 2019). Variations within a microbial population can be vital for risk assessment models, especially when the variations relate to pathogen virulence, growth and inactivation rates, and response to stresses encountered in the environment and foods.

It has long been known that these gene-level variations could affect risk assessment, even in closely related organisms (Buchanan et al., 2000; Coleman & Marks, 1999). Over the last decade, several studies have called for the integration of some form of genetic data into microbial risk assessments to capture gene-level differences in pathogens and have suggested approaches including the identification of genetic biomarkers of phenotypic properties and the identification of bacterial subpopulations that share specific genes, among others (Chen et al., 2020; Guillier et al., 2022; Haddad et al., 2018; EFSA Panel on Biological Hazards, 2019; NACMCF, 2024; Rantsiou et al., 2018). A few studies have implemented differing approaches in the attempt to incorporate pathogen heterogeneity into microbial risk assessment. Fritsch et al. (2018) categorized *Listeria* isolates into virulence groups based on clonal complexes to develop a cold-smoked salmon-related listeriosis risk model. Njage et al. (2020) also used machine learning to predict subpopulations of *Listeria* isolates when exposed to stressors and used this in a risk model for cultured milk. For *Salmonella*, Godínez-Oviedo et al. (2022) developed a risk model for salmonellosis in chicken meat that categorized *Salmonella* serovars into groups based on multidrug resistance and the presence of virulence genes. Strickland et al. (2023) also used different dose-response parameters for low and high-virulence serovars in their study on beef.

In this study, our objective was to develop a risk model for *Salmonella* in chicken meat. Our approach to modeling *Salmonella* heterogeneity focused on two key aspects: (i) analyzing *Salmonella* growth data using finite mixture modeling to identify any unobserved subpopulations growing at different rates, and (ii) using different dose-response parameters for low- and high-virulence *Salmonella* serovars.

4.3 Materials and Methods

4.3.1 Study overview

We developed a model to investigate the risk of human salmonellosis from the consumption of chicken meat. The major stages in the baseline model were on-farm contamination, processing, retail storage, transportation to home, home storage, cooking, and consumption. The model considered *Salmonella* in chicken feces as the main source of contamination in live birds. During processing, most of the steps result in a reduction in *Salmonella* levels on carcasses apart from the evisceration step where contamination may occur if the intestines are ruptured. During retail storage, transportation to home, and home storage, temperature plays a crucial role in the growth or decline of *Salmonella*. For the cooking stage, the model considered the decimal reduction of *Salmonella* during cooking and modeled the influence of food preparation actions such as handwashing and dishwashing on *Salmonella* levels. The consumption stage modeled the number of *Salmonella* cells that are ingested in each serving of chicken meat. For each of these stages, literature reviews were conducted to identify appropriate values and distributions that describe the variables under each stage. A summary of variables and parameters that were used in the baseline model has been provided in **Table 4.1**.

Since our study sought to incorporate *Salmonella* heterogeneity into the QMRA model, we have used different dose-response parameters to describe the probability of illness of both low- and high-virulence *Salmonella* serovars. *Salmonella* serovars can have considerable inter and intra-variability in growth and survival profiles even when exposed to similar conditions (Aspridou et al., 2019; Capita, 2007; Díez-García et al., 2012; Nair et al., 2015; Pye et al., 2023). Therefore, we assumed that for any given population there could be distinct subpopulations with different growth rates. A finite mixture modeling approach was used to analyze *Salmonella* growth data to identify these subpopulations, their proportions, and their mean growth rates. The resulting proportions and growth rates were then used in the models.

4.3.2 *Salmonella* growth data

Salmonella growth data was collected from the ComBase database (https://combase.errc.ars.usda.gov/). The search criteria used, "*Salmonella* spp" as the organism and "Poultry" as the food category, returned more than 1400 records. These records were downloaded and filtered based on these inclusion criteria: (i) data was from experiments on chicken and (ii) data included microbial counts at 'time zero' (this was required for growth curve fitting in ComBase). After further exploration of the data that met the above criteria, data on experiments carried out at 30 °C were selected for further analysis. This temperature, 30 °C, was selected because it falls within the growth temperature range of *Salmonella* and had the most available experimental data ($n = 68$). The 68 experiments were from studies by Bovill et al. (2000) and (Oscar, 2006, 2007, 2009a, 2009b, 2011, 2012). In the current study, each of the 68 observations was assumed to represent a *Salmonella* isolate, providing growth data on 68 individual *Salmonella* isolates. The data were then fitted to primary growth curves, using the

DMFit web-based application in ComBase, to determine the maximum growth rate, μ_{max}. For all 68 experimental data, ComBase used the (Baranyi & Roberts, 1994) model:

$$y(t) = y_0 + \mu_{max}A(t) - \frac{1}{m}\ln\left(1 + \frac{e^{m\mu_{max}A(t)} - 1}{e^{m(y_{max}-y_0)}}\right) \quad (4.1)$$

$$A(t) = t + \frac{1}{\mu_{max}}\ln(e^{-vt} + e^{-h_0} - e^{-vt-h_0}) \quad (4.2)$$

where *y(t)* is the cell concentration (ln CFU/g) at time, *t*; y_0 is the initial (time 0) cell concentration (ln CFU/g); μ_{max} is the maximum specific growth rate (h^{-1}); *A(t)* is an adjustment function that describes the adjustment of bacterial culture to a new environment (lag phase) and the transition into an exponential phase; *m* is a curvature parameter that describes the transition from the exponential phase; y_{max} is the maximum cell concentration (ln CFU/g); *v* is the rate of production of the limiting substrate; and h_0 is a parameter that describes the initial physiological state of bacterial cells and is typically constant. Mathematically, h_0 is defined as the product of the maximum specific growth rate, μ_{max}, and the lag time, λ. Parameters *m* and *v* are assumed to be equal to 1 and μ_{max}, respectively (Baranyi, 1997; Grijspeerdt & Vanrolleghem, 1999). Following the curve fitting the growth rates, μ, obtained were analyzed via finite mixture modeling.

4.3.3 Finite mixture modeling

Homogeneity is often assumed for many datasets when in reality they may be heterogeneous and have distinct unobserved clusters within the population. Histograms of such heterogeneous data usually show a multimodal distribution. Finite mixture modeling is a statistical method used to model datasets that are composed of multiple underlying subpopulations or components (Bohning et al., 1992; Figueiredo & Jain, 2002; Mclachlan et al., 2019; Schlattmann

et al., 2015). It is assumed that a random variable, Y (growth rate, μ, in this study) follows a finite mixture distribution if its probability density function takes the form:

$$p(y|\theta) = \sum_{i=1}^{k} \pi_i p(y|\theta_i) \tag{4.3}$$

where $\pi_1,\ldots,\pi_k$ are the mixing proportions (or probabilities); $p(y|\theta_i)$ is the probability density function of the i-th component; and θ_i is the parameter vector of i-th component. Since π_i are probabilities, they must be non-negative and sum up to 1 (**Equation 4.4**).

$$\pi_i \geq 0, k = 1, \ldots, k \text{ and } \sum_{i=1}^{k} \pi_i = 1 \tag{4.4}$$

The objective of finite mixture modeling is to estimate both the mixing proportions, π_i, and the parameters, θ_i, of each component (subpopulation), k. The solution for these parameters uses the Expectation-Maximization (EM) and Variational Expectation-Maximization (VEM) algorithms (Bohning et al., 1992; Figueiredo & Jain, 2002; Mclachlan et al., 2019). The EM algorithm alternates between two steps: the expectation (E) step, which constructs a function for the conditional expectation of the log-likelihood [log $p(y|\theta)$] (**Equation 4.5)** that corresponds to k-components of an observed variable Y, and the maximization (M) step which computes parameters to maximize the log-likelihood in the E-step (Bohning et al., 1992; Figueiredo & Jain, 2002; Mclachlan et al., 2019). The algorithm iterates between these two steps allowing for the computation of the local maxima of the maximum a posteriori (**Equation 4.6)** or maximum likelihood (**Equation 4.7**) estimates of the components, k. The VEM algorithm generally behaves similarly, but in the E-step, a variational inference is computed to approximate the posterior distributions of the parameters (Beal & Ghahramani, 2000).

$$\log p(y|\theta) = \log \prod_{i=1}^{n} p\left(y^{(i)}|\theta\right) = \sum_{i=1}^{n} \log \sum_{i=1}^{k} \pi_i p(y|\theta) \tag{4.5}$$

$$\hat{\theta}_{\mathrm{MxL}} = \arg\max_{\{\theta\}} \{\log p(y|\theta)\} \tag{4.6}$$

$$\hat{\theta}_{\mathrm{MAP}} = \arg\max_{\{\theta\}} \{\log p(y|\theta) + \log p(\theta)\} \tag{4.7}$$

Finite mixture modeling to determine the number of components, *k*, in the *Salmonella* growth data, was performed in R (v. 4.4.1, R Core Team, 2024) using the CAMAN package (v.0.7.8, Schlattmann et al., 2023). First, the 'mixalg' function, a hybrid algorithm that combines both the EM and VEM algorithms, for variable support size and fixed number of components, respectively was employed. All 'mixalg' arguments were set to their default values except the number of iterations which was set to 10,000 and the underlying density function which was set to Gaussian since the *Salmonella* growth rate was continuous data. Briefly, 'mixalg' creates an approximating grid and applies the VEM algorithm to find grid points with positive support to estimate *k* (Schlattmann et al., 2023). The *k* components and mixing proportions, π, are then used as seed values for the EM algorithm. 'Mixalg' then collapses similar components leaving distinct ones only. This process is iterated until convergence is achieved. The results from the 'mixalg' function were then used as starting values to construct five mixture models, A – E, using the 'mixalg.EM' function. Each model tested for a different number of components (*k*) in the data: Model A (*k* = 1), Model B (*k* = 2), Model C (*k* = 3), Model D (*k* = 4), and Model E (*k* = 5).

During model fitting, the EM and VEM algorithms calculated posterior probabilities (Equation 4.8) to classify each of the data points (*Salmonella* isolates; *n* = 68) into various mixture components to indicate which component each isolate is most likely to belong to. The posterior probability that a *Salmonella* isolate belongs to a specific component is

$$\gamma_i(y) = P(Z = i|y) = \frac{\pi_i p(y|\theta_i)}{\sum_{j=1}^{k} \pi_j p(y|\theta_j)} \quad (4.8)$$

where $\gamma_i(y)$ is the posterior probability that a given observation (a *Salmonella* isolate), y, belongs to the i-th mixture component; $P(Z = i|y)$ is another form of $\gamma_i(y)$, the posterior probability that component Z equals i, given an observation, y; and $\sum_{j=1}^{k} \pi_j p(y|\theta_j)$ is the weighted sum of the likelihoods of y under each component, ensuring that $\gamma_i(y) = 1$ across all k components of the mixture. For models B – E, where more than one component was evaluated, *Salmonella* isolates were assigned to components for which it had the highest posterior probability.

The Bayesian Information Criterion (BIC) was used for model selection:

$$\text{BIC} = -2 \ln L + k \ln N \quad (4.9)$$

where L is the likelihood of the model; k is the number of parameters, and N is the number of observations. During finite mixture modeling, the likelihood can be improved by the addition of more parameters; however, this can lead to overfitting. The BIC does a trade-off between model complexity and model fit by penalizing the number of parameters. A lower BIC value indicates a better model. Additionally, an analysis of variance (ANOVA) was performed to compare model fits.

4.3.4 Secondary growth modeling

Temperature varies widely across the chicken production and supply chain. Therefore, to determine *Salmonella* growth rates at specific temperatures during retail storage, transportation,

home storage, and thawing, we used the Ratkowsky et al. (1982) square root model for secondary growth modeling:

$$\sqrt{\mu} = b(T - T_{min}) \quad (4.10)$$

where b is the growth rate coefficient; T is the temperature (°C) and T_{min} is the hypothetical minimum growth temperature, considered an inherent property of bacteria. To implement the Ratkowsky model, we first estimated b. For this study, we used a T_{min} value of 5.7 °C based on Oscar (2002). The temperature, T, was set to 30 °C, which corresponds to the temperature in the primary growth data used in our study. The growth rate, μ, was based on the mean growth rates obtained from the finite mixture modeling. We then calculated the b values which were subsequently used to compute *Salmonella* growth rates under different temperature conditions.

4.3.5 Baseline QMRA model

4.3.5.1 Farm

Poultry farms are a significant source of *Salmonella* contamination. *Salmonella* can enter and colonize the gastrointestinal tracts of chickens through sources such as litter, feed, water, pests, and wild animals (Wang et al., 2023). *Salmonella* can be shed through chicken feces, which then contaminates other birds in the same house. In a study on *Salmonella* risk profiles in broiler farms, Obe et al. (2023) reported an average level of 3.6 log CFU/sample, with 85% of the farms being positive for *Salmonella*. Gutierrez et al. (2020) also reported a *Salmonella* prevalence of 61% in poultry litter. *Salmonella* prevalence of 46% and levels of 0.45 to >280,000 MPN/g of litter have also been reported by Dunn et al. (2022). These studies suggest a wide variation in *Salmonella*

levels across broiler farm complexes in the U.S., with various factors such as sampling location, bird age, litter age, and farm environment influencing *Salmonella* levels.

For this study, our model for the farm began at the pre-harvest stage, i.e., immediately before transporting birds for slaughter. The exterior (feathers and feet) of chickens on the farm is typically soiled with feces, which is a reservoir for *Salmonella*. Collineau (2020a) estimated the amount of feces on the exterior of broilers at pre-harvest to be PERT distributed with minimum, most likely, and maximum values of 1 g, 10 g, and 50 g, respectively. Collineau et al. (2020b) developed a within-flock model for the transmission of *Salmonella* and estimated the concentration in broiler houses at pre-harvest to be 3.579 ± 3.906 log CFU/g of feces. In this study, the *Salmonella* concentration has been described with a normal distribution and truncated at the minimum (0) and maximum (4.294) values reported by Collineau (2020b).

The transportation of chickens from farms to processing facilities typically results in increased contamination either through bird-to-bird contact or bird-to-truck contact. Furthermore, transportation-induced stress in birds may also result in an increase in bacterial shedding in feces, thus increasing contamination levels (Whyte et al., 2001). Dogan et al. (2019) used a normal distribution (μ = 0.807 log CFU, σ = 0.076 log CFU) for the change in concentration of *Campylobacter* during transport. In this study, we assumed that the change in concentration of *Salmonella* on the exterior chickens during transport would be the same as that for *Campylobacter*.

Table 4.1. Summary of variables and distributions used in the baseline QMRA model

Symbol	Variable	Distribution/Value/Formula	Unit	Source
Farm				
W_{fc}	Amount of feces on bird	RiskTriang(1,10,50)	g	Collineau et al. (2020a)
C_{fc}	*Salmonella* level in feces	RiskNormal(3.579,3.906, RiskTruncate(0,4.294))	log CFU/g	Collineau et al. (2020b).
C_{sb}	*Salmonella* level on bird exterior	$(10\text{^}C_{fc}) \times W_{fc}$	CFU/bird	Calculated
C_{tran}	Change in level during transport	10^RiskNormal(0.807,0.0706)	CFU	Dogan et al. (2019)
C_{stran}	*Salmonella* level in birds after transport	$C_{sb} + C_{tran}$	CFU/bird	Calculated
Processing				
Scalding				
R_{sc}	Reduction in *Salmonella* level due to scalding	RiskLogistic(1.12,0.06)	log CFU	Xiao et al. (2019)
C_{sc}	*Salmonella* level after scalding	LOG $C_{stran} - R_{sc}$	log CFU/carcass	Calculated
Defeathering				
R_{df}	Reduction in *Salmonella* level due to defeathering	RiskLoglogistic(-0.56132,1.9336,4.3467, RiskTruncate(,4.35))	log CFU	Allen et al. (2003); Clouser et al. (1995)
C_{df}	*Salmonella* level after defeathering	$C_{sc} - R_{df}$	log CFU/carcass	Calculated
Evisceration				
Ev	Contamination during evisceration	RiskUniform(0.7,5.1)	log CFU	González et al. (2019)
L_{Ev}[a]	Likelihood of contamination	RiskBinomial(1,0.3)		Rivera-Pérez et al. (2014)
C_{ev}	*Salmonella* level after evisceration	IF($L_{Ev} = 1$, $C_{df} + Ev$, C_{df})	log CFU/carcass	Calculated
Washing				
R_{wa}	Reduction in *Salmonella* level due to washing	RiskUniform(0.12708,2.8029)	log CFU	Northcutt et al. (2005); Wang et al. (1997);

				Xiong et al. (1998); Yang et al. (1998)
C_{wa}	*Salmonella* level after washing	$C_{ev} - R_{wa}$	log CFU/carcass	Calculated
Chilling				
R_{ch}	Reduction in *Salmonella* level during chilling	RiskTriang(-0.55, 1.35, 3.24)	log CFU	Collineau et al. (2020a)
C_{ch}	*Salmonella* level after chilling	$C_{wa} - R_{ch}$	log CFU/carcass	Calculated
W_{carc}	Ready-to-cook weight per carcass	2262.9	g	USDA NASS (2024)
C_{chl}	*Salmonella* level after chilling (log CFU/g)	LOG((10^C_{ch}) ÷ W_{carc})	log CFU/g	Calculated
Cold storage				
R_{cs}	Reduction in *Salmonella* level during cold storage	RiskPert(0,0.3,0.7)	log CFU	Xiao et al. (2021)
C_{proc}	*Salmonella* level at the end of processing	$C_{chl} - R_{cs}$	log CFU/carcass	Calculated
Retail storage				
T_{min}	Minimum growth temperature	5.7	°C	Oscar (2002)
b	Growth model parameter	0.0332		Calculated
T_{rs}	Retail storage temperature	RiskLaplace(3.3333,2.6623, RiskTruncate(-6.67,19.44))	°C	Ecosure (2007a)
t_{rs}	Retail storage time	RiskTriang(1, 72,168)	h	Authors' estimate
μ_{retail}	Retail growth rate	IF($T_{rs} > T_{min}$,(b × ($T_{rs} - T_{min}$))^2,0)	log CFU/g/h	Calculated
G_{retail}	Growth during retail storage	$t_{rs} \times \mu_{retail}$	log CFU/g	Calculated
C_{retail}	*Salmonella* level after retail storage	$C_{proc} + G_{retail}$	log CFU/g	Calculated
Prev	*Salmonella* prevalence in retail chicken	7.67	%	Sodagari et al. (2024)
Transportation to home				
T_{home}	Temperature at end of transport to home	RiskLoglogistic(-22.915, 29.371,16.763, RiskTruncate(-5.56,20))	°C	Ecosure (2007a)

T_{tran}	Temperature during transport	$(T_{rs} + T_{home}) \div 2$	°C	Calculated
t_{tran}	Transportation time	RiskExtvalue(0.96904,0.35127)	h	Ecosure (2007a)
μ_{tran}	Transportation growth rate	IF($T_{tran} > T_{min}$,(b × ($T_{tran} – T_{min}$))^2,0)	log CFU/g/h	Calculated
G_{tran}	Growth during transportation	$t_{tran} \times \mu_{tran}$	log CFU/g	Calculated
C_{tran}	*Salmonella* level after transportation	$C_{retail} + G_{tran}$	log CFU/g	Calculated
Home storage				
Refrigeration				
T_{rfg}	Home refrigerator temperature	RiskLaplace(4.4444,2.5231, RiskTruncate(-0.56,15.56))	°C	Bruhn (2014)
t_{rfg}	Refrigeration storage time	RiskWeibull(1.0279,2.1976,RiskShift(0.24392),RiskTruncate(,7)) × 24	h	Mazengia et al. (2015)
μ_{rfg}	Growth rate during refrigerator storage	IF($T_{rfg} > T_{min}$,(b × ($T_{rfg} – T_{min}$))^2,0)	log CFU/g/h	Calculated
G_{rfg}	Growth during refrigerated storage	$t_{rfg} \times \mu_{rfg}$	log CFU/g	Calculated
C_{rfg}	*Salmonella* level after refrigerated storage	$C_{tran} + G_{rfg}$	log CFU/g	Calculated
Thawing				
Frz[b]	Proportion of consumers that freeze chicken	RiskBinomial(1,0.4)		Mazengia et al. (2015)
Surv[c]	Probability of *Salmonella* survival during freezing	RiskBinomial(1,0.33)		Collineau et al. (2020a)
Thaw[d]	How is frozen chicken thawed?	RiskDiscrete({1,2,3,4}, {0.14,0.14,0.24,0.48})		Mazengia et al. (2015)
tt_1	Thawing time 1	RiskUniform(1,6)	h	Authors' estimate
T_{room}	Room temperature	RiskPert(13.89,21.11,30)	°C	Booten et al. (2017)
μ_{t1}	Growth rate during countertop thawing	IF($T_{room} > T_{min}$,(b × ($T_{room} – T_{min}$))^2,0)	log CFU/g/h	Calculated
G_{t1}	*Salmonella* growth during countertop thawing	IF(AND(Surv = 1, Thaw = 1), $tt_1 \times \mu_{t1}$,0)	log CFU/g	Calculated
tt_2	Thawing time 2	RiskUniform(0.5,2)	h	Authors' estimate
T_{run}	Running water temperature	RiskUniform(15,37)	°C	Authors' estimate

μ_{t2}	Growth rate during water thawing	IF($T_{run} > T_{min}$,(b × ($T_{run} – T_{min}$))^2,0)	log CFU/g/h	Calculated
G_{t2}	*Salmonella* growth during water thawing	IF(AND(Surv = 1, Thaw = 2), $tt_2 \times \mu_{t2}$,0)	log CFU/g	Calculated
tt_3	Thawing time 3	RiskUniform(0.17,0.5)	h	Authors' estimate
T_{mic}	Temperature during thawing	RiskUniform(-4,7)	°C	Kim et al. (2011)
μ_{t3}	Growth rate during microwave thawing	IF($T_{mic} > T_{min}$,(b × ($T_{mic} – T_{min}$))^2,0)	log CFU/g/h	Calculated
G_{t3}	*Salmonella* growth during microwave thawing	IF(AND(Surv = 1, Thaw = 3), $tt_3 \times \mu_{t3}$,0)	log CFU/g	Calculated
tt_4	Thawing time 4	RiskPert(6,24,48)	h	Authors' estimate
μ_{t4}	Growth rate during refrigerator thawing	IF($T_{rfg} > T_{min}$,(b × ($T_{rfg} – T_{min}$))^2,0)	log CFU/g/h	Calculated
G_{t4}	*Salmonella* growth during refrigerator thawing	IF(AND(Surv = 1, Thaw = 4), $tt_4 \times \mu_{t4}$,0)	log CFU/g	Calculated
C_{thaw}	*Salmonella* level after thawing	$C_{tran} + G_{t1} + G_{t2} + G_{t3} + G_{t4}$	log CFU/g	Calculated
C_{home}	*Salmonella* level at end of home storage	IF(Frz = 1, 10^C_{thaw}, 10^C_{rfg})	CFU/g	Calculated
Cooking				
Trf_{r2h}	*Salmonella* transfer rate from raw chicken to hands	RiskTriang(0.0045,0.0045,0.11667)		Montville et al. (2001)
N_{hand}	Number of cells on hands	$C_{home} \times Trf_{r2h}$	CFU	Calculated
C_{chick}	Level remaining on chicken	$C_{home} – N_{hand}$	CFU/g	Calculated
Trf_{r2u}	*Salmonella* transfer rate from raw chicken to utensils	RiskNormal(0.0125,0.0009)		Ravishankar et al. (2010)
N_{uten}	*Salmonella* level on utensils	$C_{chick} \times Trf_{r2u}$	CFU	Calculated
C_{chick2}	Level remaining on chicken	LOG($C_{chick} – N_{uten}$)	log CFU/g	Calculated
T_{cook}	Cooking temperature	RiskExtvalueMin(78.5098,10.5962, RiskTruncate(24.4444,104.4444))	°C	Ecosure (2007b)
t_{cook}	Cooking time	RiskUniform(15,50)	min	Authors' estimate
z	z-value	7.6	°C	Murphy et al. (2002)

D_{60}	D-value at reference temperature (60 °C)	8.5	min	Murphy et al. (2002)
D_{Tcook}	D-value at cooking temperature	$D_{60} \times 10^{((60 - T_{cook}) \div z)}$	min	
$Prop_{under}$[e]	Proportion of consumers who undercook	RiskBinomial(1,0.4034)		Ecosure (2007b)
C_{cook}	*Salmonella* level after cooking	$IF(Prop_{under} = 1,10^{(C_{chick2} - (T_{cook} \div t_{cook}))},0)$	CFU/g	Calculated
$Prop_{hw}$[f]	Proportion of consumers who wash hands properly	RiskBinomial(1,0.804)		Bruhn (2014); Kosa et al. (2015, 2019)
Trf_{h2c}	*Salmonella* transfer rate from hands to cooked chicken	RiskPert(0.00145, 0.0893, 0.5295)		Smadi & Sargeant (2013)
C_{ckhw}	*Salmonella* level on cooked chicken due to improper handwash	$IF(Prop_{hw} = 1,0, Trf_{h2c} \times N_{hand})$	CFU/g	Calculated
$Prop_{uten}$[g]	Proportion of consumers who properly wash or use separate utensils	RiskBinomial(1,0.8249)		DeDonder et al. (2009); Kosa et al. (2015)
Trf_{u2c}	Transfer rate from utensils to cooked chicken	RiskPert(0.0302, 0.075, 0.3096)		Smadi & Sargeant (2013)
C_{ckut}	*Salmonella* level on cooked chicken due to utensil use	$IF(Prop_{uten} = 1,0, Trf_{u2c} \times N_{uten})$	CFU/g	Calculated
C_{cons}	*Salmonella* level before consumption	$C_{cook} + C_{ckhw} + C_{ckut}$	CFU/g	Calculated
Consumption				
S	Serving size	85	g	9 CFR. 381.412
D	Dose per serving	$C_{cons} \times S$	CFU/serving	Calculated
Av_{chick}	Loss-adjusted availability of chicken in the US	24766.14	g/person-yr	USDA ERS (2021)
US_{pop}	US population	335,888,625		US Census Bureau (2024)

N_{serv}	Consumed servings per year in the US	$(Av_{chick} \div S) \times US_{pop}$	Calculated
Dose-response and Risk characterization			
α	Beta-Poisson parameter	0.1324	WHO/FAO (2002)
β	Beta-Poisson parameter	51.45	WHO/FAO (2002)
P_{ill}	Probability of illness per serving	$(1 – (1 + D \div \beta)^{\wedge} - \alpha) \times (Prev \div 100)$	Calculated
N_{ill}	Number of illnesses per year	$N_{serv} \times P_{ill}$	Calculated

[a] 1: Contamination occurs; 0: No contamination occurs.
[b] 1: Consumer freezes chicken; 0: Consumer does not freeze chicken.
[c] 1: Cell survives during freezing; Cell does not survive freezing.
[d] 1: Counter top; 2: Running water; 3: Microwave; 4: Refrigerator.
[e] 1: Undercooking; 0: Adequate cooking.
[f] 1: Proper handwashing; 0: No or improper handwashing.
[g] 1: Proper washing of utensils or separate utensils used; 0: Improper washing or utensils reused.

4.3.5.2 Processing

At processing facilities, chicken processing begins with stunning which may be achieved using an electrical water bath, dry electrodes, a controlled atmosphere of gas mixtures, or low atmosphere pressure systems. Stunning renders birds unconscious and is followed by slaughtering and bleeding the chickens for 60 – 133 s (Berg & Raj, 2015; Wabeck 2002). In this study, we have assumed that no significant changes in *Salmonella* concentration occur during stunning, slaughtering, and bleeding; hence these steps were not modeled.

Scalding is the immersion of carcasses in warm water to enlarge feather follicles and increase feather density as well as friction with picker fingers to make feather removal easier (Dogan et al., 2022; Wabeck, 2002). Scalding can be done at water temperatures of 51.4 – 53.1 °C (soft scald) or 55 – 58 °C (hard scald) for anywhere between 1.5 to 2.5 min (Dogan et al., 2022; Wabeck, 2002). This process typically results in a reduction in *Salmonella* numbers due to the washing of fecal material and dirt from feathers and the thermal inactivation of *Salmonella* at the temperatures used. Xiao et al. (2019) observed a log reduction of logistic (-1.12 log CFU, 0.06 log CFU). This distribution has been used in our QMRA model.

The defeathering step makes use of vibrating rubber fingers (pickers) to remove feathers from the carcasses. Defeathering results in a reduction in *Salmonella* numbers on the exterior of the carcass as the feathers are removed. Clouser et al. (1995) reported log reductions of 0.8 to 1.35 log CFU during defeathering. Another study by Allen et al. (2003) reported log reductions of 1.06 to 4.35 log CFU. The results from these studies were fitted to a log-logistic distribution with parameters γ = -0.56132 log CFU, β = 1.9336 log CFU, and α = 4.3467 log CFU.

The next step, evisceration, is where the removal of the carcass viscera takes place. *Salmonella* in the internal parts of chicken is resident in the gastrointestinal tract. This makes evisceration a crucial step in chicken processing as the rupture of visceral organs would result in contamination of the carcass, thus increasing the *Salmonella* numbers on the carcass. Not all instances of evisceration result in rupturing of the viscera. Rivera-Pérez et al. (2014) in a study on *Salmonella*-contaminated carcasses in a commercial processing plant, reported a 30% chance of contamination. Therefore, the likelihood of contamination was represented with a binomial distribution (n = 1, p = 0.3). If contamination occurred, a uniform distribution (min = 0.7 log CFU, max = 5.1 log CFU) from González et al. (2019) was used to describe the increase in *Salmonella* concentration during evisceration.

Carcasses are then washed to further remove any fecal material and reduce *Salmonella* levels. Washing can be achieved using equipment such as spray washers or inside-outside bird washers and may use chemical additives to enhance the reduction of bacterial load (Collineau et al., 2020a; Dogan et al., 2022). Several studies have investigated the effect of water temperature, washer type, and chemical aids on the reduction of *Salmonella* (Northcutt et al., 2005; Wang et al., 1997; Xiong et al., 1998; Yang et al., 1998). Data on log reduction due to washing were collected from these studies and fitted to a uniform distribution (min = 0.12708 log CFU, max = 2.8029 log CFU).

Following washing, carcasses are chilled. Chilling rapidly reduces the temperature of carcasses to about 2.5 – 5 °C (Wabeck, 2002). The reduction in carcass temperature ensures a low rate of microbial growth and is also important for maintaining meat quality. Chilling can be achieved using air chilling (cold air) or immersion chilling (cold water), where chemical aids can

be added to the water to aid in bacterial load reduction. Immersion chilling is more common in the U.S. while air chilling is common in other countries (Dogan et al., 2022; Wabeck, 2002). A fitted triangular distribution from Collineau et al. (2020a) was used to describe the reduction of *Salmonella* concentration on chicken carcasses during chilling.

To estimate the level of *Salmonella* in log CFU/g at the end of chilling, we considered the ready-cook-weight of broiler carcasses. From USDA Poultry Slaughter data for December 2023, the average live weight of a broiler was 6.61 lb, the total live weight of broilers slaughtered was 4,844,740,000 lb, and the total ready-cook-weight of broilers was 3,656,530,000 lb (USDA NASS, 2024). Using these values, we calculated a product yield of 75.47% and determined the ready-to-cook weight of a broiler to be 2262.9 g (75.47% of 6.61 lb). It was assumed that the *Salmonella* cells remaining on carcasses after the chilling step were evenly distributed over 2262.9 g of carcass. After chilling, chicken carcasses may go through various steps of portioning, where carcasses are cut up into parts before being packed and cold-stored. A cold chain is also maintained during the distribution of chicken from processing facilities to retailers. Cold storage is vital for food safety as it ensures chicken meat is maintained at temperatures that do not support *Salmonella* growth. The log reduction during cold storage was estimated to be PERT (min = 0 log CFU, most likely = 0.3 log CFU, max = 0.7 log CFU) (Xiao et al., 2021).

4.3.5.3 Retail storage

The growth of *Salmonella* in chicken meat during retail storage depends on the storage time and temperature. The growth rate was determined using Ratkowsky's equation (**Equation 4.10)**. It was estimated that chicken was stored in retail refrigerators between one hour and seven

days before being purchased by a consumer. Retail refrigerator temperatures were extracted from the EcoSure Cold Temperature Report (EcoSure 2007a**)** and fitted to a Laplace distribution (μ = 3.3333 °C, σ = 2.6623 °C). The distribution was truncated within the bounds of the minimum (-6.67°C) and maximum (19.44 °C) temperatures found in the report. At the retail market, long storage of chicken meat at high temperatures would lead to spoiled products which would have to be discarded. To prevent the combination of these conditions (long storage time + high temperature) during simulations, a correlation factor of -0.75 was used. This approach has been used by Golden and Mishra (2021), Smadi and Sargeant (2013), and WHO/FAO (2002).

4.3.5.4 Transportation to home

Salmonella growth rate during transportation to home as well as subsequent stages where growth was modeled was calculated using **Equation 4.10**. Transportation times were collected from the EcoSure Cold Temperature Report (EcoSure 2007a) and fitted to an extreme value distribution with a and b parameters of 0.96904 and 0.35127, respectively. The EcoSure report was also the data source for the temperature of chicken meat at the end of transportation to the home. Since the actual temperature during transportation was not known, it was assumed to be the mean of the retail temperature and temperature at the end of transportation to home based on Pang et al. (2017) and Collineau et al. (2020a). Similar to the retail storage stage, a correlation matrix was set up to define the relationship between transportation temperature and time.

4.3.5.5 Home storage

According to an observational study by Mazengia et al. (2015), 58% of consumers stored raw poultry in the refrigerator, 40% of consumers stored it in the freezer, and 2% used the poultry on the day of purchase. Since only 2% of consumers used poultry on the same day, this occurrence was not modeled. It was assumed that all consumers would either refrigerate or freeze chicken upon reaching home, thus a binomial distribution (1, 0.4) was used to represent the proportion of consumers who froze chicken. Mazengia et al. (2015) also reported that almost half of consumers (49%) stored poultry in the refrigerator for less than two days while 10% of consumers stored poultry for more than five days. This data was fitted to a Weibull distribution and truncated at seven days (1.0279, 2.1976, Shift(0.24392), Truncate(,7)). Refrigerator temperatures obtained from Bruhn (2014) were used in estimating *Salmonella* growth during refrigerator storage. Once again, a correlation factor of -0.75 was used to define the relationship between refrigeration temperature and time.

A binomial distribution (1, 0.33) was used to model *Salmonella* survival during freezing (Collineau et al., 2020a). If a *Salmonella* cell survived freezing, it may grow during thawing when the thawing temperature falls within the 'danger zone' (4.4 – 60 °C). This makes thawing a crucial step in maintaining food safety. In the Mazengia et al. (2015) study, it was reported that consumers thaw frozen chicken at room temperature (14%), under running water (14%), in a microwave oven (24%), or in a refrigerator (48%). These proportions were described with a discrete distribution in our model. The thawing times and temperatures were either estimated by authors or obtained from previous studies (Booten et al., 2017; Bruhn, 2014; Kim et al., 2011).

4.3.5.6 Cooking

Cooking represents an important risk reduction step as inactivation of *Salmonella* occurs at elevated temperatures, however low cooking temperatures or short cooking times could lead to undercooked chicken with live *Salmonella* cells. Additionally, cross-contamination can occur during food preparation, potentially reintroducing live *Salmonella* cells to previously well-cooked chicken. For cross-contamination before cooking, our model evaluated the transfer of *Salmonella* from raw chicken to hands and then the transfer of *Salmonella* remaining on the raw chicken to utensils. *Salmonella* transfer rates from raw chicken to hand were obtained from Montville et al. (2001) and fitted to a PERT distribution (min = 0.0045, most likely = 0.0045, max = 0.11667) while the transfer rate for chicken to utensils was obtained from Ravishankar et al. (2010).

After cooking, *Salmonella* can be transferred to the cooked chicken from contaminated hands and utensils if consumers do not properly wash them after handling raw chicken. Bruhn (2014) investigated consumer food preparation practices in an observational study and reported that only 62% of consumers washed their hands after handling raw chicken. Other studies have reported handwashing rates of 87.8% and 83.1% (Kosa et al., 2015, 2019). These percentages were fitted to a PERT distribution, and the mean of the distribution, 80.4%, was used to represent the proportion of consumers who washed their hands after handling raw chicken. DeDonder et al. (2009) reported that 69.23% of consumers used separate utensils to handle raw and cooked chicken while Kosa et al. (2015) reported that 95.75% of consumers used separate utensils or properly washed utensils before reuse. The mean of these two studies, 82.49%, was used to represent the proportion of consumers who either washed or used separate utensils for raw and cooked chicken. During simulations, when handwashing or utensil washing occurred, it was assumed that no cross-

contamination occurred. Otherwise, rates for *Salmonella* transfer from contaminated hands and utensils to cooked chicken were applied (Smadi & Sargeant, 2013).

Cooking temperatures were extracted from the EcoSure Cooking Temperature Report (Ecosure 2007b) and fitted to a minimum extreme value distribution. The proportion of consumers who undercook chicken was estimated to be 40.34% based on temperatures in the EcoSure report that were below the USDA FSIS-recommended minimum internal temperature for cooked chicken, 73.89 °C (USDA FSIS, 2019). During simulations, if there was adequate cooking, it was assumed that all *Salmonella* cells were inactivated, thus a value of 0 was returned. When undercooking occurred *Salmonella* inactivation was estimated using the decimal reduction time (D-value) at cooking temperature, DT_{cook} **(Equation 4.11)**;

$$D_{Tcook} = D_{60} \times 10^{\frac{60-T_{cook}}{z}} \quad (4.11)$$

where D_{60} = 8.5 min, the D-value at 60 °C; and z = 7.6 °C, temperature required for a log reduction in D-value (Murphy et al., 2002).

4.3.5.7 Dose-response modeling and risk characterization

The serving size used in this study was 85 g of cooked chicken meat, based on the reference amounts customarily consumed per eating occasion (9 CFR. 381.412). This serving size was multiplied by the *Salmonella* level before consumption to obtain the dose of *Salmonella* ingested per serving of cooked chicken meat. The loss-adjusted per capita availability of poultry in the U.S. is 54.6 lb/year (24,766.14 g/year) (USDA ERS, 2021). Using the U.S. population of 335,888,625 at the end of 2023 (US Census Bureau, 2024), the annual number of servings in the U.S. was

determined to be 97.87 billion servings. The probability of illness was estimated using the beta-Poisson model **(Equation 4.12)**. The beta-Poisson model is based on a Poisson distribution of the number of pathogen cells in each dose. It assumes each pathogen has a different probability of surviving, reaching, and infecting an appropriate site in a host and uses a beta distribution to describe the non-constant probabilities (Haas et al., 2014).

$$P = 1 - \left(1 + \frac{D}{\beta}\right)^{-\alpha} \tag{4.12}$$

P is the probability of illness; D is the dose ingested in CFU and α and β are model parameters that define the shape of the dose-response curve. The baseline QMRA model assumes all serovars have similar virulence and uses α (0.1324) and β (51.45) parameters from the WHO/FAO (2002) outbreak dose-response model. The probability of illness per serving was then calculated as the product of the *Salmonella* prevalence in retail chicken and the probability of illness.

4.3.6 Scenario models

Various scenario models were constructed using combinations of low/high-virulence serovars and homogenous/heterogenous growth rates. In **Table 4.2**, 'homogenous' indicates that it was assumed that *Salmonella* isolates have the same growth rate, thus one *b* value was used when estimating the growth rate of *Salmonella.* 'Finite mixture' indicates that it was assumed that *Salmonella* had distinct subpopulations growing at different rates, so different *b* values were used based on results from the finite mixture modeling analysis **(see Section 4.4.1)**. 'High-virulence' here was defined as the USDA FSIS's Food Safety Key Performance Indicator (KPI) serovars (Kim et al., 2024). The USDA FSIS has designated *Salmonella* serovars Infantis, Enteritidis, and

Typhimurium as KPI serovars for its Poultry Safety Program for the 2022 to 2026 fiscal years (USDA FSIS, 2022). These serovars were selected by the USDA FSIS because of their frequent association with human illnesses. 'Low-virulence' refers to all serovars other than the KPI serovars. Using USDA FSIS quarterly sampling reports on *Salmonella* from January 2019 (FY 2019; Q2) to December 2023 (FY 2024; Q1), we found that the USDA FSIS isolated 11,234 *Salmonella* serovars from domestic chicken samples (USDA FSIS, 2024a). Of this number, 6,422 (57.17%) were KPI serovars. Based on this, scenarios 2 and 3, used a ratio of 57% high-virulence serovars to 43% low-virulence serovars for modeling. Scenarios 4 and 5 modeled a situation where the fraction of high-virulence serovars has been reduced to 50%.

The 'high-virulence' scenarios used α (0.157) and β (9.17) parameters for highly virulent *Salmonella* strains from Latimer et al. (2001). 'Low-virulence' used model parameters α = 0.318 and N_{50} = 37,100, which were obtained from **(**QMRA WIKI, 2017**)** and based on *Salmonella* Anatum I and II human challenge studies by Mccullough and Eisele (1951). Parameter N_{50}, the median infectious dose, was reparameterized to obtain β = 4729.9 using **Equation (4.13)**.

$$\beta = \frac{N_{50}}{\left(2^{\frac{1}{\alpha}} - 1\right)} \tag{4.13}$$

A comparison of the three beta-Poisson dose-response models shows that the WHO/FAO (2002) α and β parameters result in a probability of illness that is higher than that of Mccullough and Eisele (1951) but lower than that of Latimer et al. (2001). Thus, the use of 'moderate' in Table 4.2 to describe virulence in the baseline and scenario 1 models.

Table 4.2. Conditions for the baseline and scenario QMRA models

Model	Virulence	Growth behavior
Baseline	Moderate	Homogenous
Scenario 1	Moderate	Finite mixture
Scenario 2[a]	High + Low-virulence	Homogenous
Scenario 3[a]	High + Low-virulence	Finite mixture
Scenario 4[b]	High + Low-virulence	Homogenous
Scenario 5[b]	High + Low-virulence	Finite mixture
Scenario 6	Low-virulence	Homogenous
Scenario 7	Low-virulence	Finite mixture

[a] Uses a proportion of 57%:43% for high- and low-virulence serovars.
[b] Uses a proportion of 50%:50% for high- and low-virulence serovars.

4.3.7 Model simulation and sensitivity analysis

The risk models were developed in the @RISK software (v. 8.7.0 Lumivero, Denver, CO, USA), an Excel add-in. The models were run at 100,000 iterations in a Monte Carlo simulation. The Latin Hypercube Sampling method was chosen to sample input distributions. This approach stratified input distributions and randomly selected a value in each stratum for every iteration. The Mersenne Twister algorithm was chosen for random number generation and all simulations used the same seed. Sensitivity analysis was performed by using Spearman's rank correlation coefficient to determine the impact of the variables on the probability of illness per serving of chicken meat and the number of illnesses in the U.S.

4.4 Results and Discussion

4.4.1 Finite mixture modeling of *Salmonella* growth data

Data on all 68 *Salmonella* isolates obtained from ComBase fitted the Baranyi-Roberts primary growth model satisfactorily with R-square values ranging from 0.74 to 0.99. The growth

rates obtained ranged from 0.0164 to 1.803 log CFU/g/h (mean = 0.6506 log CFU/g/h). A visualization of the data using a histogram and density plot showed multiple peaks, indicating a multimodal distribution and the possibility of distinct components or subpopulations of *Salmonella* isolates (**Figure 4.1**).

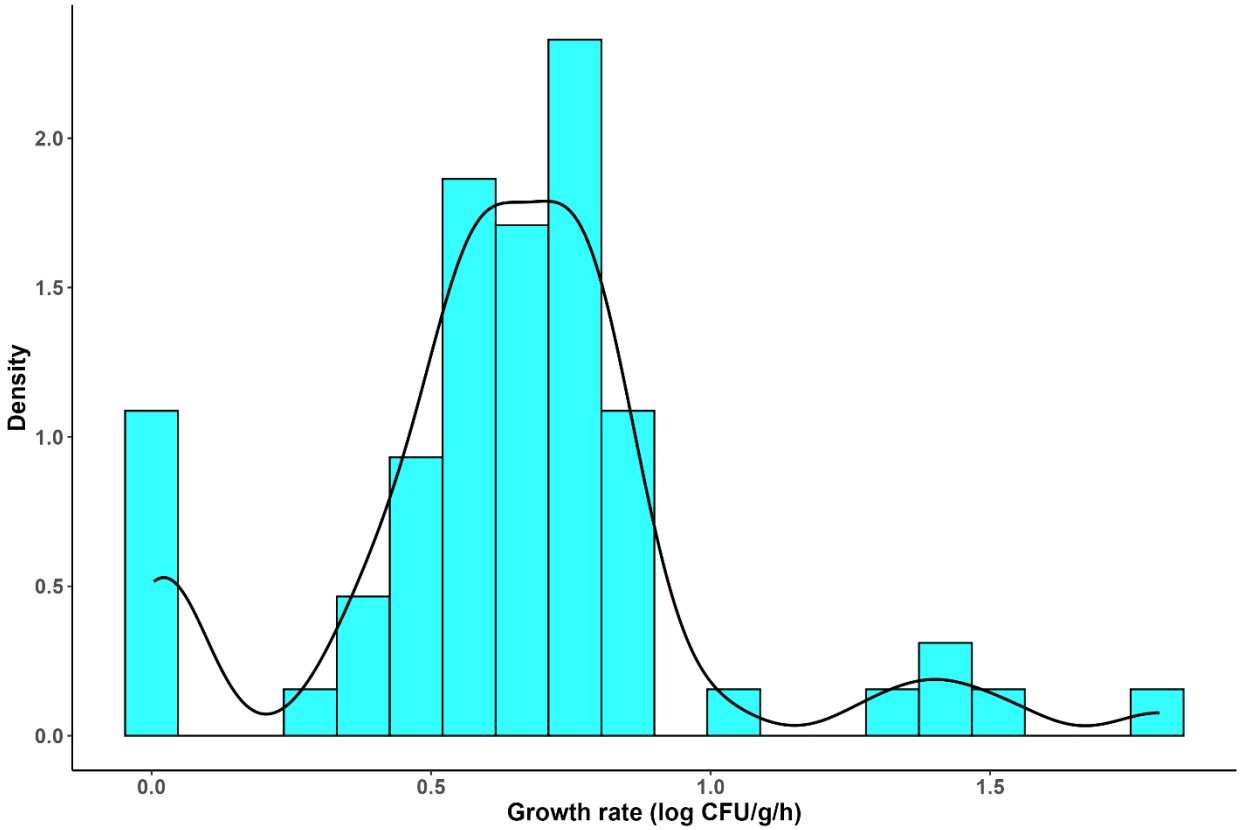


Figure 4.1. Histogram and density plot of the growth rates of *Salmonella* isolates.

Therefore, we performed finite mixture modeling to determine the number of components in the data. The results have been presented in **Table 4.3**. The simplest model, A (k = 1), had a single component, i.e., assumed there were no subpopulations and had a mean growth rate of 0.6411 log CFU/g/h and BIC of 48.4964. Model B (k = 2) introduced a second component, suggesting that there were two subpopulations with mean growth rates of 0.6411 and 0.6511 log

CFU/g/h. However, the similar growth rates produced by model B as well as the higher BIC (56.9280) indicated that even the one-component system (model A) could describe the *Salmonella* population better. Model C (k = 3) had three components indicating significant heterogeneity in the *Salmonella* population. The BIC reduced to 27.6337 indicating that a three-component model describes *Salmonella* growth dynamics better than the previous models. Models D (k = 4) and E (k = 5) had lower variance than Model C, but higher BICs. Models C, D, and E had similar variances (< 0.025), which were lower than the variances observed for models A and B.

Table 4.3. Mixing proportions, mean, variance, and BIC values of various components identified from computer-assisted mixture analysis (finite mixture modeling).

Model	Components (k)	Mixing proportions (π)	Mean (μ; log CFU/g/h)	Variance	BIC
A	1	1	0. 6506	0.1139	48.4964
B	1	0.0502	0.6411		
	2	0.9498	0.6511	0.1123	56.9280
C	1	0.1088	0.0378		
	2	0.8132	0.6563		
	3	0.0780	1.5963	0.0203	27.6337
D	1	0.1079	0.0370		
	2	0.8180	0.6561		
	3	0.0589	1.4069		
	4	0.0152	1.7788	0.0186	34.2880
E	1	0.1031	0.0222		
	2	0.3941	0.5340		
	3	0.4293	0.7652		
	4	0.0587	1.4073		
	5	0.0148	1.8015	0.0073	34.1906

ANOVA results showed that though model A had a better fit than model B (simulated p-value > 0.05), it was not better than models C, D, and E (simulated p-value < 0.05). A simulated p-value below 0.05 means that the addition of more components significantly improves a model fit, but a value above 0.05 means that there is not enough evidence to justify the addition of more

components to a model. A comparison of models C, D, and E showed that model E could be a better fit for the *Salmonella* growth data, however, the addition of more components in model E resulted in a more complex model and higher penalty, hence the higher BIC than model C.

Based on the variance, complexity, and fit, model C was chosen as the model that most adequately describes *Salmonella* growth. Moreover, it would be more interpretable. Model C had three components with most of the isolates clustered in the middle (0. 8132) of the distribution, and fewer isolates in the left (0.1088) and right tails (0.0780). This was interpreted as three different groups of isolates in the dataset: a low-growth subpopulation which is about 11% of the population and growing at a rate of 0.0378 log CFU/g/h, a medium-growth subpopulation which is 81% of the population and growing at a rate of 0.6563 log CFU/g/h and a high-growth subpopulation which is about 8% of the population and growing at a rate of 1.5963 log CFU/g/h **(Figure 4.2**). Using these growth rates, growth model parameter, *b*, was calculated using Ratkowsky's model (**Equation 4.10)**. For the homogenous population (single component), *b* was 0.0332. This result is consistent with a value of 0.0312 which was estimated by Golden and Mishra (2021). For the three-component heterogenous population, *b* values were 0.008, 0.0333, and 0.052 for the low, medium, and high-growth subpopulations, respectively. These values were used in the QMRA models when estimating growth rates in scenarios 1, 3, 5, and 7.

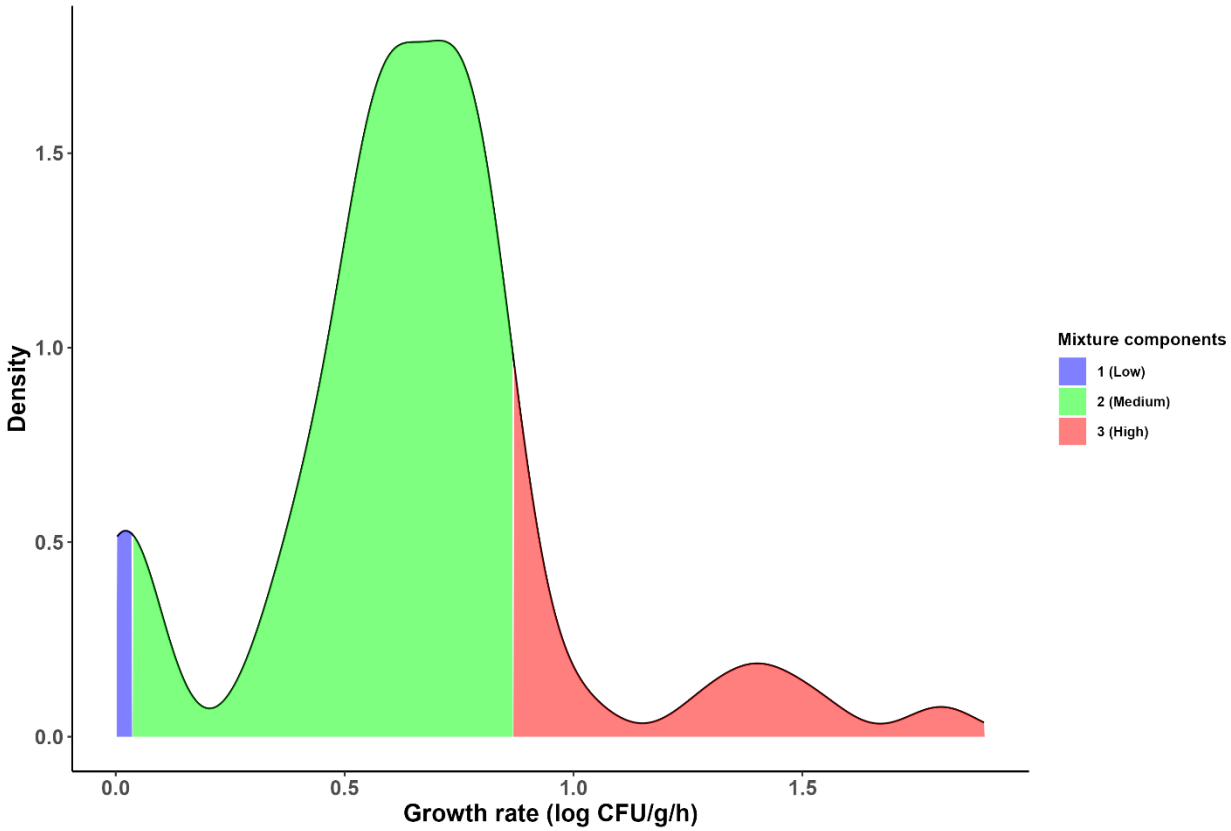


Figure 4.2. Density plot showing the mixture components of Model C.

A comparison of *Salmonella* growth rates in the baseline model (assumes homogenous growth behavior) and scenario model 1 (assumes heterogenous growth behavior) revealed higher growth rates in scenario 1. For instance, during retail storage, transportation to home, and home (refrigerator) storage in the baseline model, growth rates of 0 – 0.21 log CFU/g/h, 0 – 0.11 log CFU/g/h, and 0 – 0.11 log CFU/g/h, respectively were estimated. For the same stages in scenario 1, growth rates of 0 – 0.50 log CFU/g/h, 0 – 0.27 log CFU/g/h, and 0 – 0.21 log CFU/g/h were estimated. The difference between growth rates was not statistically significant ($p > 0.05$) during retail storage, but significant ($p < 0.05$) during transportation to home and home storage. The difference in the estimated growth rates between the baseline and scenario model 1 was even more pronounced at relatively higher temperatures (14 – 37 °C) such as during countertop thawing

(**Figure 4.3**) or running water thawing. Heterogeneity within a microbial population can be vital for risk assessment models, especially when the heterogeneity relates to their growth rate as it could indicate differing responses to stresses encountered in foods and food processing environments. Furthermore, microbial growth rates determine the population density of pathogens in foods and ultimately, the number of cells (dose) that are ingested, thus accurately modeling these differences could provide more accurate risk estimates.

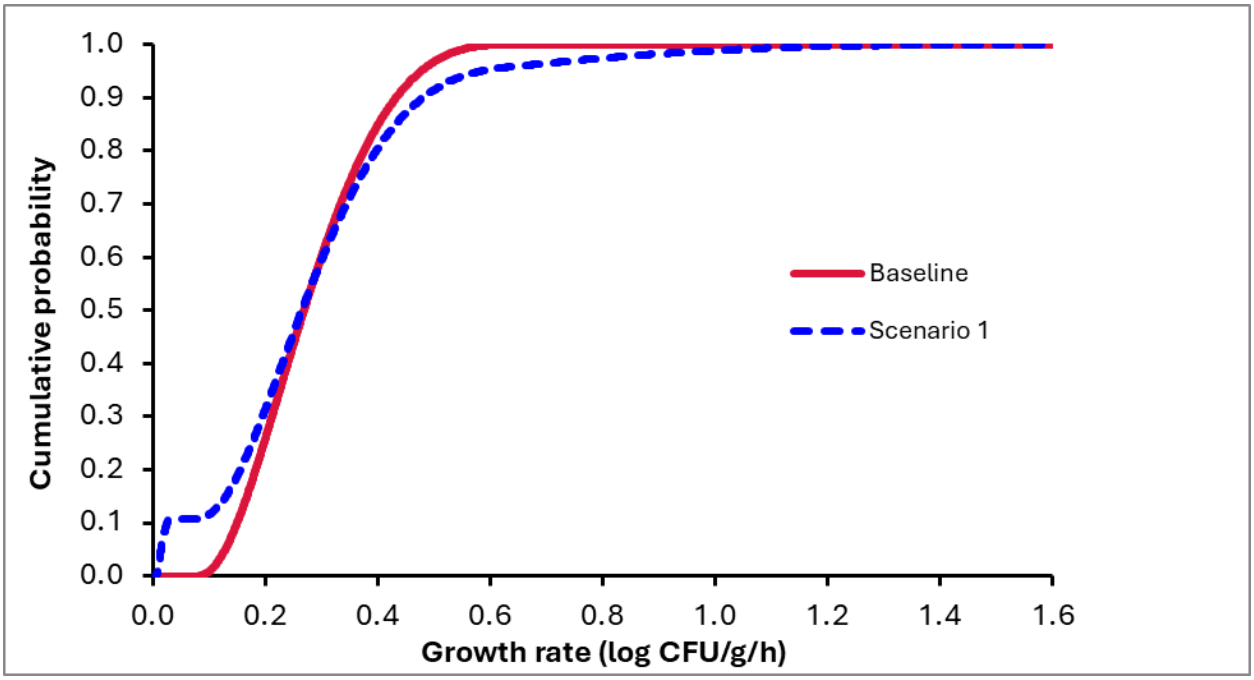


Figure 4.3. *Salmonella* growth rate during countertop thawing.

4.4.2 *Salmonella* levels across the chicken farm-to-fork chain

The density distribution of *Salmonella* concentration in chicken from the farm to the cooking stage has been presented using a ridgeline plot in **Figure 4.4.** *Salmonella* concentration was highest at the farm stage, with a mean of 3.54 log CFU on the exterior of birds. The distribution shifts to the left during the processing stage, signifying a reduction in *Salmonella* concentration.

The mean concentration at the end of processing was -4.67 log CFU/g and increased slightly to -4.63 log by the end of retail storage. Even when low or no growth is observed during retail storage, the retail stage still facilitates *Salmonella* growth by shortening the lag time, thereby enabling higher growth in subsequent stages of the farm-to-fork chain (Stathas et al., 2024). Further increases in *Salmonella* concentration occurred during transportation to home and home storage before being reduced during cooking to a mean of -6.79 log CFU/g.

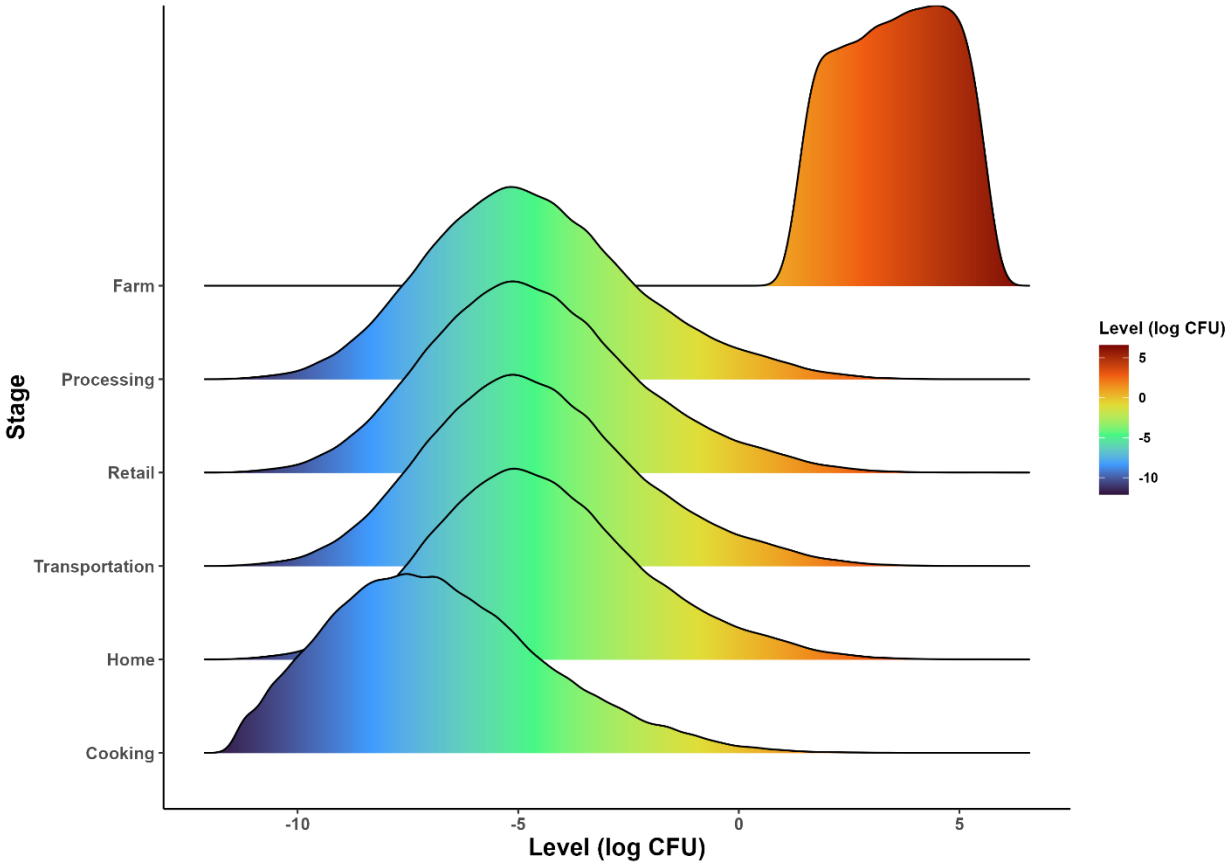


Figure 4.4. Simulated *Salmonella* levels at the end of each stage in the baseline QMRA model. Levels are in log CFU/bird for the farm stage and log CFU/g for other stages.

The estimated *Salmonella* levels from this study are consistent with results from Kim et al. (2024) and Zhu et al. (2017). Kim et al. (2024) determined the mean *Salmonella* concentration in

raw chicken parts, using two USDA FSIS processing establishment sampling datasets, to be -3.19 and -5.53 log CFU/g. Our estimated *Salmonella* concentration at the processing stage falls within this range. Additionally, Kim et al. (2024) estimated that 8.22% – 20.6%, 0.65% – 1.18%, and 0.06% – 0.38% of raw chicken part samples were contaminated with > 0.0074 CFU/g, > 1 CFU/g, and > 10 CFU/g, respectively. Our simulated results at these limits were 14.26%, 3.67%, and 1.5%. The estimated *Salmonella* concentration at the end of cooking was 1 and 2 logs lower than the concentrations estimated by Golden and Mishra (2021) (-5.78) and Stathas et al. (2024) (-4.78), respectively. Despite the relatively low mean *Salmonella* concentrations, **Figure 4.4** shows a small right tail in the distribution for the cooking stage, representing a few but highly contaminated chicken samples.

4.4.3 Risk estimates and sensitivity analysis

The current QMRA model follows the typical chicken production and supply chain in the U.S. The median probability of illness per serving was zero for the baseline model and all scenario models. In some scenarios, the probability of illness per serving was zero even at the 60th percentile. Technically, the risk cannot be considered zero; therefore, these estimates suggest that up to 60% of chicken servings consumed in the U.S. carry an extremely small residual risk of salmonellosis. The mean probability of illness per serving for the baseline model (moderately virulent isolates growing at the same rate) was estimated to be 6.61×10^{-5} (about 7 illnesses per 100,000 servings of chicken meat) **(Table 4.4)**. For the scenario models, the highest mean probability of illness per serving was observed in scenario 3 (about 1 illness per 100,000 servings of chicken meat) which models the *Salmonella* growth rate based on the finite mixture subpopulations as well as assumes that 57% of *Salmonella* serovars are high-virulent and 43% are

low-virulent. Scenario 6 which models *Salmonella* growth based on a homogenous population and assumes all *Salmonella* serovars are low-virulent recorded the lowest mean probability of illness per serving of 1.43×10^{-5} (less than 2 illnesses per 100,000 servings of chicken meat). When all the mean estimates for all models were compared, the order for the probability of illness per serving was Scenario 3 > Scenario 5 > Scenario 2 > Scenario 4 > Scenario 1 > Baseline > Scenario 7 > Scenario 6. All the estimates found for the baseline and scenario models in this study fit within the ranges of previous studies.

Using the baseline model, this study estimated a lower risk of salmonellosis as compared to Golden and Mishra (2021), Smadi and Sargeant (2013), and Zhu et al. (2017). Golden and Mishra (2021) estimated a mean probability of illness per serving of 1.80×10^{-4} for conventionally produced chicken and 2.1×10^{-4} for alternatively produced chicken. Smadi and Sargeant (2013), Stathas et al. (2024), and Zhu et al. (2017) estimated a probability of illness per serving of 1.99×10^{-4}, $1.19 \times 10^{-4,}$ and 1.20×10^{-4}, respectively. However, other studies by WHO/FAO (2002), and Xiao et al. (2021) estimated lower probabilities of illness per serving of 1.13×10^{-5} and 1.18×10^{-8}, respectively. Beyond the baseline model, there are no related *Salmonella*-chicken QMRA studies, to the best of our knowledge, which have explored population heterogeneity using both different dose-response models and growth rates. Thus, we are unable to adequately contextualize the estimates from the scenario models within those of other studies.

Although Godínez-Oviedo et al. (2022) also modeled QMRA for *Salmonella* in chicken using different dose-response models and reported a lower probability of illness per serving, the differences in the two studies make them not directly comparable. While we defined virulence using the USDA KPI serovars, Godínez-Oviedo et al. (2022) defined virulence based on multidrug

resistance and the presence of virulence genes. Furthermore, there are differences in the approaches to dose-response modeling. While the current study used parameters from Mccullough and Eisele (1951) for low-virulence serovars, Godínez-Oviedo et al. (2022) use parameters from Latimer et al. (2001) that result in an essentially flat dose-response curve until doses above 5 log CFU.

Table 4.4. Probability of illness per serving and number of illnesses per year due to the consumption of *Salmonella*-contaminated chicken

Model	Probability of illness per serving			Number of illnesses per year		
	Mean	75th percentile	95th percentile	Mean	75th percentile	95th percentile
Baseline	6.61×10^{-5}	1.52×10^{-10}	1.50×10^{-6}	6.47×10^{6}	1.50×10^{1}	1.47×10^{5}
Scenario 1	7.88×10^{-5}	1.50×10^{-10}	1.62×10^{-6}	7.71×10^{6}	1.50×10^{1}	1.58×10^{5}
Scenario 2	9.98×10^{-5}	6.22×10^{-10}	1.77×10^{-6}	9.77×10^{6}	6.10×10^{1}	1.73×10^{5}
Scenario 3	1.08×10^{-4}	6.92×10^{-10}	1.88×10^{-6}	1.06×10^{7}	6.80×10^{1}	1.84×10^{5}
Scenario 4	8.94×10^{-5}	4.54×10^{-10}	1.30×10^{-6}	8.75×10^{6}	4.40×10^{1}	1.27×10^{5}
Scenario 5	1.03×10^{-4}	4.28×10^{-10}	1.36×10^{-6}	1.00×10^{7}	4.20×10^{1}	1.34×10^{5}
Scenario 6	1.43×10^{-5}	4.01×10^{-12}	4.01×10^{-8}	1.40×10^{6}	1.00×10^{0}	3.93×10^{3}
Scenario 7	1.87×10^{-5}	4.10×10^{-12}	4.30×10^{-8}	1.83×10^{6}	1.00×10^{0}	4.20×10^{3}

The predicted number of illnesses per year follows the same trend as the probability of illness per serving since it is simply a product of the probability of illness per serving and the number of servings per year. Zero illnesses were estimated at the median, and 60th percentile for the baseline and all scenario models. The baseline model predicted a mean of 6.47 million illnesses per year in the U.S. **(Table 4.4)**. Scenarios 3 and 6 predicted the highest and lowest mean number of illnesses per year at 10.6 million and 1.4 million, respectively. Presently, the USDA estimates that the consumption of *Salmonella*-contaminated chicken results in 125,115 illnesses each year (USDA 2024b). When adjusted using the current IFSAC estimate of *Salmonella* illnesses attributed to chicken (18.6%) (IFSAC, 2023), the number of chicken-associated *Salmonella* illnesses would be 134,517 per year. Therefore, our QMRA models estimated a much higher mean number of

illnesses than the real-world surveillance data. Golden and Mishra (2021) estimated the number of salmonellosis cases to be 3.88 million. Smadi and Sargeant (2013), and Zhu et al. (2017) also reported such high estimates in their studies Combining various factors such as undercooking, improper utensils, and handwashing into one model could lead to risk overestimation Smadi and Sargeant (2013), since all these factors may not necessarily occur together.

Considering that the median and even 60th percentile number of illnesses per year were zero, but the mean was in the millions, it is apparent that the distribution of the number of illnesses is highly right-skewed and has been disproportionately affected by a small number of extremely high values. These values arose from the right tail of the *Salmonella* concentration at the cooking stage distribution **(Figure 4.4)** and when integrated into the dose-response model and multiplied by the ~98 billion servings resulted in very high numbers of illnesses. Furthermore, the 95th percentiles **(Table 4.4)**, indicate that most data points are well below the means. Since the 95th percentiles exclude the top 5% values, they were taken as realistic estimates of the number of illnesses per year. The 95th percentile estimates in the baseline and scenarios 1 – 5 ranged from 127,000 to 184,000 and are 5.6% lower to 36.8% higher than the estimated 134,517 chicken-associated *Salmonella* illnesses that occur yearly. The estimates at scenarios 6 and 7 (3,930 – 4,200) are expectedly much lower as they simulate a hypothetical scenario where all high-virulence *Salmonella* serovars have been eliminated from chicken flocks.

On average, models assuming heterogeneous growth (scenarios 1, 3, 5, and 7) estimated a 6.6% higher number of illnesses per year compared to models assuming homogeneous growth (baseline, scenarios 2, 4, and 6). Scenarios 2 and 3 model the current situation where 57% of *Salmonella* serovars isolated from domestic chicken samples are USDA KPI serovars (high-

virulence) and estimated 173,000 to 184,000 annual illnesses. If the proportion of KPI serovars were reduced to 50% (scenarios 4 and 5), it would result in a 27% reduction in the number of illnesses. A further reduction in the proportion of KPI serovars to zero (scenarios 6 and 7) would result in a 97.7% reduction in the number of illnesses. Assuming all serovars were moderately virulent (baseline and scenario 1), the number of illnesses per year would be 147,000 – 158,000, a 14.6% decrease from the estimates in scenarios 2 and 3. These findings clearly show that population heterogeneity affects the risk of salmonellosis. As we have shown here, even minute differences in pathogens can lead to considerable variations in risk estimates. These phenotypic differences are regulated at the gene level, hence the many calls for the inclusion of genomic data into QMRA using biomarkers, serovar-specific dose-response models, stress response, and growth profiles, among other pathogen properties (Chen et al., 2020; Guillier et al., 2022; Haddad et al., 2018; EFSA Panel on Biological Hazards, 2019; NACMCF, 2024, Rantsiou et al., 2018). In this study, we incorporated pathogen heterogeneity into QMRA by focusing on the exposure assessment (adjusting growth rates based on distinct subpopulations) and the dose-response modeling steps of the QMRA framework. Although heterogeneity in growth rate has a lower impact on chicken-associated salmonellosis than heterogeneity in virulence, it is important that both factors are modeled to achieve relatively more accurate risk estimates for effective risk management.

In a bid to reduce the number of chicken-associated *Salmonella* illnesses, the USDA FSIS has targeted a 2% annual reduction in the proportion of chicken samples contaminated with the KPI *Salmonella* serovars (USDA FSIS, 2022). Also, it has recently been declared that not ready-to-eat breaded stuffed chicken products containing more than 1 CFU/g of *Salmonella* are adulterated (USDA FSIS, 2024c). The USDA FSIS is proposing a related regulation to classify

certain raw poultry products as adulterated if they contain specific levels and/or serovars of *Salmonella* (USDA FSIS, 2024b). Further regulations that would require processing facilities to designate *Salmonella* as a hazard that is reasonably likely to occur during the receiving of flocks, hence mandating flock testing have also been proposed (USDA FSIS, 2024b) The implementation of these regulations could drive ongoing industry efforts, such as vaccination, omics surveillance, logistic slaughter, and pre/probiotic administration to reduce *Salmonella* levels in both chicken flocks and products.

The results of a sensitivity analysis to determine the effects of the model inputs on the number of illnesses (in scenario 3) have been presented in **Figure 4.5.** The top four inputs were all related to consumer food preparation, i.e., the proportion of consumers who wash their hands, the proportion of consumers who wash utensils properly or use separate ones, cooking temperature, and undercooking. The effect of the processing (evisceration) and farm (proportion of high virulence isolates and *Salmonella* concentration in feces) stages were relatively lower. If 90% of consumers cooked chicken to a minimum internal temperature of 73.89 °C (165 °F) and avoided cross-contamination during food preparation, the estimated number of illnesses would reduce by 97%, which is almost as much as eliminating all high-virulence *Salmonella* serovars from chicken flocks. These results highlight the importance of food preparation practices. Studies have shown that consumers often overestimate or are overconfident about their food safety practices, while their actual behaviors fall short (Evans et al., 2020; Mazengia et al., 2015; Sutherland et al., 2020), underscoring the need for continued consumer education on safe food preparation. It could be argued that if *Salmonella* levels were zero or very low in chicken flocks, to begin with, the risk of illness would be negligible. While this is true, the significance of proper food preparation cannot

be overstated. Reducing the number of chicken-associated *Salmonella* illnesses requires efforts from regulators, industry, and consumers.

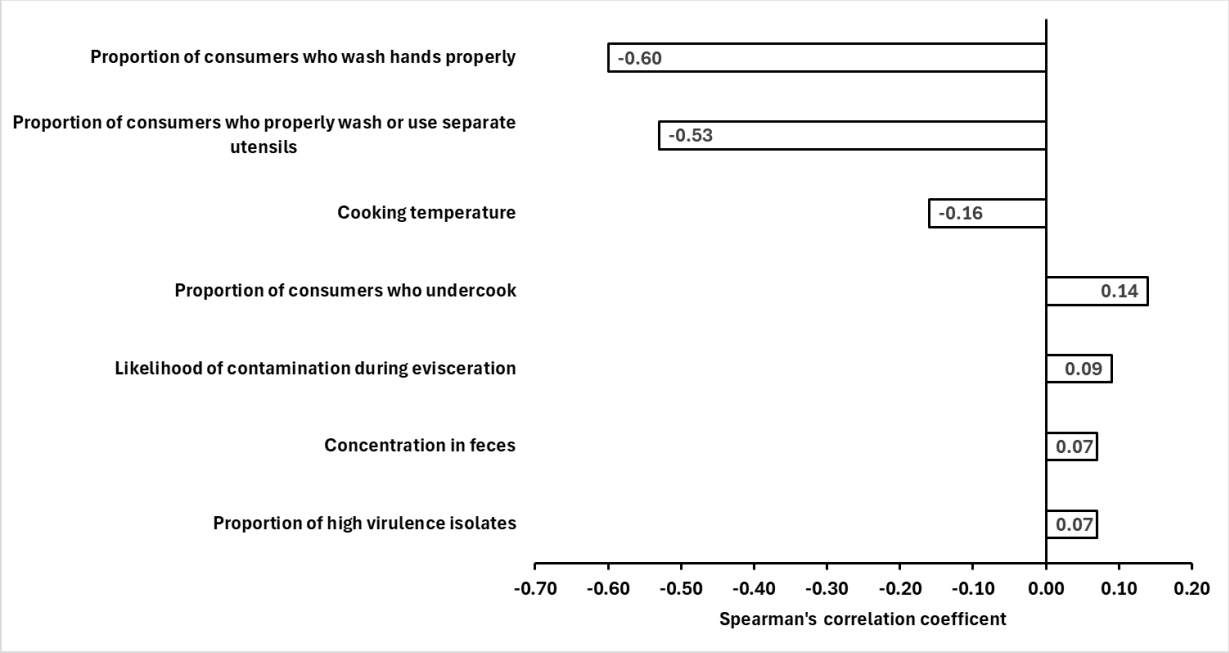


Figure 4.5. Tornado plot showing the model input variables with the largest impact on the number of illnesses per year.

4.5 Conclusions

This study provides a farm-to-fork QMRA model that incorporates *Salmonella* heterogeneity by using different growth rates and dose-response model parameters. Our findings indicate that the presence of distinct subpopulations with varying growth rates substantially impacts the risk of salmonellosis; however, this impact is less significant than that of differences in *Salmonella* serovar virulence. Poor food safety practices in the home such as undercooking, and cross-contamination after handling raw chicken were associated with an increased risk of illness. Improving these practices was almost as effective in reducing the number of illnesses as

eliminating all high-virulence *Salmonella* serovars. Consumers as well as regulators and industry are all important stakeholders in ensuring food safety.

Despite providing valuable insights into the effects of population heterogeneity, as with any QMRA model, our study has some limitations. The initial dataset for identifying distinctly growing subpopulations was small; therefore, the proportions of the different subpopulations could vary in a larger dataset, potentially influencing risk estimates. Also, while we focused on differences in growth capacity and virulence, other factors such as antimicrobial resistance and stress response exist but were not modeled in our study. Additionally, although we employed different dose-response models based on *Salmonella* virulence, they may not necessarily capture the full virulence potential of every single *Salmonella* serovar. Nonetheless, we believe that this study offers accurate estimates of the risk of chicken-associated salmonellosis when incorporating pathogen differences. Our study significantly complements ongoing efforts to fine-tune risk models by incorporating pathogen heterogeneity.

Chapter 5: Summary and future studies

Whole genome sequencing already has numerous benefits for food safety and risk assessment and even holds more promise when combined with machine learning for food safety applications. Despite the focus on *Salmonella* and chicken, the methods here can be extended to other pathogens and foods. It is even likely that the integration of genomics into QMRA would have a bigger impact on fresh produce safety than on poultry and meats since fresh produce is typically consumed raw.

In Chapter 2, the potential of machine learning to identify genes associated with *Salmonella* stress response during poultry processing was explored. Six machine learning algorithms were trained on *Salmonella* WGS data, and model fit was assessed using metrics such as the AUROC and confusion matrix statistics. All models achieved high performances with logit boost showing the best performance with an AUROC of 0.904, sensitivity of 0.889, and specificity of 0.920. The significant genes identified included *ybtX*, which encodes a Yersiniabactin-associated zinc transporter, and the transferase-encoding genes *yccK* and *thiS*. Additionally, genes coding for cold (*cspA*, *cspD*, and *cspE*) and heat shock (*rpoH* and *rpoE*) responses were identified. Other significant genes included those involved in lipopolysaccharide biosynthesis (*irp1*, *waaD, rfc,* and *rfbX*), DNA repair and replication (*traI*), biofilm formation (*ccdA* and *fyuA*), and cellular metabolism (*irtA*). A major highlight here is the identification of genes such as *ccdA*, *fyuA*, *ybtX*, and others which would typically not be classified as stress response genes.

Chapter 3 focused on the development of a machine learning model to predict *Salmonella* resistance to amoxicillin-clavulanic acid, ampicillin, ceftiofur, ceftriaxone, sulfisoxazole, streptomycin, tetracycline, and cefoxitin. The machine learning models showed high sensitivity

(≥0.833), specificity (≥0.833), and balanced accuracy (≥0.866), across all the antimicrobials. The models predicted resistance prevalences ranging from 1% (ceftriaxone) to 65.5% (streptomycin). When the AMR phenotype predictions of the machine learning models were compared to predictions from ResFinder, the predictions from this study were accurate (>95%).

Chapter 4 focused on the incorporation of *Salmonella* heterogeneity into a QMRA model by using different dose-response model parameters to account for differences in serovar virulence and finite mixture modeling to identify differences in growth rates. Models estimated the annual number of chicken-associated salmonellosis cases to range from 3,930 to 184,000 (95th percentile). The results showed that heterogeneity in *Salmonella* virulence had a much higher impact than heterogeneity in growth rates on the risk of salmonellosis. Sensitivity analysis showed that at-home food preparation practices had the highest effect on the risk of chicken-associated salmonellosis. A comparison of risk estimates showed that if 90% of consumers cooked chicken adequately and avoided cross-contamination, the reduction in illnesses was nearly as substantial as eliminating all high-virulence serovars.

WGS data, which provides information on gene presence or absence, was used in this research. Future studies may benefit from the use of transcriptomic and proteomic data to provide a more comprehensive understanding of gene expression profiles. This would be vital in linking phenotypes observed as a result of food processing conditions to the up or downregulation of specific genes. Genome-scale metabolic models (microbial networks that link genes, proteins, and biochemical reactions) may help in this regard; however, their application in food safety research is still in its early stages. Another challenge is the translation of omics data into quantitative forms

that can be used in a QMRA. Continued research into advanced data analytic methods applicable to genetic data is still needed to address this challenge.

In Chapters 2 and 3 of this dissertation, the models used all *Salmonella* serovars present in the datasets. Future studies could consider focusing only on serovars such as Enteritidis, Heidelberg, Infantis, Thompson, and Typhimurium, which have been linked to multiple outbreaks involving chicken. Additionally, though we have used different dose-response models for varying virulence levels in Chapter 4, future risk assessment models could consider the use of serovar-specific dose-response models where data is available. Furthermore, modeling the transmission dynamics of *Salmonella* will be important for understanding the genetic or environmental factors that influence changes in serovar prevalence across the poultry production and supply chain. Finally, machine learning is data-intensive; models are as good as their underlying data. This is even more crucial as we aim to integrate genomic data into QMRA. Proper curation of genomic data, with comprehensive metadata, would greatly enhance the predictive power of models, leading to improved risk assessment and risk management actions.

References

Ades, S. E. (2008). Regulation by destruction: design of the σE envelope stress response. Current *Opinion in Microbiology, 11*(6), 535–540. https://doi.org/10.1016/j.mib.2008.10.004

Agyare, C., Etsiapa Boamah, V., Ngofi Zumbi, C., Boateng Osei, F. (2019). Antibiotic Use in Poultry Production and Its Effects on Bacterial Resistance. In Kumar, Y. (Ed.). *Antimicrobial Resistance - A Global Threat*. IntechOpen, London. https://doi.org/10.5772/intechopen.79371

Alcock, B. P., Huynh, W., Chalil, R., Smith, K. W., Raphenya, A. R., Wlodarski, M. A., Edalatmand, A., Petkau, A., Syed, S. A., Tsang, K. K., Baker, S. J. C., Dave, M., Mccarthy, M. C., Mukiri, K. M., Nasir, J. A., Golbon, B., Imtiaz, H., Jiang, X., Kaur, K., … Mcarthur, A. G. (2023). CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Research*, *51*(1 D), D690–D699. https://doi.org/10.1093/nar/gkac920

Allard, M. W., Bell, R., Ferreira, C. M., Gonzalez-Escalona, N., Hoffmann, M., Muruvanda, T., Ottesen, A., Ramachandran, P., Reed, E., Sharma, S., Stevens, E.,Timme, R., Zheng, J., & Brown, E. W. (2018). Genomics of foodborne pathogens for microbial food safety. *Current Opinion in Biotechnology*, *49*, 224-229. https://doi.org/10.1016/j.copbio.2017.11.002

Allard, M. W., Strain, E., Melka, D., Bunning, K., Musser, S. M., Brown, E. W., & Timme, R. (2016). Practical value of food pathogen traceability through building a whole-genome sequencing network and database. In *Journal of Clinical Microbiology* (Vol. 54, Issue 8, pp. 1975–1983). American Society for Microbiology. https://doi.org/10.1128/JCM.00081-16

Allen, V. M., Tinker, D. B., Hinton, M. H., & Wathes, C. M. (2003). Dispersal of micro-organisms in commercial defeathering systems. *British Poultry Science*, *44*(1), 53–59. https://doi.org/10.1080/0007166031000085436

Álvarez-Ordóñez, A., Fernández, A., Bernardo, A., & López, M. (2010). Arginine and lysine decarboxylases and the Acid Tolerance Response of *Salmonella* Typhimurium. *International Journal of Food Microbiology*, *136*(3), 278–282. https://doi.org/10.1016/j.ijfoodmicro.2009.09.024

Álvarez-Ordóñez, A., Fernández, A., López, M., Arenas, R., & Bernardo, A. (2008). Modifications in membrane fatty acid composition of *Salmonella* Typhimurium in response to growth conditions and their effect on heat resistance. *International Journal of Food Microbiology*, *123*(3), 212–219. https://doi.org/10.1016/j.ijfoodmicro.2008.01.015

Alzahrani, K. O., AL-Reshoodi, F. M., Alshdokhi, E. A., Alhamed, A. S., Al Hadlaq, M. A., Mujallad, M. I., Mukhtar, L. E., Alsufyani, A. T., Alajlan, A. A., Al Rashidy, M. S., Al Dawsari, M. J., Al-Akeel, S. I., AL-Harthi, M. H., Al Manee, A. M., Alghoribi, M. F., & Alajel, S. M. (2023). Antimicrobial resistance and genomic characterization of *Salmonella enterica* isolates from chicken meat. *Frontiers in Microbiology*, *14*. https://doi.org/10.3389/fmicb.2023.1104164

Andino, A., & Hanning, I. (2015). *Salmonella enterica*: Survival, colonization, and virulence differences among serovars. *Scientific World Journal*, *2015*. https://doi.org/10.1155/2015/520179

Aspridou, Z., Balomenos, A., Tsakanikas, P., Manolakos, E., & Koutsoumanis, K. (2019). Heterogeneity of single cell inactivation: Assessment of the individual cell time to death and implications in population behavior. *Food Microbiology*, *80*, 85–92. https://doi.org/10.1016/j.fm.2018.12.011

Aviles, B., Klotz, C., Eifert, J., Williams, R., & Ponder, M. (2013). Biofilms promote survival and virulence of *Salmonella enterica* sv. Tennessee during prolonged dry storage and after passage through an in vitro digestion system. *International Journal of Food Microbiology*, *162*(3), 252–259. https://doi.org/10.1016/j.ijfoodmicro.2013.01.026

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, *19*(5), 455–477. https://doi.org/10.1089/cmb.2012.0021

Baranyi, J. (1997). Simple is good as long as it is enough. *Food Microbiology*, *14*(189–192).

Baranyi, J., & Roberts, T. A. (1994). A dynamic approach to predicting bacterial growth in food. *International Journal of Food Microbiology*, *23*, 277–294.

Barlow, S. M., Boobis, A. R., Bridges, J., Cockburn, A., Dekant, W., & Nauta, M. J. (2015). The role of hazard- and risk-based approaches in ensuring food safety. *Trends in Food Science & Technology*, *46*, 176–188. https://doi.org/10.1016/j.tifs.2015.10.007

Barria, C., Malecki, M., & Arraiano, C. M. (2013). Bacterial adaptation to cold. *Microbiology (United Kingdom)*, *159*(PART 12), 2437–2443. https://doi.org/10.1099/mic.0.052209-0

Beal, M. J., & Ghahramani, Z. (2000). The Variational Bayesian EM Algorithm for Incomplete Data: with Application to Scoring Graphical Model Structures. In J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, & M. West (Eds.), *Bayesian Statistics*. Oxford University Press.

Ben Hassena, A., Haendiges, J., Zormati, S., Guermazi, S., Gdoura, R., Gonzalez-Escalona, N., & Siala, M. (2021). Virulence and resistance genes profiles and clonal relationships of non-typhoidal food-borne *Salmonella* strains isolated in Tunisia by whole genome sequencing. *International Journal of Food Microbiology*, *337*. https://doi.org/10.1016/j.ijfoodmicro.2020.108941

Benefo, E. O., Karanth, S., & Pradhan, A. K. (2022). Applications of advanced data analytic techniques in food safety and risk assessment. In *Current Opinion in Food Science* (Vol. 48). Elsevier Ltd. https://doi.org/10.1016/j.cofs.2022.100937

Benefo, E. O., Karanth, S., & Pradhan, A. K. (2024). A machine learning approach to identifying *Salmonella* stress response genes in isolates from poultry processing. *Food Research International*, 113635. https://doi.org/10.1016/j.foodres.2023.113635

Berg, C., & Raj, M. (2015). A review of different stunning methods for Poultry—animal welfare aspects (stunning methods for poultry). *Animals*, *5*, 1207–1219. https://doi.org/10.3390/ani5040407

Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research,* 13, 281–305.

Bertani, B., & Ruiz, N. (2018). Function and biogenesis of lipopolysaccharides. *EcoSal Plus*, *8*(1). https://doi.org/10.1128/ecosalplus.esp-0001-2018

Bharat, A., Petkau, A., Avery, B. P., Chen, J., Folster, J., Carson, C. A., Kearney, A., Nadon, C., Mabon, P., Thiessen, J., Alexander, D. C., Allen, V., El Bailey, S., Bekal, S., German, G. J., Haldane, D., Hoang, L., Chui, L., Minion, J., … Mulvey, M. R. (2022). Correlation between Phenotypic and In Silico Detection of Antimicrobial Resistance in *Salmonella enterica* in Canada Using Staramr. *Microorganisms*, *10*(2). https://doi.org/10.3390/microorganisms10020292

Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, *25*(2), 197–227. https://doi.org/10.1007/s11749-016-0481-7

Bohning, D., Schlattmann, P., & Lindsay, B. (1992). Computer-Assisted Analysis of Mixtures (C.A.MAN): Statistical Algorithms. *Biometrics, 48*(1),283–303.

Booten, C., Robertson, J., Christensen, D., Heaney, M., Brown, D., Norton, P., & Smith, C. (2017). *Residential Indoor Temperature Study*. Technical Report NREL/TP-5500-68019. https://doi.org/10.2172/1351449

Bortolaia, V., Kaas, R. S., Ruppe, E., Roberts, M. C., Schwarz, S., Cattoir, V., Philippon, A., Allesoe, R. L., Rebelo, A. R., Florensa, A. F., Fagelhauer, L., Chakraborty, T., Neumann, B., Werner, G., Bender, J. K., Stingl, K., Nguyen, M., Coppens, J., Xavier, B. B., … Aarestrup, F. M. (2020). ResFinder 4.0 for predictions of phenotypes from genotypes. *Journal of Antimicrobial Chemotherapy*, *75*(12), 3491–3500. https://doi.org/10.1093/jac/dkaa345

Bovill, R., Bew, J., Cook, N., D'Agostino, M., Wilkinson, N., & Baranyi, J. (2000). Predictions of growth for *Listeria monocytogenes* and *Salmonella* during fluctuating temperature. *International Journal of Food Microbiology*, *59*, 157–165. https://doi.org/10.1016/s0168-1605(00)00292-0

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32.

Brown, B., Allard, M., Bazaco, M. C., Blankenship, J., & Minor, T. (2021). An economic evaluation of the Whole Genome Sequencing source tracking program in the U.S. *PLoS ONE*, *16*(10). https://doi.org/10.1371/journal.pone.0258262

Brown, E. W., Bell, R., Zhang, G., Timme, R., Zheng, J., Hammack, T. S., & Allard, M. W. (2021). *Salmonella* Genomics in Public Health and Food Safety. *EcoSal Plus*, *9*. https://doi.org/10.1128/ecosalplus.ESP-0008-2020

Brown, E., Dessai, U., McGarry, S., & Gerner-Smidt, P. (2019). Use of whole-genome sequencing for food safety and public health in the United States. *Foodborne Pathogens and Disease, 16*(7), 441-450. https://doi.org/10.1089/fpd.2019.2662

Bruhn, C. M. (2014). Chicken Preparation in the Home: An Observational Study. *Food Protection Trends*, *34*(5), 318–330.

Buchanan, R. L., Smith, J. L., & Long, W. (2000). Microbial risk assessment: Dose-response relations and risk characterization. *International Journal of Food Microbiology, 58*, 159–172. https://doi:10.1016/S0168-1605(00)00270-1.

Calderón, I. L., Morales, E. H., Collao, B., Calderón, P. F., Chahuán, C. A., Acuña, L. G., Gil, F., & Saavedra, C. P. (2014). Role of *Salmonella* Typhimurium small RNAs RyhB-1 and RyhB-2 in the oxidative stress response. *Research in Microbiology*, *165*(1), 30–40. https://doi.org/10.1016/j.resmic.2013.10.008

Campioni, F., Cao, G., Kastanis, G., Janies, D. A., Bergamini, A. M. M., Rodrigues, D. D. P., Stones, R., Brown, E., Allard, M. W., & Falcão, J. P. (2018). Changing of the genomic pattern of *Salmonella* Enteritidis strains isolated in Brazil over a 48 year-period revealed by whole genome SNP analyses. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-28844-6

Campioni, F., Gomes, C. N., Dália, &, Rodrigues, P., Alzira, Bergamini, M. M., & Falcão, J. P. (2021). Phenotypic analyses of *Salmonella enterica* serovar Enteritidis strains isolated in the pre- and post-epidemic period in Brazil. *Brazilian Journal of Microbiology*, *52*, 172–183. https://doi.org/https://doi.org/10.1007/s42770-020-00392-0

Cantalapiedra, C. P., Herņandez-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. (2021). eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Molecular Biology and Evolution*, *38*(12), 5825–5829. https://doi.org/10.1093/molbev/msab293

Capita, R. (2007). Variation in *Salmonella* Resistance to Poultry Chemical Decontaminants, Based on Serotype, Phage Type, and Antibiotic Resistance Patterns. *Journal of Food Protection*, *70*(8), 1835–1843.

Castro-Vargas, R. E., Herrera-Sánchez, M. P., Rodríguez-Hernández, R., & Rondón-Barragán, I. S. (2020). Antibiotic resistance in *Salmonella* spp. isolated from poultry: A global overview. *Veterinary World*, *13*(10), 2070–2084. https://doi.org/10.14202/vetworld.2020.2070-2084

Centers for Disease Control and Prevention (CDC). (2024a). BEAM (Bacteria, Enterics, Ameba, and Mycotics) Dashboard. Atlanta, Georgia: U.S. Department of Health and Human Services. Retrieved September 29, 2024, from https://www.cdc.gov/ncezid/dfwed/BEAM-dashboard.html.

Centers for Disease Control and Prevention (CDC). (2024b). *Salmonella.* Retrieved September 29, 2024, from https://www.cdc.gov/*Salmonella*/index.html

Chattopadhyay, M. (2006). Mechanism of bacterial adaptation to low temperature. *Journal of Biosciences*, *31*(1), 157-65.

Chaudhari, R., Singh, K., & Kodgire, P. (2023). Biochemical and molecular mechanisms of antibiotic resistance in *Salmonella* spp. *Research in Microbiology*,*174*. https://doi.org/10.1016/j.resmic.2022.103985

Chen, J., Karanth, S., & Pradhan, A. K. (2020). Quantitative microbial risk assessment for *Salmonella*: Inclusion of whole genome sequencing and genomic epidemiological studies, and advances in the bioinformatics pipeline. *Journal of Agriculture and Food Research, 2.* https://doi.org/10.1016/j.jafr.2020.100045

Cheng, R. A., Eade, C. R., & Wiedmann, M. (2019). Embracing diversity: Differences in virulence mechanisms, disease severity, and host adaptations contribute to the success of nontyphoidal *Salmonella* as a foodborne pathogen. *Frontiers in Microbiology*, *10*(JUN). https://doi.org/10.3389/fmicb.2019.01368

Clark, S. C., Egan, R., Frazier, P. I., & Wang, Z. (2013). ALE: A generic assembly likelihood evaluation framework for assessing the accuracy of genome and metagenome assemblies. *Bioinformatics*, *29*(4), 435–443. https://doi.org/10.1093/bioinformatics/bts723

Clouser, C. S., Doores, S., Mast, M. G., & Knabel, S. J. (1995). The Role of Defeathering in the Contamination of Turkey Skin by *Salmonella* species and Listeria monocytogenes. *Poultry Science*, *74*, 723–731.

Cocolin, L., Mataragas, M., Bourdichon, F., Doulgeraki, A., Pilet, M. F., Jagadeesan, B., Rantsiou, K., & Phister, T. (2018). Next generation microbiological risk assessment meta-omics: The next need for integration. *International Journal of Food Microbiology*, *287*, 10–17. https://doi.org/10.1016/j.ijfoodmicro.2017.11.008

Code of Federal Regulations (CFR). 9 CFR. 381.412. Retrieved September 29, 2024, from https://www.ecfr.gov/current/title-9/section-381.412

Codex Alimentarius Commission. (2015). *Procedural Manual. Twenty-Fourth Edition*. Rome, Italy. http://www.fao.org/3/a-i5079e.pdf

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 37–46. https://doi.org/10.1177/001316446002000104

Coleman, M. E., & Marks, H. M. (1999). Qualitative and quantitative risk assessment. *Food Control*, *10*(4–5), 289–297. https://doi.org/10.1017/9781316711408.007

Collineau, L., Boerlin, P., Carson, C. A., Chapman, B., Fazil, A., Hetman, B., McEwen, S. A., Parmley, E. J., Reid-Smith, R. J., Taboada, E. N., & Smith, B. A. (2019). Integrating whole-genome sequencing data into quantitative risk assessment of foodborne antimicrobial resistance: a review of opportunities and challenges. *Frontiers in Microbiology, 10*, 1107. https://doi.org/10.3389/fmicb.2019.01107

Collineau, L., Chapman, B., Bao, X., Sivapathasundaram, B., Carson, C. A., Fazil, A., Reid-Smith, R. J., & Smith, B. A. (2020a). A farm-to-fork quantitative risk assessment model for *Salmonella* Heidelberg resistant to third-generation cephalosporins in broiler chickens in Canada. *International Journal of Food Microbiology*, *330*. https://doi.org/10.1016/j.ijfoodmicro.2020.108559

Collineau, L., Phillips, C., Chapman, B., Agunos, A., Carson, C., Fazil, A., Reid-Smith, R. J., & Smith, B. A. (2020b). A within-flock model of *Salmonella* Heidelberg transmission in broiler chickens. *Preventive Veterinary Medicine*, *174*. https://doi.org/10.1016/j.prevetmed.2019.104823

Cook, P. W., & Nightingale, K. K. (2018). Use of omics methods for the advancement of food quality and food safety. *Animal Frontiers*, *8*(4), 33–41. https://doi.org/10.1093/af/vfy024

Csonka, L. N. & Epstein, W. (1996). Osmoregulation. In: Neidhardt, F.C. (Ed.), *Escherichia coli* and *Salmonella cellular and molecular biology*. ASM Press, Washington DC.

Csonka, L. N. (1989). Physiological and genetic responses of bacteria to osmotic stress. *Microbiological Reviews*, *53*(1), 121–147. https://doi.org/10.1128/mmbr.53.1.121-147.1989

Davis, J. J., Wattam, A. R., Aziz, R. K., Brettin, T., Butler, R., Butler, R. M., Chlenski, P., Conrad, N., Dickerman, A., Dietrich, E. M., Gabbard, J. L., Gerdes, S., Guard, A., Kenyon, R. W., MacHi, D., Mao, C., Murphy-Olson, D., Nguyen, M., Nordberg, E. K., … Stevens, R. (2020). The PATRIC Bioinformatics Resource Center: Expanding data and analysis capabilities. *Nucleic Acids Research*, *48*(D1), D606–D612. https://doi.org/10.1093/nar/gkz943

Davis, K. M., & Isberg, R. R. (2016). Defining heterogeneity within bacterial populations via single cell approaches. *BioEssays*, *38*(8),782 –790. https://doi.org/10.1002/bies.201500121

Dawoud, T. M., Davis, M. L., Park, S. H., Kim, S. A., Kwon, Y. M., Jarvis, N., O'Bryan, C. A., Shi, Z., Crandall, P. G., & Ricke, S. C. (2017). The potential link between thermal resistance and virulence in *Salmonella*: A review. *Frontiers in Veterinary Science*, *4*. https://doi.org/10.3389/fvets.2017.00093

De Mesquita Souza Saraiva, M., Lim, K., do Monte, D. F. M., Givisiez, P. E. N., Alves, L. B. R., de Freitas Neto, O. C., Kariuki, S., Júnior, A. B., de Oliveira, C. J. B., & Gebreyes, W. A. (2022). Antimicrobial resistance in the globalized food chain: a One Health perspective applied to the poultry industry. In *Brazilian Journal of Microbiology* (Vol. 53, Issue 1, pp. 465–486). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/s42770-021-00635-8

DeDonder, S., Jacob, C. J., Surgeoner, B. V., Chapman, B., Phebus, R., & Powell, D. A. (2009). Self-reported and observed behavior of primary meal preparers and adolescents during preparation of frozen, uncooked, breaded chicken products. *British Food Journal*, *111*(9), 915–929. https://doi.org/10.1108/00070700910992844

Demple, B., Hidalgo, E., & Ding, H. (1999). Transcriptional regulation via redox-sensitive iron-sulphur centres in an oxidative stress response. Biochemical Society Symposia, 64, 119–128.

Díez-García, M., Capita, R., & Alonso-Calleja, C. (2012). Influence of serotype on the growth kinetics and the ability to form biofilms of *Salmonella* isolates from poultry. *Food Microbiology*, *31*(2), 173–180. https://doi.org/10.1016/j.fm.2012.03.012

Djordjevic, S. P., Jarocki, V. M., Seemann, T., Cummins, M. L., Watt, A. E., Drigo, B., Wyrsch, E. R., Reid, C. J., Donner, E., & Howden, B. P. (2023). Genomic surveillance for antimicrobial resistance — a One Health perspective. In *Nature Reviews Genetics*. https://doi.org/10.1038/s41576-023-00649-y

Dogan, O. B., Aditya, A., Ortuzar, J., Clarke, J., & Wang, B. (2022). A systematic review and meta-analysis of the efficacy of processing stages and interventions for controlling *Campylobacter* contamination during broiler chicken processing. *Comprehensive Reviews in Food Science and Food Safety*, *21*(1), 227–271. https://doi.org/10.1111/1541-4337.12860

Dogan, O. B., Clarke, J., Mattos, F., & Wang, B. (2019). A quantitative microbial risk assessment model of *Campylobacter* in broiler chickens: Evaluating processing interventions. *Food Control*, *100*, 97–110. https://doi.org/10.1016/j.foodcont.2019.01.003

Du, X., Liu, Y., Sun, X., Zhao, J., Wang, J., Li, J., Liu, X., Wang, J., & Li, Y. (2020). The mRNA expression of ompF, invA and invE was associated with the ciprofloxacin-resistance in *Salmonella*. *Archives of Microbiology*, *202*(8), 2263–2268. https://doi.org/10.1007/s00203-020-01928-6

Du, Y., & Guo, Y. Machine learning techniques and research framework in foodborne disease surveillance system. *Food Control*, *2022*, 131. https://doi.org/10.1016/j.foodcont.2021.108448

Dunn, L. L., Sharma, V., Chapin, T. K., Friedrich, L. M., Larson, C. C., Rodrigues, C., Jay-Russell, M., Schneider, K. R., & Danyluk, M. D. (2022). The prevalence and concentration of *Salmonella enterica* in poultry litter in the southern United States. *PLoS ONE*, *17*(5 5). https://doi.org/10.1371/journal.pone.0268231

EcoSure. (2007a). *EcoSure 2007 Colding Temperature Data*. Retrieved September 29, 2024, from https://www.foodrisk.org/resources/sendFile/7?download=0

EcoSure. (2007b). *EcoSure 2007 Cooking Temperature Data*. Retrieved September 29, 2024, from https://www.foodrisk.org/resources/sendFile/8?download=0

EFSA BIOHAZ Panel. Koutsoumanis, K., Allende, A., Alvarez-Ordóñez, A., Bolton, D., Bover-Cid, S., Chemaly, M., Davies, R., De Cesare, A., Hilbert, F., Lindqvist, R., Nauta, M., Peixe, L., Ru, G., Simmons, M., Skandamis, P., Suffredini, E., Jenkins, C., Malorny, B., Ribeiro Duarte, A. S., … Herman, L. (2019). Whole genome sequencing and metagenomics for outbreak investigation, source attribution and risk assessment of food-borne microorganisms. *EFSA Journal*, *17*(12). https://doi.org/10.2903/j.efsa.2019.5898

Eng, S. K., Pusparajah, P., Ab Mutalib, N. S., Ser, H. L., Chan, K. G., & Lee, L. H. (2015). *Salmonella*: A review on pathogenesis, epidemiology and antibiotic resistance. *Frontiers in Life Science*, *8*(3), 284–293. https://doi.org/10.1080/21553769.2015.1051243

Etter, A. J., West, A. M., Burnett, J. L., Wu, S. T., Veenhuizen, D. R., Ogas, R. A., & Oliver, H. F. (2019). *Salmonella enterica* subsp. *enterica* serovar Heidelberg food isolates associated with a salmonellosis outbreak have enhanced stress tolerance capabilities. https://doi.org/10.1128/AEM

Evans, K. S., Teisl, M. F., Lando, A. M., & Liu, S. T. (2020). Risk perceptions and food-handling practices in the home. *Food Policy*, *95*. https://doi.org/10.1016/j.foodpol.2020.101939

Fang, F. C., Frawley, E. R., Tapscott, T., & Vázquez-Torres, A. (2016). Bacterial stress responses during host infection. In *Cell Host and Microbe*, *20*(2), 133–143. https://doi.org/10.1016/j.chom.2016.07.009

Food and Drug Administration (FDA). (2024). NARMS Now. Rockville, MD: U.S. Department of Health and Human Services. Retrieved September 29, 2024, from https://www.fda.gov/animal-veterinary/national-antimicrobial-resistance-monitoring-system/narms-now-integrated-data

Feldgarden, M., Brover, V., Gonzalez-Escalona, N., Frye, J. G., Haendiges, J., Haft, D. H., Hoffmann, M., Pettengill, J. B., Prasad, A. B., Tillman, G. E., Tyson, G. H., & Klimke, W. (2021). AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-91456-0

Figueiredo, M. A. T., & Jain, A. K. (2002). Unsupervised Learning of Finite Mixture Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(3), 381–396.

Finn, S., Händler, K., Condell, O., Colgan, A., Cooney, S., McClure, P., Amézquita, A., Hinton, J. C. D., & Fanning, S. (2013). ProP is required for the survival of desiccated *Salmonella enterica* serovar Typhimurium cells on a stainless steel surface. *Applied and Environmental Microbiology*, *79*(14), 4376–4384. https://doi.org/10.1128/AEM.00515-13

Florensa, A. F., Kaas, R. S., Clausen, P. T. L. C., Aytan-Aktug, D., & Aarestrup, F. M. (2022). ResFinder – an open online resource for identification of antimicrobial resistance genes in next-generation sequencing data and prediction of phenotypes from genotypes. *Microbial Genomics*, *8*(1). https://doi.org/10.1099/mgen.0.000748

Foley, S. L., Johnson, T. J., Ricke, S. C., Nayak, R., & Danzeisen, J. (2013). *Salmonella* Pathogenicity and Host Adaptation in Chicken-Associated Serovars. *Microbiology and Molecular Biology Reviews*, *77*(4), 582–607. https://doi.org/10.1128/mmbr.00015-13

Fong, K., & Wang, S. (2016). Heat resistance of *Salmonella enterica* is increased by pre-adaptation to peanut oil or sub-lethal heat exposure. *Food Microbiology*, *58*, 139–147. https://doi.org/10.1016/j.fm.2016.04.004

Foster, J. W. (2001). Acid Stress Responses of *Salmonella* and E. coli: Survival Mechanisms, Regulation, and Implications for Pathogenesis. *Journal of Microbiology*, *39*(2), 89–94.

Friedman, J., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: a statistical view of boosting. *The Annals of Statistics*, *28*(2), 337–374.

Fritsch, L., Guillier, L., & Augustin, J. C. (2018). Next generation quantitative microbiological risk assessment: Refinement of the cold smoked salmon-related listeriosis risk model by integrating genomic data. *Microbial Risk Analysis*, *10*, 20–27. https://doi.org/10.1016/j.mran.2018.06.003

Frossard, S. M., Khan, A. A., Warrick, E. C., Gately, J. M., Hanson, A. D., Oldham, M. L., Sanders, D. A., & Csonka, L. N. (2012). Identification of a third osmoprotectant transport system, the osmU system, in *Salmonella enterica*. *Journal of Bacteriology*, *194*(15), 3861–3871. https://doi.org/10.1128/JB.00495-12

Frye, J. G., & Jackson, C. R. (2013). Genetic mechanisms of antimicrobial resistance identified in *Salmonella enterica*, Escherichia coli, and Enteroccocus spp. isolated from U.S. food animals. *Frontiers in Microbiology*, *4*(MAY). https://doi.org/10.3389/fmicb.2013.00135

Gast, R. K., Dittoe, D. K., & Ricke, S. C. (2024). *Salmonella* in eggs and egg-laying chickens: pathways to effective control. *Critical Reviews in Microbiology*, *50* (1), 39–63. https://doi.org/10.1080/1040841X.2022.2156772

Gennis, R. B., & Stewart, V. (1996). Respiration. In F. C. Neidhardt, R. Curtiss III, J. L.Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, & H. E. Umbarger (Eds.), *Escherichia coli and Salmonella: Cellular and molecular biology* (pp. 217–261). Washington, D.C.: ASM Press.

Gershenson, C. (2003). Artificial neural networks for beginners. *arXiv*. https://arxiv.org/ftp/cs/papers/0308/0308031.pdf

Gmeiner, A., Njage, P. M. K., Hansen, L. T., Aarestrup, F. M., & Leekitcharoenphon, P. (2024). Predicting Listeria monocytogenes virulence potential using whole genome sequencing and machine learning. *International Journal of Food Microbiology*, *410*, 110491. https://doi.org/10.1016/j.ijfoodmicro.2023.110491

Godínez-Oviedo, A., Sampedro, F., B. J. P., Garcés-Vega, F. J., & Hernández-Iturriaga, M. (2022). Genotypic and phenotypic quantitative microbial risk assessment model of human salmonellosis related to the consumption of chicken meat in the central region of Mexico. *Food Research International*, *162*, 11901. https://doi.org/https://doi.org/10.1016/j.foodres.2022.11190

Golden, C. E., & Mishra, A. (2020). Prevalence of *Salmonella* and *Campylobacter* spp. In alternative and conventionally produced chicken in the United States: A systematic review and meta-analysis. *Journal of Food Protection*, *83*(7), 1181–1197. https://doi.org/10.4315/JFP-19-538

Golden, C. E., & Mishra, A. (2021). Assessing the Risk of Salmonellosis from Consumption of Conventionally and Alternatively Produced Broiler Meat Prepared In-Home in the United States. *Microbial Risk Analysis*, *18*. https://doi.org/10.1016/j.mran.2021.100160

González, R. J., Sampedro, F., Feirtag, J. M., Sánchez-Plata, M. X., & Hedberg, C. W. (2019). Prioritization of chicken meat processing interventions on the basis of reducing the *Salmonella* residual relative risk. *Journal of Food Protection*, *82*(9), 1575–1582. https://doi.org/10.4315/0362-028X.JFP-19-033

Google ML. (2022). Imbalanced data. https://developers.google.com/machine-learning/data-prep/construct/sampling-splitting/imbalanced-data.

Görke, B., & Stülke, J. (2008). Carbon catabolite repression in bacteria: Many ways to make the most out of nutrients. *Nature Reviews Microbiology*, *6*(8), 613–624. https://doi.org/10.1038/nrmicro1932

Grijspeerdt, K., & Vanrolleghem, P. (1999). Estimating the parameters of the Baranyi model for bacterial growth. *Food Microbiology*, *16*, 593–605.

Guillén, S., Nadal, L., Álvarez, I., Mañas, P., & Cebrián, G. (2021). Impact of the resistance responses to stress conditions encountered in food and food processing environments on the virulence and growth fitness of non-typhoidal salmonellae. In *Foods* (Vol. 10, Issue 3). MDPI AG. https://doi.org/10.3390/foods10030617

Guillier, L., Palma, F., & Fritsch, L. (2022). Taking account of genomics in quantitative microbial risk assessment: what methods? What issues? In *Current Opinion in Food Science* (Vol. 48). Elsevier Ltd. https://doi.org/10.1016/j.cofs.2022.100922

Guisbert, E., Yura, T., Rhodius, V. A., & Gross, C. A. (2008). Convergence of molecular, modeling, and systems approaches for an understanding of the *Escherichia coli* heat shock response. *Microbiology and Molecular Biology Reviews*, *72*(3), 545–554. https://doi.org/10.1128/mmbr.00007-08

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Gutierrez, A., De, J., & Schneider, K. R. (2020). Prevalence, concentration, and antimicrobial resistance profiles of *Salmonella* isolated from Florida poultry litter. *Journal of Food Protection*, *83*(12), 2179–2186. https://doi.org/10.4315/JFP-20-215

Haas, C. N., Rose, J. B., & Gerba, C. P. (2014). *Quantitative Microbial Risk Assessment*. Hoboken, NJ: Wiley.

Haddad, N., Johnson, N., Kathariou, S., Métris, A., Phister, T., Pielaat, A., Tassou, C., Wells-Bennik, M. H. J., & Zwietering, M. H. (2018). Next generation microbiological risk assessment—Potential of omics data for hazard characterisation. *International Journal of Food Microbiology*, *287*, 28–39. https://doi.org/10.1016/j.ijfoodmicro.2018.04.015

Hall, H. K., & Foster, J. W. (1996). The role of Fur in the acid tolerance response of *Salmonella* Typhimurium is physiologically and genetically separable from its role in iron acquisition. *Journal of Bacteriology*, *178*(19), 5683–5691. https://doi.org/10.1128/jb.178.19.5683-5691.1996

Hancock, V., Ferrières, L., & Klemm, P. (2008). The ferric yersiniabactin uptake receptor FyuA is required for efficient biofilm formation by urinary tract infectious *Escherichia coli* in human urine. *Microbiology*, *154*(1), 167–175. https://doi.org/10.1099/mic.0.2007/011981-0

Hardesty, L. (2017). *Explained: Neural networks.* MIT News. https://news.mit.edu/2017/explained-neural-networks-deep-learning-0414.

Hearst, M. A. (1998). SVMs - A practical consequence of learning theory. *IEEE Intelligent Systems and Their Applications*, *13*(4), 18–21. https://doi.org/10.1109/5254.708428

Hiura, S., Koseki, S., & Koyama, K. (2021). Prediction of population behavior of *Listeria monocytogenes* in food using machine learning and a microbial growth and survival database. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-90164-z

Hoffmann, M., Luo, Y., Monday, S. R., Gonzalez-Escalona, N., Ottesen, A. R., Muruvanda, T., ... & Brown, E. W. (2016). Tracing origins of the *Salmonella* Bareilly strain causing a food-borne outbreak in the United States. *The Journal of Infectious Diseases*, *213*(4), 502-508. https://doi.org/10.1093/infdis/jiv297

Huber, L., Agunos, A., Gow, S. P., Carson, C. A., & Van Boeckel, T. P. (2021). Reduction in antimicrobial use and resistance to *Salmonella, Campylobacter*, and *Escherichia coli* in broiler

chickens, Canada, 2013-2019. In *Emerging Infectious Diseases, 27*(9), 2434–2444. https://doi.org/10.3201/eid2709.204395

Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., Von Mering, C., & Bork, P. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Molecular Biology and Evolution*, *34*(8), 2115–2122. https://doi.org/10.1093/molbev/msx148

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., Letunic, I., Rattei, T., Jensen, L. J., Von Mering, C., & Bork, P. (2019). EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, *47*(D1), D309–D314. https://doi.org/10.1093/nar/gky1085

IFSAC (Interagency Food Safety Analytics Collaboration). (2023). Foodborne illness source attribution estimates for *Salmonella*, *Escherichia coli* O157, and *Listeria monocytogenes* — United States, 2021. https://www.cdc.gov/foodsafety/ifsac/pdf/P19-2021-report-TriAgency-508.pdf

Immerseel, F., Eeckhaut, V., Boyen, F., Pasmans, F., Haesebrouck, F., & Ducatelle, R. (2008). Mutations influencing expression of the *Salmonella enterica* serovar Enteritidis pathogenicity island I key regulator hilA. *Antonie van Leeuwenhoek*, *94*(3), 455–461. https://doi.org/10.1007/s10482-008-9262-y

Jefferies, D., Shearer, J., & Khalid, S. (2019). Role of O-antigen in response to mechanical stress of the *E. coli* outer membrane: insights from coarse-grained MD simulations. *Journal of Physical Chemistry B*, *123*(17), 3567–3575. https://doi.org/10.1021/acs.jpcb.8b12168

Karanth, S., Tanui, C. K., Meng, J., & Pradhan, A. K. (2022). Exploring the predictive capability of advanced machine learning in identifying severe disease phenotype in *Salmonella enterica*. *Food Research International*, *151*. https://doi.org/10.1016/j.foodres.2021.110817

Kempf, B., & Bremer, E. (1998). Uptake and synthesis of compatible solutes as microbial stress responses to high-osmolality environments. *Archives of Microbiology*, *170*(5), 319–330. https://doi.org/10.1007/s002030050649

Kenyon, W. J., Nicholson, K. L., Rezuchova, B., Homerova, D., Garcia-del Portillo, F., Finlay, B. B., Pallen, M. J., Kormanec, J., & Spector, M. P. (2007). σS-Dependent carbon-starvation induction of pbpG (PBP 7) is required for the starvation-stress response in *Salmonella enterica* serovar Typhimurium. *Microbiology*, *153*(7), 2148–2158. https://doi.org/10.1099/mic.0.2007/005199-0

Kenyon, W. J., Sayers, D. G., Humphreys, S., Roberts, M., & Spector, M. P. (2002). The starvation-stress response of *Salmonella enterica* serovar Typhimurium requires σE-, but not CpxR-regulated extracytoplasmic functions. *Microbiology*, *148*(1), 113–122. https://doi.org/10.1099/00221287-148-1-113

Khan, M. A. S., Islam, Z., Barua, C., Sarkar, M. M. H., Ahmed, M. F., & Rahman, S. R. (2024). Phenotypic characterization and genomic analysis of a *Salmonella* phage L223 for biocontrol of *Salmonella* spp. in poultry. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-024-64999-1

Kieboom, J., & Abee, T. (2006). Arginine-dependent acid resistance in *Salmonella enterica* serovar Typhimurium. *Journal of Bacteriology*, *188*(15), 5650–5653. https://doi.org/10.1128/JB.00323-06

Kim, M., Barnett-Neefs, C., Chavez, R. A., Kealey, E., Wiedmann, M., & Stasiewicz, M. J. (2024). Risk Assessment Predicts Most of the Salmonellosis Risk in Raw Chicken Parts is Concentrated in Those Few Products with High Levels of High-Virulence Serotypes of *Salmonella*. *Journal of Food Protection*, *87*(7). https://doi.org/10.1016/j.jfp.2024.100304

Kim, T. H., Choi, J. H., Choi, Y. S., Kim, H. Y., Kim, S. Y., Kim, H. W., & Kim, C. J. (2011). Physicochemical properties of thawed chicken breast as affected by microwave power levels. *Food Science and Biotechnology*, *20*(4), 971–977. https://doi.org/10.1007/s10068-011-0134-2

Kong, Q., Yang, J., Liu, Q., Alamuri, P., Roland, K. L., & Curtiss, R. (2011). Effect of deletion of genes involved in lipopolysaccharide core and O-antigen synthesis on virulence and immunogenicity of *Salmonella enterica* serovar Typhimurium. *Infection and Immunity*, *79*(10), 4227–4239. https://doi.org/10.1128/IAI.05398-11

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: Scalable and accurate long-read assembly via adaptive κ-mer weighting and repeat separation. *Genome Research*, *27*(5), 722–736. https://doi.org/10.1101/gr.215087.116

Kosa, K. M., Cates, S. C., Bradley, S., Chambers, E., & Godwin, S. (2015). Consumer-reported handling of raw poultry products at home: Results from a national survey. *Journal of Food Protection*, *78*(1), 180–186. https://doi.org/10.4315/0362-028X.JFP-14-231

Kosa, K. M., Cates, S. C., Brophy, J., Godwin, S., Chambers, D., & Chambers, E. (2019). Older adults and parents of young children have different handling practices for raw poultry. *Journal of Food Protection*, *82*(2), 200–206. https://doi.org/10.4315/0362-028X.JFP-18-323

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software, 28*(5), 1–26. https://doi.org/10.18637/jss.v028.i05

Kuhn, M. (2019). The caret Package. http://topepo.github.io/caret/.

Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling*. New York: Springer.

Lammerding, A. M. (1997). An overview of microbial food safety risk assessment. *Journal of Food Protection, 60*(11), 1420-1425. https://doi.org/10.4315/0362-028X-60.11.1420

Lammerding, A. M., & Fazil, A. (2000). Hazard identification and exposure assessment for microbial food safety risk assessment. *International Journal of Food Microbiology, 58*, 147–157. https://doi.org/10.1016/S0168-1605(00)00269-5

Lammerding, A. M., & Paoli, G. M. (1997). Quantitative risk assessment: an emerging tool for emerging foodborne pathogens. *Emerging Infectious Diseases, 3*(4), 483. https://doi:10.3201/eid0304.970411

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174.

Latimer, H. K., Jaykus, L. A., Morales, R. A., Cowen, P., & Crawford-Brown, D. (2001). A weighted composite dose-response model for human salmonellosis. *Risk Analysis*, *21*(2), 295–306. https://doi.org/10.1111/0272-4332.212112

Lee, K. Y., Atwill, E. R., Pitesky, M., Huang, A., Lavelle, K., Rickard, M., Shafii, M., Hung-Fan, M., & Li, X. (2022). Antimicrobial Resistance Profiles of Non-typhoidal *Salmonella* from Retail Meat Products in California, 2018. *Frontiers in Microbiology*, *13*. https://doi.org/10.3389/fmicb.2022.835699

Lee, Y. H., Kim, B. H., Kim, J. H., Yoon, W. S., Bang, S. H., & Park, Y. K. (2007). CadC has a global translational effect during acid adaptation in *Salmonella enterica* serovar Typhimurium. *Journal of Bacteriology*, *189*(6), 2417–2425. https://doi.org/10.1128/JB.01277-06

Leffer, A. M., Kuttel, J., Martins, L. M., Pedroso, A. C., Astolfi-Ferreira, C. S., Ferreira, F., & Ferreira, A. J. P. (2010). Vectorial Competence of Larvae and Adults of Alphitobius diaperinus in the Transmission of *Salmonella* Enteritidis in Poultry. *Vector-Borne and Zoonotic Diseases*, *10*(5), 481–487. https://doi.org/10.1089/vbz.2008.0089

Li, H. (2016). Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics, 32*(14), 2103–2110. https://doi.org/10.1093/bioinformatics/btw152

Lianou, A., & Koutsoumanis, K. P. (2011). Effect of the growth environment on the strain variability of *Salmonella enterica* kinetic behavior. *Food Microbiology*, *28*(4), 828–837. https://doi.org/10.1016/j.fm.2010.04.006

Ling, C. X., & Sheng, V. S. (2011). Cost-sensitive learning. In: Sammut, C., Webb, G.I. (eds) Encyclopedia of machine learning. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-30164-8_181

Liu, Y., Wu, H., Sun, Z., Xu, X., & Liu, F. (2021). Contamination and biofilm formation of foodborne and opportunistic pathogens in yellow-feathered chicken carcass. *Foodborne Pathogens and Disease*, *18*(3), 210–218. https://doi.org/10.1089/fpd.2020.2876

Lu, M., Liu, H., Lu, H., Liu, R., & Liu, X. (2020). Characterization and Genome Analysis of a Novel *Salmonella* Phage vB_SenS_SE1. *Current Microbiology*, *77*(7), 1308–1315. https://doi.org/10.1007/s00284-020-01879-7

Lunardon, N., Menardi, G., & Torelli, N. (2014). ROSE: A package for binary imbalanced learning. *The R Journal*, *6*.

Maguire, F., Rehman, M. A., Carrillo, C., Diarra, M. S., & Beiko, R. G. (2019). Identification of primary antimicrobial resistance drivers in agricultural nontyphoidal *Salmonella enterica* serovars by using machine learning. *mSystems*, *4*(4), 10–1128. https://doi.org/10.1128/msystems.00211-19

Mattick, K. L., Jørgensen, F., Legan, J. D., Cole, M. B., Porter, J., Lappin-Scott, H. M., & Humphrey, T. J. (2000). Survival and filamentation of *Salmonella enterica* serovar Enteritidis PT4 and *Salmonella enterica* serovar Typhimurium DT104 at low water activity. *Applied and*

Environmental Microbiology, *66*(4), 1274–1279. https://doi.org/10.1128/AEM.66.4.1274-1279.2000

Mazengia, E., Fisk, C., Liao, G., Huang, H., & Meschke, J. (2015). Direct Observational Study of the Risk of Cross-contamination during Raw Poultry Handling: Practices in Private Homes. *Food Protection Trends*, *35*(1), 8–23.

Mccullough, N. B., & Eisele, C. W. (1951). Experimental Human Salmonellosis: I. Pathogenicity of Strains of *Salmonella* Meleagridis and *Salmonella* Anatum Obtained from Spray-Dried Whole Egg. *The Journal of Infectious Diseases, 88*(3), 278–289. https://doi.org/https://doi.org/10.1093/infdis/88.3.278

McDermott, P. F., Tyson, G. H., Kabera, C., Chen, Y., Li, C., Folster, J. P., Ayers, S. L., Lam, C., Tate, H. P., & Zhao, S. (2016). Whole-genome sequencing for detecting antimicrobial resistance in nontyphoidal *Salmonella*. *Antimicrobial Agents and Chemotherapy*, *60*(9), 5515–5520. https://doi.org/10.1128/AAC.01030-16

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemica Medica*, *22*(3), 276–282.

Mclachlan, G. J., Lee, S. X., & Rathnayake, S. I. (2019). Finite Mixture Models. *Annual Review of Statistics and Its Application*, *6*, 355–378. https://doi.org/10.1146/annurev-statistics

Meerburg, B. G., & Kijlstra, A. (2007). Role of rodents in transmission of *Salmonella* and Campylobacter. *Journal of the Science of Food and Agriculture*, *87*(15), 2774–2781. https://doi.org/10.1002/jsfa.3004

Menardi, G., & Torelli, N. (2014). Training and assessing classification rules with imbalanced data. *Data Mining and Knowledge Discovery*, 28, 92–122. https://doi.org/10.1007/s10618-012-0295-5

Mercer, R. G., Walker, B. D., Yang, X., McMullen, L. M., & Gänzle, M. G. (2017). The locus of heat resistance (LHR) mediates heat resistance in *Salmonella enterica*, *Escherichia coli* and *Enterobacter cloacae*. *Food Microbiology*, *64*, 96–103. https://doi.org/10.1016/j.fm.2016.12.018

Montville, R., Chen, Y., & Schaffner, D. W. (2001). Glove Barriers to Bacterial Cross-Contamination between Hands to Food. In *Journal of Food Protection* (Vol. 64, Issue 6).

Morita, M., Kanemori, M., Yanagi, H., & Yura, T. (1999). Heat-induced synthesis of σ32 in Escherichia coli: Structural and functional dissection of *rpoH* mRNA secondary structure. *Journal of Bacteriology*, *181*(2), 401–410. https://doi.org/10.1128/jb.181.2.401-410.1999

Murphy, R. Y., Duncan, L. K., Johnson, E. R., Davis, M. D., & Smith, A. J. N. (2002). Thermal Inactivation D-and z-Values of *Salmonella* Serotypes and *Listeria innocua* in Chicken Patties, Chicken Tenders, Franks, Beef Patties, and Blended Beef and Turkey Patties. *Journal of Food Protection*, *65*(1), 53–60.

National Advisory Committee on Microbiological Criteria for Foods. (NACMCF). (2024). Response to Questions Posed by the Food Safety and Inspection Service: Enhancing *Salmonella* Control in Poultry Products. *Journal of Food Protection*, *87*(2). https://doi.org/10.1016/j.jfp.2023.100168

National Center for Biotechnology Information (NCBI). (2024). New Milestone! NCBI Pathogen Detection Reaches 2 Million Isolates. Retrieved September 29, 2024, from https://ncbiinsights.ncbi.nlm.nih.gov/2024/09/03/milestone-ncbi-pathogen-detection-2m-isolates/

National Chicken Council (NCC). (2022). Food Safety and Inspection in the U.S. Broiler Chicken Industry. https://www.nationalchickencouncil.org/industry-issues/food-safety/.

Nagel, G. M., Bauermeister, L. J., Bratcher, C. L., Singh, M., & McKee, S. R. (2013). *Salmonella* and *Campylobacter* reduction and quality characteristics of poultry carcasses treated with various antimicrobials in a post-chill immersion tank. *International Journal of Food Microbiology*, *165*(3), 281–286. https://doi.org/10.1016/j.ijfoodmicro.2013.05.016

Nahm, F. S. (2022). Receiver operating characteristic curve: overview and practical use for clinicians. *Korean Journal of Anesthesiology, 75(*1), 25-36. https://doi.org/10.4097/kja.21209

Nair, A., Rawool, D. B., Doijad, S., Poharkar, K., Mohan, V., Barbuddhe, S. B., Kolhe, R., Kurkure, N. V., Kumar, A., Malik, S. V. S., & Balasaravanan, T. (2015). Biofilm formation and genetic diversity of *Salmonella* isolates recovered from clinical, food, poultry and environmental sources. *Infection, Genetics and Evolution*, *36*, 424–433. https://doi.org/10.1016/j.meegid.2015.08.012

Ngogo, F. A., Joachim, A., Abade, A. M., Rumisha, S. F., Mizinduko, M. M., & Majigo, M. V. (2020). Factors associated with *Salmonella* infection in patients with gastrointestinal complaints seeking health care at Regional Hospital in Southern Highland of Tanzania. *BMC Infectious Diseases*, *20*(1). https://doi.org/10.1186/s12879-020-4849-7

Nikolenko, S. I., Korobeynikov, A. I., & Alekseyev, M. A. (2013). BayesHammer: Bayesian clustering for error correction in single-cell sequencing. *BMC Genomics*, *14*. https://doi.org/10.1186/1471-2164-14-S1-S7

Njage, P. M. K., Henri, C., Leekitcharoenphon, P., Mistou, M. Y., Hendriksen, R. S., & Hald, T. (2019). Machine learning methods as a tool for predicting risk of illness applying next-generation sequencing data. *Risk Analysis*, *39*(6), 1397–1413. https://doi.org/10.1111/risa.13239

Njage, P. M. K., Leekitcharoenphon, P., Hansen, L. T., Hendriksen, R. S., Faes, C., Aerts, M., & Hald, T. (2020). Quantitative microbial risk assessment based on whole genome sequencing data: Case of listeria monocytogenes. *Microorganisms*, *8*(11), 1–24. https://doi.org/10.3390/microorganisms8111772

Northcutt, J. K., Smith, D. P., Musgrove, M. T., Ingram, K. D., & Hinton, A. (2005). Microbiological Impact of Spray Washing Broiler Carcasses Using Different Chlorine Concentrations and Water Temperatures. *Poultry Science*, *84*, 1648–1652.

O'Neill, J. (2016). *Tackling drug-resistant infections globally: Final Report and Recommendations*. HM Gov Welcome Trust UK. https://amr-review.org/sites/default/files/160518_Final%20paper_with%20cover.pdf

Obe, T., Siceloff, A. T., Crowe, M. G., Morgan Scott, H., & Shariat, N. W. (2023). Combined Quantification and Deep Serotyping for *Salmonella* Risk Profiling in Broiler Flocks. *Applied and Environmental Microbiology*, *89*(4). https://doi.org/10.1128/aem.02035-22

Olson, R. D., Assaf, R., Brettin, T., Conrad, N., Cucinell, C., Davis, J. J., Dempsey, D. M., Dickerman, A., Dietrich, E. M., Kenyon, R. W., Kuscuoglu, M., Lefkowitz, E. J., Lu, J., Machi, D., Macken, C., Mao, C., Niewiadomska, A., Nguyen, M., Olsen, G. J., … Stevens, R. L. (2023). Introducing the Bacterial and Viral Bioinformatics Resource Center (BV-BRC): a resource combining PATRIC, IRD and ViPR. *Nucleic Acids Research*, *51*(1 D), D678–D689. https://doi.org/10.1093/nar/gkac1003

Oscar, T. P. (2002). Development and validation of a tertiary simulation model for predicting the potential growth of *Salmonella* Typhimurium on cooked chicken. *International Journal of Food Microbiology*, *76*, 177–190. https://doi.org/10.1016/S0168-1605(02)00025-9

Oscar, T. P. (2006). Validation of a Tertiary Model for Predicting Variation of *Salmonella* Typhimurium DT104 (ATCC 700408) Growth from a Low Initial Density on Ground Chicken Breast Meat with a Competitive Microflora *Journal of Food Protection*, *69*(9), 2048–2057. https://doi.org/10.4315/0362-028x-69.9.2048

Oscar, T. P. (2007). Predictive models for growth of *Salmonella* Typhimurium DT104 from low and high initial density on ground chicken with a natural microflora. *Food Microbiology*, *24*(6), 640–651. https://doi.org/10.1016/j.fm.2006.11.003

Oscar, T. P. (2009a). General Regression Neural Network and Monte Carlo Simulation Model for Survival and Growth of *Salmonella* on Raw Chicken Skin as a Function of Serotype, Temperature, and Time for Use in Risk Assessment. *Journal of Food Protection*, *72*(10), 2078–2087. https://doi.org/10.4315/0362-028x-72.10.2078

Oscar, T. P. (2009b). Predictive Model for Survival and Growth of *Salmonella* Typhimurium DT104 on Chicken Skin during Temperature Abuse †. In *Journal of Food Protection*, *72*(2), 304–314. https://doi.org/10.4315/0362-028x-72.2.304

Oscar, T. P. (2011). Extrapolation of a predictive model for growth of a low inoculum size of *Salmonella* Typhimurium DT104 on chicken skin to higher inoculum sizes. *Journal of Food Protection*, *74*(10), 1630–1638. https://doi.org/10.4315/0362-028X.JFP-11-127

Oscar, T. P. (2012). Growth of *Salmonella* Typhimurium DT104 at 30°C Is not affected by anatomical location on the chicken carcass. *Journal of Food Protection*, *75*(1), 164–168. https://doi.org/10.4315/0362-028X.JFP-11-248

Pang, H., Lambertini, E., Buchanan, R. L., Schaffner, D. W., & Pradhan, A. K. (2017). Quantitative microbial risk assessment for *Escherichia coli* O157:H7 in fresh-cut lettuce. *Journal of Food Protection, 80*(2), 302–311. https://doi:10.4315/0362-028X.JFP-16-246.

Papagianeli, S. D., Lianou, A., Aspridou, Z., Stathas, L., & Koutsoumanis, K. (2022). The magnitude of heterogeneity in individual-cell growth dynamics is an inherent characteristic of *Salmonella enterica* ser. Typhimurium strains. *Food Research International*, *162*. https://doi.org/10.1016/j.foodres.2022.111991

Peng, Y., Leung, H. C. M., Yiu, S. M., & Chin, F. Y. L. (2010). IDBA – A Practical Iterative de Bruijn Graph De Novo Assembler. *Research in Computational Molecular Biology*, 426–440. https://link.springer.com/chapter/10.1007/978-3-642-12683-3_28

Pérez-Rodríguez, F., Campos, D., Ryser, E., T., Buchholz, a L., Posada-Izquierdo, G., D., Marks, B. P., Zurera, G., & Todd, E. (2011). A mathematical risk model for *Escherichia coli* O157:H7 cross-contamination of lettuce during processing. *Food Microbiology 28*, 694–701. https://doi:10.1016/j.fm.2010.06.008.

Pérez-Rodríguez, F., Valero, A., Carrasco, E., García, R. M., & Zurera, G. (2008). Understanding and modelling bacterial transfer to foods: a review. *Trends in Food Science and Technology 19*, 131–144. https://doi:10.1016/j.tifs.2007.08.003.

Pornsukarom, S., Van Vliet, A. H. M., & Thakur, S. (2018). Whole genome sequencing analysis of multiple *Salmonella* serovars provides insights into phylogenetic relatedness, antimicrobial resistance, and virulence markers across humans, food animals and agriculture environmental sources. *BMC Genomics*, *19*(1). https://doi.org/10.1186/s12864-018-5137-4

Pradhan, A. K., Mishra, A., and Pang, H. (2019). Predictive Microbiology and Microbial Risk Assessment. In: Doyle, M. P., Diez-Gonzalez, F. and Hill, C. (eds.). *Food Microbiology: Fundamentals and Frontiers*. 5th Ed. Washington, D.C.: ASM Press. https://doi:10.1128/9781555819972.ch39.

Pye, H. V., Thilliez, G., Acton, L., Kolenda, R., Al-Khanaq, H., Grove, S., & Kingsley, R. A. (2023). Strain and serovar variants of *Salmonella enterica* exhibit diverse tolerance to food chain-related stress. *Food Microbiology*, *112*. https://doi.org/10.1016/j.fm.2023.104237

QMRA WIKI. (2017). *Salmonella* Anatum: Dose Response Experiments. Retrieved September 29, 2024, from https://qmrawiki.org/experiments/*Salmonella*-anatum

R Core Team (2024). R: A Language and Environment for Statistical Computing_. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rajula, H. S. R., Verlato, G., Manchia, M., Antonucci, N., & Fanos V. (2020). Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment. *Medicina (Kaunas). 8:56*(9):455. https://doi:10.3390/medicina56090455.

Rakitin, A. L., Yushina, Y. K., Zaiko, E. V., Bataeva, D. S., Kuznetsova, O. A., Semenova, A. A., Ermolaeva, S. A., Beletskiy, A. V., Kolganova, T. V., Mardanov, A. V., Shapovalov, S. O., & Tkachik, T. E. (2022). Evaluation of antibiotic resistance of *Salmonella* serotypes and whole-genome sequencing of multiresistant strains isolated from food products in Russia. *Antibiotics*, *11*(1). https://doi.org/10.3390/antibiotics11010001

Rantsiou, K., Kathariou, S., Winkler, A., Skandamis, P., Saint-Cyr, M. J., Rouzeau-Szynalski, K., & Amézquita, A. (2018). Next generation microbiological risk assessment: opportunities of whole genome sequencing (WGS) for foodborne pathogen surveillance, source tracking and risk assessment. *International Journal of Food Microbiology, 287*, 3-9. https://doi.org/10.1016/j.ijfoodmicro.2017.11.007

Ratkowsky, D. A., Olley, J., Mcmeekin, T. A., & Ball, A. (1982). Relationship Between Temperature and Growth Rate of Bacterial Cultures. *Journal of Bacteriology,149*(1), 1–5. https://doi.org/10.1128/jb.149.1.1-5.1982

Raut, R., Maharjan, P., & Fouladkhah, A. C. (2023). Practical Preventive Considerations for Reducing the Public Health Burden of Poultry-Related Salmonellosis. *International Journal of Environmental Research and Public Health*, *20*, 6654. https://doi.org/10.3390/ijerph20176654

Ravishankar, S., Zhu, L., & Jaroni, D. (2010). Assessing the cross contamination and transfer rates of *Salmonella enterica* from chicken to lettuce under different food-handling scenarios. *Food Microbiology*, *27*(6), 791–794. https://doi.org/10.1016/j.fm.2010.04.011

Ren, Y., Chakraborty, T., Doijad, S., Falgenhauer, L., Falgenhauer, J., Goesmann, A., Hauschild, A.-C., Schwengers, O., & Heider, D. (2022). Prediction of antimicrobial resistance based on whole-genome sequencing and machine learning. *Bioinformatics, 38*(2), 325–334. https://doi.org/10.1093/bioinformatics/btab681

Ricci, V., Tzakas, P., Buckley, A., Coldham, N. C., & Piddock, L. J. V. (2006). Ciprofloxacin-resistant *Salmonella enterica* serovar Typhimurium strains are difficult to select in the absence of AcrB and TolC. *Antimicrobial Agents and Chemotherapy*, *50*(1), 38–42. https://doi.org/10.1128/AAC.50.1.38-42.2006

Richards, A. K., Kue, S., Norris, C. G., & Shariat, N. W. (2023). Genomic and phenotypic characterization of *Salmonella enterica* serovar Kentucky. *Microbial Genomics*, *9*(9). https://doi.org/10.1099/mgen.0.001089

Ricke, S. C., Dawoud, T. M., Kim, S. A., Park, S. H., & Kwon, Y. M. (2018). *Salmonella* cold stress response: mechanisms and occurrence in foods. *Advances in Applied Microbiology*, *104*, 1–38. https://doi.org/10.1016/bs.aambs.2018.03.001

Rivera-Pérez, W., Barquero-Calvo, E., & Zamora-Sanabria, R. (2014). *Salmonella* contamination risk points in broiler carcasses during slaughter line processing. *Journal of Food Protection*, *77*(12), 2031–2034. https://doi.org/10.4315/0362-028X.JFP-14-052

Ryan, M. P., O'Dwyer, J., & Adley, C. C. (2017). Evaluation of the Complex Nomenclature of the Clinically and Veterinary Significant Pathogen *Salmonella*. *BioMed Research International*, *2017.* https://doi.org/10.1155/2017/3782182

Sabbagh, S. C., Forest, C. G., Lepage, C., Leclerc, J. M., & Daigle, F. (2010). So similar, yet so different: Uncovering distinctive features in the genomes of *Salmonella enterica* serovars Typhimurium and Typhi. *FEMS Microbiology Letters*, 305, 1–13). https://doi.org/10.1111/j.1574-6968.2010.01904.x

Scallan, E., Hoekstra, R. M., Angulo, F. J., Tauxe, R. V., Widdowson, M. A., Roy, S. L., Jones, J. L., & Griffin, P. M. (2011a). Foodborne illness acquired in the United States—major pathogens. *Emerging Infectious Diseases, 17*(1), 7. https://doi:10.3201/eid1701.p11101

Scallan, E., Griffin, P. M., Angulo, F. J., Tauxe, R. V., & Hoekstra, R. M. (2011b). Foodborne illness acquired in the United States—unspecified agents. *Emerging Infectious Diseases, 17*(1), 16. https://doi:10.3201/eid1701.091101p2

Schlattmann P, Hoehne J, Verba M (2023). CAMAN: Finite Mixture Models and Meta-Analysis Tools – Based on C.A.MAN. R package version 0.78, https://CRAN.R-project.org/package=CAMAN

Schlattmann, P., Verba, M., Dewey, M., & Walther, M. (2015). Mixture models in diagnostic meta-analyses - Clustering summary receiver operating characteristic curves accounted for heterogeneity and correlation. *Journal of Clinical Epidemiology*, *68*(1), 61–72. https://doi.org/10.1016/j.jclinepi.2014.08.013

Schroeder, C. M., Jensen, E., Miliotis, M. D., Dennis, S. B., & Morgan, K. M. (2007). Microbial Risk Assessment. In: Simjee, S. (Ed.) *Foodborne Diseases. Infectious Disease*. Humana Press. https://doi.org/10.1007/978-1-59745-501-5_18

Shah, J., Desai, P. T., Chen, D., Stevens, J. R., & Weimer, B. C. (2013). Preadaptation to cold stress in *Salmonella enterica* serovar Typhimurium increases survival during subsequent acid stress exposure. *Applied and Environmental Microbiology*, *79*(23), 7281–7289. https://doi.org/10.1128/AEM.02621-13

Shaji, S., Selvaraj, R. K., & Shanmugasundaram, R. (2023). *Salmonella* Infection in Poultry: A Review on the Pathogen and Control Strategies. *Microorganisms, 11*. https://doi.org/10.3390/microorganisms11112814

Sia, C. M., Baines, S. L., Valcanis, M., Lee, D. Y. J., Gonçalves da Silva, A., Ballard, S. A., Easton, M., Seemann, T., Howden, B. P., Ingle, D. J., & Williamson, D. A. (2021). Genomic diversity of antimicrobial resistance in non-typhoidal *Salmonella* in Victoria, Australia. *Microbial Genomics*, *7*(12). https://doi.org/10.1099/mgen.0.000725

Siceloff, A. T., Waltman, D., & Shariat, N. W. (2022). Regional *Salmonella* Differences in United States Broiler Production from 2016 to 2020 and the Contribution of Multiserovar Populations to *Salmonella* Surveillance. *Applied and Environmental Microbiology*, *88*(8). https://doi.org/10.1128/aem.00204-22

Singer, R. S., Schrag, N. F. D., Ricke, I., & Apley, M. D. (2023). Antimicrobial usage in broiler chicken production in the United States. *Frontiers in Veterinary Science*. https://doi.org/https://doi.org/10.3389/fvets.2023.1139908

Singh, G., Yadav, M., Ghosh, C., & Rathore, J. S. (2021). Bacterial toxin-antitoxin modules: classification, functions, and association with persistence. In *Current Research in Microbial Sciences* (Vol. 2). Elsevier Ltd. https://doi.org/10.1016/j.crmicr.2021.100047

Smadi, H., & Sargeant, J. M. (2013). Quantitative Risk Assessment of Human Salmonellosis in Canadian Broiler Chicken Breast from Retail to Consumption. *Risk Analysis*, *33*(2), 232–248. https://doi.org/10.1111/j.1539-6924.2012.01841.x

Sodagari, H. R., Sohail, M. N., & Varga, C. (2024). Differences in the prevalence and antimicrobial resistance among non-typhoidal *Salmonella* serovars isolated from retail chicken meat across the United States of America, 2013–2020. *Food Control*, *165*. https://doi.org/10.1016/j.foodcont.2024.110701

Spector, M. P., & Kenyon, W. J. (2012). Resistance and survival strategies of *Salmonella enterica* to environmental stresses. *Food Research International*, *45*(2), 455–481. https://doi.org/10.1016/j.foodres.2011.06.056

Spector, M. P., DiRusso, C. C., Pallen, M. J., Garcia Del Portillo, F., Dougan, G., & Finlay, B. B. (1999). The medium-/long-chain fatty acyl-CoA dehydrogenase (fadF) gene of *Salmonella* Typhimurium is a phase 1 starvation-stress response (SSR) locus. *Microbiology*, *145*(1), 15–31. https://doi.org/10.1099/13500872-145-1-15

Staes, I., Passaris, I., Cambré, A., & Aertsen, A. (2019). Population heterogeneity tactics as driving force in *Salmonella* virulence and survival. *Food Research International*, *125*. https://doi.org/10.1016/j.foodres.2019.108560

Stathas, L., Aspridou, Z., & Koutsoumanis, K. (2024). Quantitative microbial risk assessment of *Salmonella* in fresh chicken patties. *Food Research International*, *178*. https://doi.org/10.1016/j.foodres.2024.113960

Stevens, E. L., Carleton, H. A., Beal, J., Tillman, G. E., Lindsey, R. L., Lauer, A. C., Pightling, A., Jarvis, K. G., Ottesen, A., Ramachandran, P., Hintz, L., Katz, L. S., Folster, J. P., Whichard, J. M., Trees, E., Timme, R. E., Mcdermott, P., Wolpert, B., Bazaco, M., … Braden, C. (2022). Use of Whole Genome Sequencing by the Federal Interagency Collaboration for Genomics for Food and Feed Safety in the United States. *Journal of Food Protection*, *85*(5), 755–772. https://doi.org/10.4315/JFP-21-437

Storz, G., & Zheng, M. (2000). Oxidative stress. In G. Storz, & R. Hengge-Aronis (Eds.), *Bacterial stress responses* (pp. 47–59). Washington, D.C.: ASM Press.

Strickland, A. J., Sampedro, F., & Hedberg, C. W. (2023). Quantitative Risk Assessment of *Salmonella* in Ground Beef Products and the Resulting Impact of Risk Mitigation Strategies on Public Health. Journal of Food Protection, 86(6), 100093. https://doi.org/10.1016/j.jfp.2023.100093

Sun, P., Reid, M. D., & Zhou, J. (2014). An improved multiclass LogitBoost using Adaptive-One-Vs-One. *Machine Learning*, *97*(3), 295–326. https://doi.org/10.1007/s10994-014-5434-3

Sutherland, C., Sim, C., Gleim, S., & Smyth, S. J. (2020). Consumer insights on Canada's food safety and food risk assessment system. *Journal of Agriculture and Food Research*, *2*. https://doi.org/10.1016/j.jafr.2020.100038

Tan, S. J., Nordin, S., Esah, E. M., & Mahror, N. (2022). *Salmonella spp*. in Chicken: Prevalence, Antimicrobial Resistance, and Detection Methods. In *Microbiology Research* (Vol. 13, Issue 4, pp. 691–705). MDPI. https://doi.org/10.3390/microbiolres13040050

Tanui, C. K., Benefo, E. O., Karanth, S., & Pradhan, A. K. (2022b). A Machine Learning Model for Food Source Attribution of *Listeria monocytogenes*. *Pathogens*, *11*(6). https://doi.org/10.3390/pathogens11060691

Tanui, C. K., Karanth, S., Njage, P. M. K., Meng, J., & Pradhan, A. K. (2022a). Machine learning-based predictive modeling to identify genotypic traits associated with *Salmonella enterica* disease endpoints in isolates from ground chicken. *LWT*, *154*. https://doi.org/10.1016/j.lwt.2021.112701

Tate, H., Folster, J. P., Hsu, C.-H., Chen, J., Hoffmann, M., Li, C., Morales, C., Tyson, G. H., Mukherjee, S., Brown, A. C., Green, A., Wilson, W., Dessai, U., Abbott, J., Joseph, L., Haro, J., Ayers, S., Mcdermott, P. F., & Zhao, S. (2017). *Comparative Analysis of Extended-Spectrum-Lactamase CTX-M-65-Producing Salmonella enterica Serovar Infantis Isolates from Humans, Food Animals, and Retail Chickens in the United States*. https://doi.org/10.1128/AAC

Tate, H., Hsu, C. H., Chen, J. C., Han, J., Foley, S. L., Folster, J. P., Francois Watkins, L. K., Reynolds, J., Tillman, G. E., Nyirabahizi, E., & Zhao, S. (2022). Genomic Diversity, Antimicrobial Resistance, and Virulence Gene Profiles of *Salmonella* Serovar Kentucky Isolated from Humans, Food, and Animal Ceca Content Sources in the United States. *Foodborne Pathogens and Disease*, *19*(8), 509–521. https://doi.org/10.1089/fpd.2022.0005

Thai-Nghe, N., Gantner, Z., & Schmidt-Thieme, L. (2010). Cost-sensitive learning methods for imbalanced data. *Proceedings of the International Joint Conference on Neural Networks*. https://doi.org/10.1109/IJCNN.2010.5596486

Thakur, S., & Gray, G. C. (2019). The mandate for a global "one health" approach to antimicrobial resistance surveillance. In *American Journal of Tropical Medicine and Hygiene* (Vol. 100, Issue 2, pp. 227–228). American Society of Tropical Medicine and Hygiene. https://doi.org/10.4269/ajtmh.18-0973

Tsoy, O., Ravcheev, D., & Mushegian, A. (2009). Comparative genomics of ethanolamine utilization. *Journal of Bacteriology*, *191*(23), 7157–7164. https://doi.org/10.1128/JB.00838-09

U.S. Census Bureau. (2024). *U.S. and World Population Clock*. Retrieved September 29, 2024, from https://www.census.gov/popclock/

U.S. Department of Agriculture Economic Research Service (USDA ERS). (2021). *Food Availability (Per Capita) Data System*. Retrieved September 29, 2024, from https://www.ers.usda.gov/data-products/food-availability-per-capita-data-system/food-availability-per-capita-data-system

U.S. Department of Agriculture Food Safety Inspection Service (USDA FSIS). (2019). *Chicken from Farm to Table*. Retrieved September 29, 2024, from https://www.fsis.usda.gov/food-safety/safe-food-handling-and-preparation/poultry/chicken-farm-table

U.S. Department of Agriculture Food Safety Inspection Service (USDA FSIS). (2022). *FY 2022-2026 Food Safety Key Performance Indicator*. Retrieved September 29, 2024, from https://www.fsis.usda.gov/inspection/inspection-programs/inspection-poultry-products/reducing-*Salmonella*-poultry/*Salmonella*-0

U.S. Department of Agriculture Food Safety Inspection Service (USDA FSIS). (2024a). *Quarterly Sampling Reports on Salmonella and Campylobacter*. Retrieved September 29, 2024, from https://www.fsis.usda.gov/science-data/data-sets-visualizations/microbiology/microbiological-testing-program-rte-meat-and-7

U.S. Department of Agriculture Food Safety Inspection Service (USDA FSIS). (2024b). *Salmonella Framework for Raw Poultry Products*. Retrieved September 29, 2024, from

https://www.fsis.usda.gov/policy/federal-register-rulemaking/federal-register-rules/*Salmonella*-framework-raw-poultry-products

U.S. Department of Agriculture Food Safety Inspection Service (USDA FSIS). (2024c). *Salmonella in Not-Ready-To-Eat Breaded Stuffed Chicken Products*. Retrieved September 29, 2024, from https://www.fsis.usda.gov/policy/federal-register-rulemaking/federal-register-rules/*Salmonella*-not-ready-eat-breaded-stuffed

U.S. Department of Agriculture National Agricultural Statistics Service (USDA NASS) (2024). *Poultry Slaughter*. Retrieved September 29, 2024, from https://usda.library.cornell.edu/concern/publications/3197xm04j?locale=en

UniProt Consortium. (2017). UniProt: the universal protein knowledgebase. *Nucleic Acids Research*, 45, D158–D169. https://doi:10.1093/nar/gkw1099

Vestby, L. K., Møretrø, T., Ballance, S., Langsrud, S., & Nesse, L. L. (2009). Survival potential of wild type cellulose deficient *Salmonella* from the feed industry. *BMC Veterinary Research*, *5*, 1–10. https://doi.org/10.1186/1746-6148-5-43

Wabeck, C. J. (2002). Processing Chicken Meat. In. Bell, D. D., & Weaver, W. D. (Eds.), *Commercial Chicken Meat and Egg Production* (pp. 899 – 920). Springer.

Wang, J., Vaddu, S., Bhumanapalli, S., Mishra, A., Applegate, T., Singh, M., & Thippareddi, H. (2023). A systematic review and meta-analysis of the sources of *Salmonella* in poultry production (pre-harvest) and their relative contributions to the microbial risk of poultry meat. *Poultry Science*, *102*(5). https://doi.org/10.1016/j.psj.2023.102566

Wang, W.-C., Li, Y., Slavik, M. F., & Xiong, H. (1997). Trisodium Phosphate and Cetylpyridinium Chloride Spraying on Chicken Skin to Reduce Attached *Salmonella* Typhimurium. In *Journal of Food Protection* (Vol. 60, Issue 8).

Wang, X., Bouzembrak, Y., Lansink, A. G. J. M. O., & van der Fels-Klerx, H. J. (2022). Application of machine learning to the monitoring and prediction of food safety: A review. *Comprehensive Reviews in Food Science and Food Safety*, *21*(1), 416–434. https://doi.org/10.1111/1541-4337.12868

Wang, Y., Ge, H., Wei, X., & Zhao, X. (2022). Research progress on antibiotic resistance of *Salmonella*. *Food Quality and Safety*, *6*. https://doi.org/10.1093/fqsafe/fyac035

Wang, Z., Zhu, S., Li, C., Lyu, L., Yu, J., Wang, D., Xu, Z., Ni, J., Gao, B., Lu, J., & Yao, Y. F. (2022). Gene essentiality profiling reveals a novel determinant of stresses preventing protein aggregation in *Salmonella*. *Emerging Microbes and Infections*, *11*(1), 1554–1571. https://doi.org/10.1080/22221751.2022.2081618

WBG (World Bank Group). (2017). *Drug-Resistant Infections: A Threat to Our Economic Future.* https://www.worldbank.org/en/topic/health/publication/drug-resistant-infections-a-threat-to-our-economic-future

Whiting, R. C., & Buchanan, R. L. (2008). Using Risk Assessment Principles in an Emerging Paradigm for Controlling the Microbial Safety of Foods. In: Schaffner, D. W. (ed.). *Microbial Risk Analysis of Foods*. Washington, D.C.: ASM Press. https://doi:10.1128/9781555815752.ch2.

Whyte, P., Collins, J. D., Mcgill, K., Monahan, C., & O'mahony, H. (2001). The Effect of Transportation Stress on Excretion Rates of *Campylobacters* in Market-Age Broilers. *Poultry Science*, *80*, 817–820.

World Health Organization (WHO). (2022). *Food Safety: Key Facts 2022*. https://www.who.int/news-room/fact-sheets/detail/food-safety

World Health Organization (WHO). (2023). *Antimicrobial resistance.* https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance

World Health Organization/ Food and Agriculture Organization of the United Nations (WHO/FAO). (2002). Risk assessments of *Salmonella* in eggs and broiler chickens. Retrieved September 29, 2024, from https://openknowledge.fao.org/server/api/core/bitstreams/72d30fc6-2c38-440a-a9b3-de18f8d9c025/content

World Organization for Animal Health (WOAH). (2023). *Seventh Annual Report on Antimicrobial Agents Intended for Use in Animals.* https://www.woah.org/en/document/seventh-annual-report-on-antimicrobial-agents-intended-for-use-in-animals-2/

Wright, G. D. (2005). Bacterial resistance to antibiotics: Enzymatic degradation and modification. *Advanced Drug Delivery Reviews*, *57*(10), 1451–1470). https://doi.org/10.1016/j.addr.2005.04.002

Wu, F., & Rodricks, J. V. (2020). Forty Years of Food Safety Risk Assessment: A History and Analysis. *Risk Analysis*, *40*, 2218–2230. https://doi.org/10.1111/risa.13624

Wu, H., Wang, Y., Wu, Y., Qiao, J., Li, H., Zheng, S., Xia, X., Cui, S., Wang, X., Xi, M., Meng, J., & Yang, B. (2015). Emergence of β-lactamases and Extended-Spectrum β-Lactamases (ESBLs) producing *Salmonella* in retail raw chicken in China. *Foodborne Pathogens and Disease*, *12*(3), 228–234. https://doi.org/10.1089/fpd.2014.1859

Wu, J., & Weiss, B. (1991). Two divergently transcribed genes, soxR and soxS, control a superoxide response regulon of Escherichia coli. *Journal of Bacteriology*, *173*(9), 2864–2871. https://doi.org/10.1128/jb.173.9.2864-2871.1991

Xiao, X., Wang, W., Zhang, J., Liao, M., Rainwater, C., Yang, H., & Li, Y. (2021). A quantitative risk assessment model of *Salmonella* contamination for the yellow-feathered broiler chicken supply chain in China. *Food Control*, *121*. https://doi.org/10.1016/j.foodcont.2020.107612

Xiao, X., Wang, W., Zhang, X., Zhang, J., Liao, M., Yang, H., Zhang, Q., Rainwater, C., & Li, Y. (2019). Modeling the reduction of *Salmonella* spp. On chicken breasts and wingettes during scalding for QMRA of the poultry supply chain in China. *Microorganisms*, *7*(6). https://doi.org/10.3390/microorganisms7060165

Xiong, H., Li, Y., Slavik, M. F., & Walkeri, J. T. (1998). Spraying Chicken Skin with Selected Chemicals to Reduce Attached *Salmonella* Typhimurium. In *Journal of Food Protection* (Vol. 61, Issue 3).

Yang, Y., Khoo, W. J., Zheng, Q., Chung, H. J., & Yuk, H. G. (2014). Growth temperature alters *Salmonella* Enteritidis heat/acid resistance, membrane lipid composition and stress/virulence related gene expression. *International Journal of Food Microbiology*, *172*, 102–109. https://doi.org/10.1016/j.ijfoodmicro.2013.12.006

Yang, Z., Li, Y., & Slavik2, I. (1998). Use of Antimicrobial Spray Applied with an Inside-Outside Birdwasher To Reduce Bacterial Contamination on Prechilled Chicken Carcasses. In *Journal of Food Protection* (Vol. 61, Issue 7).

Ye, B., He, S., Zhou, X., Cui, Y., Zhou, M., & Shi, X. (2019). Response to acid adaptation in *Salmonella enterica* serovar Enteritidis. *Journal of Food Science*, *84*(3), 599–605. https://doi.org/10.1111/1750-3841.14465

Yin, X., M'ikanatha, N. M., Nyirabahizi, E., McDermott, P. F., & Tate, H. (2021). Antimicrobial resistance in non-Typhoidal *Salmonella* from retail poultry meat by antibiotic usage-related production claims – United States, 2008–2017. *International Journal of Food Microbiology*, *342*. https://doi.org/10.1016/j.ijfoodmicro.2021.109044

Yoon, Y., Lee, H., Lee, S., Kim, S., & Choi, K. H. (2015). Membrane fluidity-related adaptive response mechanisms of foodborne bacterial pathogens under environmental stresses. *Food Research International*, *72*, 25–36. https://doi.org/10.1016/j.foodres.2015.03.016

Yura, T., & Nakahigashi, K. (1999). Regulation of the heat-shock response. *Current Opinion in Microbiology*, *2*(2), 153–158. https://doi.org/10.1016/S1369-5274(99)80027-7

Zerbino, D. R., & Birney, E. (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, *18*(5), 821–829. https://doi.org/10.1101/gr.074492.107

Zhang, C. Z., Ren, S. Q., Chang, M. X., Chen, P. X., Ding, H. Z., & Jiang, H. X. (2017). Resistance mechanisms and fitness of *Salmonella* Typhimurium and *Salmonella* Enteritidis mutants evolved under selection with ciprofloxacin in vitro. *Scientific Reports*, *7*(1). https://doi.org/10.1038%2Fs41598-017-09151-y

Zhou, K., Sun, L., Zhang, X., Xu, X., Mi, K., Ma, W., Zhang, L., & Huang, L. (2023). *Salmonella* antimicrobials inherited and the non-inherited resistance: mechanisms and alternative therapeutic strategies. *Frontiers in Microbiology*, *14*. https://doi.org/10.3389/fmicb.2023.1176317

Zhu, J., Bai, Y., Wang, Y., Song, X., Cui, S., Xu, H., Jiao, X., & Li, F. (2017). A risk assessment of salmonellosis linked to chicken meals prepared in households of China. *Food Control*, *79*, 279–287. https://doi.org/10.1016/j.foodcont.2017.04.003