FOODBORNE PATHOGENS AND DISEASE
Volume 21, Number 3, 2024
Mary Ann Liebert, Inc.
DOI: 10.1089/fpd.2023.0075

Review

A Review on Microbiological Source Attribution Methods of Human Salmonellosis: From Subtyping to Whole-Genome Sequencing

Rebeca Cardim Falcao,[1,2] Megan R. Edwards,[1,2] Matt Hurst,[3] Erin Fraser,[1,2] and Michael Otterstatter[1,2]

Abstract

Salmonella is one of the main causes of human foodborne illness. It is endemic worldwide, with different animals and animal-based food products as reservoirs and vehicles of infection. Identifying animal reservoirs and potential transmission pathways of *Salmonella* is essential for prevention and control. There are many approaches for source attribution, each using different statistical models and data streams. Some aim to identify the animal reservoir, while others aim to determine the point at which exposure occurred. With the advance of whole-genome sequencing (WGS) technologies, new source attribution models will greatly benefit from the discriminating power gained with WGS. This review discusses some key source attribution methods and their mathematical and statistical tools. We also highlight recent studies utilizing WGS for source attribution and discuss open questions and challenges in developing new WGS methods. We aim to provide a better understanding of the current state of these methodologies with application to *Salmonella* and other foodborne pathogens that are common sources of illness in the poultry and human sectors.

Keywords: source attribution, frequency-matching methods, *Salmonella*, whole-genome sequence, microbiological, machine learning, population genetic methods

Introduction

In studies of human foodborne illnesses, such as those caused by *Salmonella*, *Campylobacter*, and *Escherichia coli*, attributing the source of the pathogen is essential for better understanding transmission dynamics and developing efficient control strategies. Source attribution methods attribute human cases caused by a foodborne disease to different sources (Mughini-Gras et al., 2019; Pires et al., 2009).

They quantify the contribution of each source to the human disease burden through their linkage. This can help with the prioritization of intervention strategies. Source is a broad term, meaning the origin of the pathogen, and includes a range of groups, such as animal reservoirs and vehicles, depending on the attribution problem being tackled (Pires et al., 2009). For zoonotic pathogens, like *Salmonella*, animals may be hosts (organisms that harbour the pathogen) or carriers (hosts without discernible illness), where the pathogen lives and multiplies and are known as animal reservoirs. The transmission vehicles represent ways pathogens can travel from the reservoirs to humans.

Food, environment, and direct contact with animals are examples of a vehicle (Mughini-Gras et al., 2019; Pires et al., 2009; Wagenaar et al., 2013). These components can be potential sources in source attribution studies (Carstens et al., 2019; Ferrari et al., 2019). Pires et al. (2009) define points of

[1] British Columbia Centre for Disease Control, Vancouver, Canada.
[2] School of Population and Public Health, The University of British Columbia, Vancouver, Canada.
[3] Public Health Agency of Canada, Guelph, Canada.

This article has been updated on December 12, 2023 after first online publication of November 29, 2023 to reflect Open Access, with copyright transferring to the author(s), and a Creative Commons License (CC-BY) added (http://creativecommons.org/licenses/by/4.0/)

attribution as ''points in the food chain where human illness attribution can take place, such as production, distribution, and consumption.'' There are many approaches to source attribution, depending on the goal, questions of the study, data availability, and point of attribution (Mather et al., 2015; Mughini-Gras et al., 2019; Pires et al., 2014). The diversity of potential transmission sources can greatly complicate attempts at attribution. One of the reasons is the need to have robust and representative samples from all true sources (Mather et al., 2015; Mughini-Gras et al., 2019; Pires et al., 2014; Pires et al., 2009).

Salmonella is one of the most common causes of foodborne illness in the world (Kirk et al., 2015). *Salmonella* is a genus of Gram-negative rod-shaped bacteria comprising two species: *Salmonella enterica* and *Salmonella bongori*. *Salmonella enterica* is further categorized into six subspecies: *Salmonella enterica* subsp. *enterica*; *Salmonella enterica* subsp. *salamae*; *Salmonella enterica* subsp. *arizonae*; *Salmonella enterica* subsp. *diarizonae*; *Salmonella enterica* subsp. *houtenae*; and *Salmonella enterica* subsp. *indica* (Brenner et al., 2000; Eng et al., 2015). *Salmonella enterica* subsp. *enterica* is responsible for most *Salmonella* infections in humans (Eng et al., 2015; Kirk et al., 2015).

More than 2600 serotypes of *Salmonella* have been identified to date, and at least 50% of these serotypes belong to *Salmonella enterica* subsp. *enterica* (Eng et al., 2015; Thames and Theradiyil Sukumaran, 2020; World Health Organization, 2018). The most frequent Salmonella serovar isolated from reported human cases in Canada was *Salmonella* Enteritidis, corresponding to 35% of cases (Government of Canada, 2020). *Salmonella* outbreaks are frequently linked to animal reservoirs such as chickens (Hoelzer et al., 2011; Wessels et al., 2021) and pigs (Bearson, 2022; Hoelzer et al., 2011), and table eggs (vehicle) (Chousalkar et al., 2018; Popa and Papa, 2021).

In addition, food products such as cheese pasta, infant formula, mayonnaise, cucumbers, and other vegetables have been linked to cases of salmonellosis (Carstens et al., 2019; Laughlin et al., 2019; Popa and Papa, 2021), likely due to environmental or cross-contamination during processing. *Salmonella* can cause illness in hosts ranging from poultry to humans, with clinical manifestations of disease spanning enteric fever, gastroenteritis, bacteremia, and an asymptomatic chronic carrier state (Eng et al., 2015; World Health Organization, 2018). In livestock, including poultry, asymptomatic and persistent infection in the animal's digestive tract result in a carrier state, further facilitating transmission to humans (Hoelzer et al., 2011; Silva et al., 2014). Moreover, rodents, known carriers of *Salmonella*, can contaminate barn and farm environments (Anderson et al., 2006; Hoelzer et al., 2011).

Salmonella can also colonize plants in the field (Holden et al., 2009; Jechalke et al., 2019), survive in fresh produce (Beuchat, 2002; Critzer and Doyle, 2010), and survive in soil and water (Jechalke et al., 2019), underscoring the breadth of potential transmission sources (Silva et al., 2014). Understanding the complex interactions between humans, animals and their environments that may lead to disease spread is necessary to identify and address transmission pathways of pathogens such as *Salmonella*. A one health approach using multidisciplinary collaborative efforts is essential to develop effective methods, public policy, and interventions aimed at source attribution and disease control in host populations (Destoumieux-Garzón et al., 2018; Silva et al., 2014).

Source attribution provides several methods for evaluating interventions and pathways of transmission and infection, with varying data sources and data quality requirements. These methodologies include microbiological, epidemiological, expert elicitation, and intervention studies, with the choice of method often driven by the type and quality of available data (Pires et al., 2014; Pires et al., 2009). Aiding in the detection of local and global outbreaks of foodborne diseases is PulseNet, a laboratory network comprising standardized subtyping data of foodborne pathogens around the globe.

Most of its data consist of pulsed-field gel electrophoresis (PFGE) subtyping analysis, although globally and in Canada since 2017, PFGE has been replaced by whole-genome sequencing (WGS) (Government of Canada, 2020). It is noteworthy that, in the poultry industry, a primary source of *Salmonella*, several interventions have been tried to reduce *Salmonella* prevalence, including vaccination, cleaning and sanitation of barns, separation of flocks, and testing, with the aim to lessen human cases (Dórea et al., 2010; Totton et al., 2012; Trampel et al., 2014). Similar interventions have also been used in pigs and cattle (da Costa et al., 2021; Holschbach and Peek, 2018). However, assessing the efficacy of these interventions is challenging (Taylor et al., 2018), in part, due to the requirement for high-quality data from both human cases and from potential sources along the farm-to-fork continuum.

In this study, we review source attribution methodologies within the microbiological approach, which is the most frequently used approach for *Salmonella* (Barco et al., 2013; Mughini-Gras et al., 2018), and their challenges and limitations. We also discuss new challenges raised by recent advances in WGS used in source attribution methods, including the benefits and limitations of incorporating WGS into source attribution models. We focus this review particularly on methods for source attribution relevant to *Salmonella*, including methods used on *Campylobacter*, given that *Campylobacter* is also a foodborne pathogen that can populate different animal reservoirs. These methods could potentially be used with *Salmonella*.

Microbiological Source Attribution

Attribution of human cases of salmonellosis to sources of transmission and/or infection is essential for the identification of transmission hotspots, the development of control strategies, and the implementation and assessment of interventions. Source attribution methodologies can be performed at three levels in the farm-to-fork continuum: (1) Point of production, that is, animal reservoirs in farms, (2) point of distribution, that is, processing industries and retail; and (3) point of exposure, that is, food preparation and consumption (EFSA, 2008; Ravel et al., 2017).

Most of the previous work on *Salmonella* has focused on the point of production (Barco et al., 2015; David et al., 2013; De Knegt et al., 2015; Hald et al., 2007; Hald et al., 2004; Mullner et al., 2009; Ravel et al., 2017), enabling the assessment of different interventions in the control of pathogens at the reservoir level before reaching other possible transmission routes. Fewer studies have worked with the point of distribution (Guo et al., 2011), the point of exposure

(Christidis et al., 2020; Ravel et al., 2017), or utilizing a combination of data from different points (Boysen et al., 2014; Hurst et al., 2023; Mughini-Gras and van Pelt, 2014; Mughini-Gras et al., 2018; Mughini-Gras et al., 2014; Ravel et al., 2017).

Microbiological methodologies of source attribution can be further divided into microbial subtyping methods and comparative exposure assessment. Described in greater detail below, these approaches encompass several methods of source attribution and have both distinct and shared strengths and limitations (Pires et al., 2014).

Microbial subtyping methods

Microbial subtyping is a technique that allows for differentiation among bacterial isolates (Barco et al., 2013; Wiedmann, 2002). This method compares the set of subtypes of the pathogen in each source with the set of subtypes from human cases. The method relies on accurately matching and distributing the cases of human illness from a particular subtype over the possible sources where that subtype is found.

These so-called ''frequency-matching'' methods are the most commonly used for foodborne pathogen source attribution (EFSA, 2008; Mughini-Gras et al., 2018). One goal of subtyping is to identify strains that can discriminate among the potential sources or, said differently, identify strains that are source specific. The Kentucky serotype of *Salmonella*, for instance, is almost exclusively found in poultry manure, so any case found with the Kentucky strain is very likely to have originated from this reservoir (Dunn et al., 2022; Murray et al., 2023). A work by Hurst et al. (2023) provides potential metrics to evaluate the subtype definition used in the attribution model. These metrics can aid in achieving optimal balance among source specificity of subtypes, missing data, and level of discrimination power of subtypes.

When using a frequency-matching algorithm, it is important to include all possible sources of transmission of a pathogen so that cases may be matched to their true source (Barco et al., 2013; Mughini-Gras et al., 2019; Pires et al., 2014). Similarly, cases whose subtypes are not found in any of the sources should be discarded in frequency-matching methods. Thus, using a subtyping method that minimizes these unmatched cases is an important consideration (Pires et al., 2014; Pires et al., 2009). Subtypes are defined by either phenotyping methods (e.g., serotyping, phage-type, and antimicrobial resistance) or by genotyping methods (e.g., PFGE and comparative genomic fingerprinting [CGF]) (Barco et al., 2013; Ferrari et al., 2017; Yan et al., 2004). For *Salmonella* isolates, PFGE is the most extensively used method, standardized to support the comparison of isolates between human cases and sources and between and within countries.

Frequency-matching methods.

The Dutch model and the Hald model, and their modifications, have been extensively used for source attribution of foodborne pathogens (David et al., 2013; De Knegt et al., 2015; Guo et al., 2011; Hald et al., 2007; McLure et al., 2022; Mughini-Gras et al., 2018; Mughini-Gras et al., 2014; Vieira et al., 2016). The Dutch model is a frequentist model that compares the number of human cases of a pathogen subtype i with the number of isolates of the same subtype in each source s (Hald et al., 2004). The Hald model is based on the Dutch model and relies on estimating the expected number of human cases of subtype i from each source s, λ_{is}. To allow for appropriate uncertainty in the parameter estimation process, Bayesian inference is applied to the process (Hald et al., 2004; Mullner et al., 2009), with λ_{is} equal to

$$\lambda_{\mathrm{is}} = \mathrm{p_{is} q_i a_s},$$

where p_{is} is the proportion of subtype i in source s, q_i is the subtype-dependent factor that summarizes survivability, virulence, and transmissibility of the pathogen, and a_s is the food source-dependent factor representing the source ability to act as a vehicle for the foodborne pathogen and the differences in monitoring systems of each source. The observed number of human cases of subtype i, o_i, follows a Poisson distribution:

$$\mathrm{o_i} = \mathrm{Poisson}\left(\sum_{\mathrm{s}} \lambda_{\mathrm{is}}\right)$$

p_{is} is given by the data, where q_i and a_s are parameters of the model.

This results in an overspecified model, given that T_i is the total number of subtypes and T_s is the total number of sources, then there are $T_i + T_s$ parameters, but only T_i independent data points (David et al., 2013; Miller et al., 2017; Mullner et al., 2009). Extra assumptions on q_i and a_s are introduced to reduce the number of parameters, such as q_i is equal for some subtypes and a_s is equal for some food types or sources, yielding *a priori* grouping of subtypes (Hald et al., 2004).

However, no quantification of uncertainty among all possible groupings is available (Miller et al., 2017). The modified Hald model tackles this issue by modeling q_i as random observations from the distribution of characteristics of the pathogen given by $\log(\mathrm{q_i}) \sim \mathrm{N}(0, \tau)$, where τ controls the variation of these characteristics. The following prior distribution is used for $\tau \sim \mathrm{Gamma}(0.01, 0.01)$ (Mullner et al., 2009). Moreover, data can be split into different periods to achieve identifiability, given that q_i and a_s are assumed to be constant over time. However, the parameters of this model are still weakly identifiable, resulting in slow convergence of Markov chain Monte Carlo (MCMC) algorithms used to fit the method (David et al., 2013; Miller et al., 2017). MCMC algorithms are used to sample values from a probability distribution (posterior) using prior information and an underlying model (likelihood).

To make the original model more robust, the modified Hald model incorporates uncertainty on the prevalence parameter by defining $\mathrm{p_{is}} = \pi_{\mathrm{s}} \mathrm{r_{is}}$, where π_{s} is the prevalence over all types in source s and r_{is} is the relative occurrence of subtype i among the subtyped isolates from source s. A beta distribution is used as prior for $\mathrm{p_{is}}$ ($\pi_{\mathrm{s}} \sim \mathrm{Beta}(1,\ 1)$), and a Dirichlet distribution of size T_i is used as a prior for $\mathrm{r_{is}}$ ($\mathrm{r_{is}} \sim \mathrm{Dir}(1,\ 1, \ldots, 1)$) (Mullner et al., 2009).

Recent work by Miller et al. (2017) developed a new method that fits a joint model for both human cases and source samples. This method addresses the weak identifiability issue by using a nonparametric Bayesian clustering method to group the subtypes, thus incorporating uncertainty on all possible groups of subtypes. This approach reduces the number of parameters and does not make any strong assumption on τ.

With the advent of WGS, greater power of discrimination among isolates can be achieved, resulting in higher accuracy for source attribution inference. Moreover, WGS techniques have become more practical and less expensive. Studies have even shown that the benefits of WGS outweigh the cost (Alleweldt et al., 2021; Brown et al., 2021; Glass-Kaastra et al., 2022). As a result, WGS methods are gradually replacing phenotyping and other genotyping methods worldwide.

However, in resource-poor settings, it is still challenging to implement WGS at its full power and with robust sampling (Mather et al., 2015). WGS isolates can be compared across various degrees of similarity, providing different resolutions for genotyping. For example, single nucleotide polymorphism (SNP) analysis compares SNPs across each aligned genome, while core-genome multilocus sequence typing (cgMLST) and whole-genome multilocus sequence typing (wgMLST) use gene-to-gene comparisons. Defining the optimal resolution for genotyping is not clear and may depend on the sources examined (Collineau et al., 2019; Mughini-Gras et al., 2018). Once the sequenced isolates are subtyped, frequency-matching methods can be used. However, WGS data also bring the possibility of new approaches to link sources with cases, as exemplified in the next sections.

Population genetic methods. Population genetic methods are a powerful tool with applications to source attribution inference utilizing WGS data. These methods model the organism's evolutionary history and have been extended for source attribution applications for *Salmonella* and *Campylobacter* (Barco et al., 2015; Mughini-Gras et al., 2014; Wilson et al., 2008). One example is the asymmetric island model, which models the DNA sequence evolution and zoonotic transmission of the pathogen (Wilson et al., 2008). Each source is considered a population of pathogens, that is, an island.

Pathogens can migrate among populations and evolve through mutation and recombination. The model estimates migration rate, mutation, and recombination parameters and uses those to assign the probability of each human case isolate having originated from one of the source populations (Wilson et al., 2008). Given that population genetics model the evolution of the pathogen, unique strains (in humans) may be assigned to a source rather than be excluded from the dataset as is necessary for frequency-matching methods.

Another population genetic method for source attribution is STRUCTURE (Jehanne et al., 2020; Mughini-Gras et al., 2021; Mulder et al., 2020; Saif et al., 2022). STRUCTURE uses model-based clustering, which assigns a cluster (population) to each sample, while simultaneously estimating the allele frequency in each population (Pritchard et al., 2000). STRUCTURE assumes that the allele frequency within a population is constant and that the association between different genes is completely random (independent) (Pritchard et al., 2000). In other words, the model associates variability among alleles with population grouping by structuring the samples into clusters.

When extending the application of this model to source attribution, each population (cluster) would be a source, and the human case isolates would be classified among these clusters. The STRUCTURE algorithm can also consider admixture and, as such, the possibility of the introduction of new lineages into a population (or in human cases). The initial algorithm has transformed over the years to address ancestry, dominant marker, and prior information on the groups. In addition, there have been extensions developed to address issues ranging from computational speed to properties of the model, such as considering the spatial distribution of the populations (November, 2016).

However, with the large datasets generated by WGS, the computation time required for these algorithms can be substantial, usually increasing linearly with the number of loci (Pérez-Reche et al., 2020). STRUCTURE, for example, works on short genotypes consisting of, at most, only hundreds of loci, which does not encapsulate all available information of the WGS data.

Many methods have been developed to select markers (features) on the genome that provide higher discrimination among strains and reduce the size of the datasets (Banks et al., 2003; Manel et al., 2002; Pérez-Reche et al., 2020; Storer et al., 2012). These features can be used as input for the source attribution algorithms, resulting in less computational time. Recent work proposed a minimal multilocus distance method to attribute cases to sources, which is fast enough to deal with thousands of loci, while other work suggests a method to select optimal markers from the genotype using information theory (Pérez-Reche et al., 2020).

Novel methods of source attribution. Finding hidden complex patterns through Machine Learning (ML) algorithms usually requires a larger amount of data (James et al., 2022). Therefore, ML algorithms are suitable for analyzing WGS data (Lupolova et al., 2019). ML algorithms applied to source attribution can be either unsupervised or supervised learning techniques (Lupolova et al., 2019). For unsupervised learning, there is no label in the data, and the algorithm will group the data into clusters based on their similarities, also known as clustering methods (James et al., 2022). Each cluster can then be associated with a source.

On the other hand, supervised learning uses labeled data to learn hidden characteristics and patterns to categorize the data based on these labels (sources) (James et al., 2022). The model learns using a training dataset and utilizes a testing dataset (data not seen before) for performance evaluation.

Models are often trained on the isolates with known sources of the pathogen. Then, the final model can be used to predict labels (or sources) for unlabeled data. For optimal model hyperparameter tuning and establishing a more robust and unbiased model development process, it is recommended to perform validation procedures, such as cross-validation (James et al., 2022). This involves partitioning the training dataset into distinct subsets, where some are used to train the model and others for model validation (Lupolova et al., 2019). Features or predictors are variables in the input data mapped to the labels through an empirical relationship learned by the model.

There has been an increase in source-attribution studies using ML and classification algorithms (Lupolova et al., 2019; Lupolova et al., 2017; Munck et al., 2020). Recent analyses with *Salmonella* isolates were developed using supervised learning algorithms such as random forests (RF), logit boost, and support vector machines (SVM) for source-attribution problems, and supervised multiclass classification algorithms such as multinomial logistic regression (MLR) (Duarte et al. 2021; Guillier et al., 2020; Lupolova et al., 2017; Munck et al., 2020; Zhang et al., 2019).

Zhang et al. used genomic data of *Salmonella* Typhimurium from different countries to develop a RF algorithm to attribute sources across animal reservoirs to outbreak cases. Their data consisted of genomes from different countries, spanning a significant period (2007–2013, 2015–2017) and focusing on outbreak data. The final dataset comprised 1473 isolates after removing 744 redundant isolates to avoid bias due to sampling similar strains. From these, 1041 genomes were from animals and used to train the classifier. Their final input data comprised 3137 features—1882 core genome SNPs, 150 quality indels, and 1105 source discriminatory accessory genes. The model predicts four animal reservoirs: poultry, wild birds, bovine, and swine, with an accuracy rate of 82.9%.

Because an ML classifier is restricted to the classes represented in the training data, Zhang et al. added a tool to further classify its prediction as precise or imprecise. Thus, isolates from sources not present in the training data are classified as one of the sources included in the model, but its prediction may be identified as imprecise. In addition, they build an extra RF classifier with humans as a source to compare their results with Lupolova et al.'s (2017) work (Wheeler, 2019; Zhang et al., 2019).

They found that only 36.96% of human cases were assigned to humans, as opposed to 90% of Lupolova et al.'s studies. This was due to Lupolova et al. having closely related human isolates (around 85% shared their most recent common ancestor) in their training dataset, resulting in the high accuracy of human host prediction. Given Zhang et al. removed redundant isolates, their data have only 36.9% of human isolates sharing the same most recent common ancestor.

Lupolova et al. used SVM to predict the isolation host for each genome and analyze host specificity. They want to determine whether genetic content can discriminate among interspecies transmission. Their genome data span an extended range of years (1945–2016) and countries and contain different serovars: *Salmonella* Typhimurium (human, bovine, swine, and poultry), *Salmonella* Typhi (human), and *Salmonella* Dublin (human and bovine). They build an SVM model for each host with final input data of protein variants as features and hosts as labels. Thus, an isolate could be assigned for multiple sources—making it a generalist strain.

However, the majority (94%) was assigned to only one host. The model misclassified some isolates; this could be because the data do not incorporate all the genetic features of a source or the strain is transient between hosts. The final model predictions were highly accurate (ranging from 67% to 90%) (Lupolova et al., 2017). These two studies highlight the importance of model building, feature selection, and appropriate data processing when dealing with ML models and WGS data. It is possible to achieve different conclusions by following other procedures.

Recent work by Munck et al. developed a boosting algorithm (logit boost) to classify sporadic human cases of *Salmonella* Typhimurium in Denmark among the following sources: Broilers, layers, cattle (domestic), cattle (import), ducks (import), pigs (domestic), and pigs (import). They used human, food, and animal isolates collected from an integrated surveillance system in Denmark over 2 years to ensure the data represent all true sources. Their input data consisted of cgMLST, which was further reduced to only 17 loci using feature selection techniques.

All sources' isolates were correctly predicted, except for 38% of domestic pigs and 27% of imported pigs, which were wrong classified as poultry. Their final model accuracy was 92%. Of all human sporadic cases, 81% were attributed. The human cases not attributed were either infected from a source not in the training dataset or a strain not captured in the training data. They compared their model against the Bayesian Hald model (Hald et al., 2004) in the same dataset. The input data for the Hald model were the isolates' multilocus variable-number tandem-repeat analysis (MLVA) profile and resistance profile. The results were similar, but only 49% of human cases were attributed (Munck et al., 2020). Both models draw similar conclusions regarding the sources, corroborating ML as a new, robust, and efficient tool for source attribution.

Guillier et al. developed an MLR, an extension of logistic regression to allow multiclass classification to predict the source of environmental strains of *Salmonella* Typhimurium and its monophasic variant. Ninety-eight bacterial isolates were collected from 2010 to 2015; 69 were from animals (pigs, poultry, and ruminants) and 19 were from the environment (no source). They first calculated the accessory genes (noncore genome) enriched in each source to use as input data in the MLR. Then, they use Aikake information criteria to decide which accessory genes to include as features in the final model (eight genes). The chosen model had an accuracy of 74% (Guillier et al., 2020). Table 1 summarizes the main properties of each model.

A study compared three ML source attribution models, SVM, RF, and neural networks (NN), on the same dataset of *Salmonella* Typhimurium. All models arrived at similar results with similar accuracy (75–90%). RF is the most user-friendly, can predict multiple classes at once, and provides a list of the most relevant features. NN is highly scalable and can predict various classes; however, it requires technical knowledge. In summary, any of these models effectively attributes the source for *Salmonella* Typhimurium (Lupolova et al., 2019).

It is possible to expand source attribution even further. Recent work developed a source attribution model based on hierarchical clustering to rapidly identify and trace salmonellosis' geographical sources, rather than points in the food chain, from WGS data (Bayliss et al., 2023).

It is noteworthy to mention the work by Arning et al. (2021), which provides a comparison of performances of different ML algorithms for attributing sources of campylobacteriosis cases. In summary, they identified the best-performing ML algorithms for different resolutions of sequence data: multilocus sequence typing (MLST), cgMLST, and WGS. They tested 14 supervised learning algorithms, ranging from simple learners such as K-nearest neighbors, decision tree-based algorithms to deep learning algorithms such as NN, and the asymmetric island model, iSource. They found ML outperforms iSource.

In addition, some studies have applied weighted network analyses, a clustering method, to perform source attribution. In this methodology, each node in the network is an isolate, and links between isolates represent their genetic distance. Isolates from the same sources would then be clustered together. It has been found that this method remains robust independent of the resolution of WGS data used, whether SNP, cgMLST, or wgMLST (Merlotti et al., 2020; Wainaina et al., 2022).

There are still challenges to applying source attribution to WGS data. Following, we cover some of them:

Table 1. Summary of Properties of Each Machine Learning Source Attribution Model Using Whole-Genome Sequencing: Base Model, Data Collection, Input Features, Sources (Labels), Percentage of Attributed Human Cases, Comparison with Other Models (if Existent), and *Salmonella* Serotype

Author	Munck et al.	Zhang et al.	Lupolova et al.	Guillier et al.
Models	Logit Boost	Random Forest	Support Vector Machines	Multinomial logistic regression
Data collection	Human, food, and animal isolates collected from integrated surveillance system in Denmark	Outbreak data on human, food, environment, wild and livestock animal from different countries	Human and animal isolates over many countries and ranging from 1945 to 2016.	98 Bacterial isolates collected over 2010–2015 (19 without label)
Input Features	cgMLST (17 loci after feature reduction)	Core genome SNPs, high quality indels, and source discriminatory accessory genes	Protein variants (up to 1000–1500)	Eight accessory genes (noncore genome)
Sources	Broilers, layers, cattle, cattle (import), ducks (import), pigs, pigs (import)	Bovine, swine, wild bird, and poultry	Isolation host: avian, bovine, human, swine	Pigs, poultry, and ruminants
Attribution of cases	81%	42%	Not applicable	25 Out of 29 environmental strains
Accuracy	92%	83%	67–90%	74%
Model comparison	Hald model using MLVA profile and resistance profile: fit of 0.9 and 49% attribution of human cases	Not applicable	Not applicable	Not applicable
Serovar	*Salmonella* Typhimurium	*Salmonella* Typhimurium	Multiple serovars (Typhi, Dublin, Typhimurium)	*Salmonella enterica* Typhimurium and *Salmonella enterica* 1,4,[5],12:i:-

cgMLST, core-genome multilocus sequence typing; MLVA, multilocus variable-number tandem-repeat analysis; SNP, single nucleotide polymorphism.

(1) A well-known issue is unique strains in unlabelled data. For Bayesian models, subtypes not included in the sources can be removed. ML models can only classify strains that are in the training data so predictions can further be labeled as precise or imprecise to ensure accurate classification (Munck et al., 2020; Zhang et al., 2019). Unique isolates may be assigned to an existent source for population genetics, given that pathogen evolution is considered in the model.

(2) Another common issue is when some sources are poorly sampled, which could generate incorrect predictions. A potential solution is to upsample or downsample the dataset (Lupolova et al., 2019; Munck et al., 2020). Moreover, for WGS data, it is essential to remove redundant genomes by analyzing some genetic features, such as the number of SNPs separating each isolate (Zhang et al., 2019), to avoid overinflating model accuracy (such as similar strains from outbreak data in a population-level study).

(3) Predictions need to be adjusted for ''unknown sources'' not included in the data, for example, by allowing the classification into an extra source using isolates from other sources (environment) to inform the ''unknown source'' or adding a tool to identify imprecise predictions (Zhang et al., 2019).

All the above highlight the need for having a robust sampling process of the true sources.

Comparative exposure assessment

Comparative exposure assessment is a microbiological methodology that focuses on the point of exposure and transmission routes rather than animal reservoirs. There have been studies that apply comparative exposure assessment to attribute sources of exposure for some foodborne pathogens such as *Salmonella* (Christidis et al., 2020; Fajardo-Guerrero et al., 2020) and *Campylobacter* (Evers et al., 2008; Pintar et al., 2017). The comparative exposure assessment estimates the average number of pathogens that individuals in a population are exposed to in each source and route per day.

The exposure, E, is defined as the average number of organisms that individuals are exposed to in a day (units of cells/person/day) for a pathway of a specific source. It can be formulated as following:

$$\mathrm{E} = \mathrm{f} \times \mathrm{i} \times \mathrm{p} \times \mathrm{c},$$

where *f* is the frequency of ingestion events (events/day), *i* is the total mass (or volume) consumed per individual per event (mass/event/person), *p* is the probability that the ingested item is contaminated with the pathogen, and *c* is the concentration of pathogen cells per mass (volume) in the ingested item, given it is contaminated (cells/mass) (Christidis et al., 2020).

Exposure is estimated separately for all relevant transmission routes within the categories of food, animal contact, and environment (EFSA, 2008). For each transmission route, adaptations to the calculation of each component of the exposure equation may need to be implemented. For example, for food contamination, an extra term indicating raw, undercooked, and cooked consumption may be included. Comparative exposure assessment estimates the relative contribution of each transmission route to the population's total exposure, which is directly related to the likely sources of cases of human illness (Pintar et al., 2017; Ravel et al., 2017). In this way, one can assess which sources, transmission pathways, and points along the pathway have a larger risk for the population and implement interventions to decrease this risk.

However, the possibility of cross-contamination and different transmission routes make directly linking the point of exposure to animal reservoirs more difficult. There are many techniques to achieve this linkage or provide a better understanding of possible routes. For instance, a meta-analysis combined results from attribution studies across reservoirs and transmission routes and estimated attribution proportions for the transmission pathways (Mughini-Gras et al., 2022). A more precise estimate is possible by either combining frequency-matching methods with comparative exposure assessment or case–control studies, allowing for the control of exposure when estimating the frequency of cases in each source.

This generates a complete picture of the transmission pathways from the point of production to exposure, which can better inform risk management in the prioritization of control strategies for each transmission route (EFSA, 2008; Hurst et al., 2023; Mughini-Gras and van Pelt, 2014; Mughini-Gras et al., 2019; Mughini-Gras et al., 2018; Mughini-Gras et al., 2014; Ravel et al., 2017). A study by Mughini-Gras et al. combined multiple microbial subtyping frequency-matching methods with a comparative exposure assessment to estimate the contribution of each point of exposure to salmonellosis. They incorporated an exposure term in the calculation of λ_{is}. For the Hald model, they had,

$$\lambda_{is} = m_s c_s p_{is} q_i a_s,$$

where m_s represents the consumption of source s, c_s is the probability of the source being eaten raw/undercooked, p_{is} is the proportion of subtype i in source s, q_i is the subtype-dependent factor, and a_s is the food source-dependent factor, as previously defined. For the Dutch model, controlling the consumption of the source without considering the probability of eating raw/undercooked food led to pig as the highest contributing source and table eggs as second, which is inconsistent with common knowledge in *Salmonella* epidemiology. Thus, it is necessary to consider the consumption weight and the likelihood of the food being undercooked to properly estimate the contribution of each source when using the Dutch model. The Hald model grants expected results regardless of the inclusion of food consumption data.

Impact on Salmonellosis

In the case of *Salmonella*, efforts to reduce incidence have shown positive results. In 2019, it was estimated that illnesses caused by *Salmonella* in Canada decreased by more than 25,000 cases relative to the previous 5 years (Glass-Kaastra et al., 2022). Successful source attribution through genomic-based surveillance contributed to the implementation of new effective, targeted interventions, driving the reduction of cases (Glass-Kaastra et al., 2022; Morton et al., 2019). WGS source attribution implementation generated a more accurate and specific linkage to products, providing the evidence needed for new control requirements (Morton et al., 2019). The implementation of WGS in the United States has prevented around 25,000 cases of foodborne illness, saving around 500 million U.S. dollars (Brown et al., 2021; Glass-Kaastra et al., 2022). Work on case studies on WGS implementation across Europe and America found that the benefits of WGS outweigh the cost (Alleweldt et al., 2021).

In brief, the implementation of WGS provides better accuracy, more specificity on outbreak linkages, generate better evidence to inform control policy, and improve understanding of disease transmission. Recent work by Hurst et al. (2023) shows a decline in the percentage of cases attributed to chicken breasts by one-third from 2015 to 2019 and in the incidence rate of salmonellosis by one-third in the same period in Canada. However, despite the observed reduction in cases, the incidence of *Salmonella* infections remains high, with an estimated 70,833 cases of illness in Canada in 2019 (Glass-Kaastra et al., 2022), highlighting that further efforts are needed.

Conclusion

Source attribution methods have been extensively applied to identify transmission routes and animal reservoirs of foodborne pathogens such as *Salmonella*. Frequency-matching approaches have been widely utilized for microbial subtyped data to estimate the probability of a human case originating from an animal reservoir. The growth of WGS and its popularization for source attribution studies has increased the development and application of novel methods. WGS provides high-resolution power to discriminate isolates, which can increase the accuracy of frequency-matching approaches. Evolutionary and population genetics algorithms may also be used to link sources to human case isolates. The large size of WGS data further allows for introducing ML classification methods in source attribution, where each source is a class. This work summarizes well-known source attribution methods and novel methods.

However, there are still challenges to overcome, such as the computational efficiency of these methods, given the large data size. To address this issue, one may select a few features (genetic markers) with high discrimination power among sources and reduce the input data size. Another well-known problem is that available data often lack complete information about the various sources, which highlights the importance of having a solid and integrated surveillance system encompassing all one health spheres—animal, human, and environmental (Mather et al., 2015; Mughini-Gras et al., 2019; Pires et al., 2014). The field of source attribution is still evolving, with new methods arising, which improve on the older ones. Furthermore, the richness of WGS data has not yet been fully utilized, although progress is being made.

Including WGS data in source attribution can provide better evidence to inform policy development and prioritize intervention strategies to control salmonellosis. In addition, they help better understand the complex interactions of pathogens with animals, humans, and the environment, such as determining genetic features responsible for host specificity and adaptability and geographical distribution of *Salmonella*. Therefore, continued improvement, development, and generalization of source attribution methods are essential to advance our understanding and control of *Salmonella* transmission.

Acknowledgment

The authors would like to acknowledge Samantha Hoch for performing a literature review search and providing a selection of relevant articles for this work.

Authors' Contributions

R.C.F.: Conceptualization, writing—original draft preparation, and writing—review and editing. M.R.E.: Writing—original draft preparation and writing—review and editing. M.H.: Writing—review and editing. E.F.: Writing—review and editing and funding acquisition. M.O.: Conceptualization and writing—review and editing.

Disclosure Statement

The authors have no conflicts of interest to declare.

Funding Information

This work was supported by funds from Genome Canada, and Genome BC as part of the PeptAid (291PEP) project.

References

Alleweldt F, Kara Ş, Best K, et al. Economic evaluation of whole genome sequencing for pathogen identification and surveillance—Results of case studies in Europe and the Americas 2016 to 2019. Euro Surveill 2021;26(9):1900606; doi: 10.2807/1560-7917.es.2021.26.9.1900606

Anderson LA, Miller DA, Trampel DW. Epidemiological investigation, cleanup, and eradication of pullorum disease in adult chickens and ducks in two small-farm flocks. Avian Dis 2006;50(1):142–147; doi: 10.1637/7397-062105R.1

Arning N, Sheppard SK, Bayliss S, et al. Machine learning to predict the source of campylobacteriosis using whole genome data. PLoS Genet 2021;17(10):e1009436; doi: 10.1371/journal.pgen.1009436

Banks MA, Eichert W, Olsen JB. Which genetic loci have greater population assignment power? Bioinformatics 2003; 19:1436–1438; doi: 10.1093/bioinformatics/btg172

Barco L, Barrucci F, Cortini E, et al. Ascertaining the relationship between *Salmonella* Typhimurium and *Salmonella* 4,[5], 12: I:-by MLVA and inferring the sources of human salmonellosis due to the two serovars in Italy. Front Microbiol 2015;6:301; doi: 10.3389/fmicb.2015.00301

Barco L, Barrucci F, Olsen JE, et al. *Salmonella* source attribution based on microbial subtyping. Int J Food Microbiol 2013;163:193–203; doi: 10.1016/j.ijfoodmicro.2013.03.005

Bayliss SC, Locke RK, Jenkins C, et al. Rapid geographical source attribution of *Salmonella enterica* serovar Enteritidis genomes using hierarchical machine learning. eLife 2023;12: e84167; doi: 10.7554/elife.84167

Bearson S. *Salmonella* in swine: Prevalence, multidrug resistance, and vaccination strategies. Annu Rev Anim Biosci 2022; 10:373–393; doi: 10.1146/annurev-animal-013120-043304

Beuchat LR. Ecological factors influencing survival and growth of human pathogens on raw fruits and vegetables. Microbes Infect 2002;4(4):413–423; doi: 10.1016/s1286-4579(02)01555-1

Boysen L, Rosenquist H, Larsson JT, et al. Source attribution of human campylobacteriosis in Denmark. Epidemiol Infect 2014;142:1599–1608; doi: 10.1017/s0950268813002719

Brenner FW, Villar RG, Angulo FJ, et al. *Salmonella* nomenclature. J Clin Microbiol 2000;38(7):2465–2467; doi: 10.1128/JCM.38.7.2465-2467.2000

Brown B, Allard M, Bazaco MC, et al. An economic evaluation of the whole genome sequencing source tracking program in the U.S. PLoS One 2021;16(10):e0258262; doi: 10.1371/journal.pone.0258262

Carstens CK, Salazar JK, Darkoh C. Multistate outbreaks of foodborne illness in the United States associated with fresh produce from 2010 to 2017. Front Microbiol 2019;10:2667; doi: 10.3389/fmicb.2019.02667

Chousalkar K, Gast R, Martelli F, et al. Review of egg-related salmonellosis and reduction strategies in United States, Australia, United Kingdom and New Zealand. Crit Rev Microbiol 2018;44:290–303; doi: 10.1080/1040841x.2017.1368998

Christidis T, Hurst M, Rudnick W, et al. A comparative exposure assessment of foodborne, animal contact and waterborne transmission routes of *Salmonella* in Canada. Food Control 2020;109:106899; doi: 10.1016/j.foodcont.2019.106899

Collineau L, Boerlin P, Carson CA, et al. Integrating whole-genome sequencing data into quantitative risk assessment of foodborne antimicrobial resistance: A review of opportunities and challenges. Front Microbiol 2019;10:1107; doi: 10.3389/fmicb.2019.01107

Critzer FJ, Doyle MP. Microbial ecology of foodborne pathogens associated with produce. Curr Opin Biotechnol 2010; 21(2):125–130; doi: 10.1016/j.copbio.2010.01.006

David JM, Sanders P, Bemrah N, et al. Attribution of the French human salmonellosis cases to the main food-sources according to the type of surveillance data. Prev Vet Med 2013; 110:12–27; doi: 10.1016/j.prevetmed.2013.02.002

De Knegt LV, Pires SM, Hald T. Attributing foodborne salmonellosis in humans to animal reservoirs in the European Union using a multi-country stochastic model. Epidemiol Infect 2015; 143:1175–1186; doi: 10.1017/s0950268814001903

Destoumieux-Garzón D, Mavingui P, Boetsch G, et al. The One Health concept: 10 Years old and a long road ahead. Front Vet Sci 2018;5:14; doi: 10.3389/fvets.2018.00014

Dórea FC, Cole DJ, Hofacre C, et al. Effect of *Salmonella* vaccination of breeder chickens on contamination of broiler chicken carcasses in integrated poultry operations. Appl Environ Microbiol 2010;76:7820–7825; doi: 10.1128/aem.01320-10

Duarte ASR, Röder T, Van Gompel L, et al. Metagenomics-based approach to source-attribution of antimicrobial resistance determinants—Identification of reservoir resistome signatures. Front Microbiol 2021;11:601407; doi: 10.3389/fmicb.2020.601407

Dunn LL, Sharma V, Chapin TK, et al. The prevalence and concentration of *Salmonella enterica* in poultry litter in the Southern United States. PLoS One 2022;17:e0268231; doi: 10.1371/journal.pone.0268231

Eng S-K, Pusparajah P, Mutalib N-SA, et al. *Salmonella*: A review on pathogenesis, epidemiology and antibiotic resistance. Front Life Sci 2015;8:284–293; doi: 10.1080/21553769.2015.1051243

European Food Safety Authority (EFSA). Overview of methods for source attribution for human illness from food-borne microbiological hazards—Scientific Opinion of the Panel on Biological Hazards. EFSA J 2008;6:764.

Evers E, Van Der Fels-Klerx H, Nauta M, et al. *Campylobacter* source attribution by exposure assessment. Int J Risk Assess Manage 2008;8:174–190; doi: 10.1504/ijram.2008.016151

Fajardo-Guerrero M, Rojas-Quintero C, Chamorro-Tobar I, et al. Exposure assessment of *Salmonella* spp. in fresh pork meat from two abattoirs in Colombia. Food Sci Technol Int 2020;26:21–27; doi: 10.1177/1082013219864746

Ferrari RG, Panzenhagen PH, Conte-Junior CA. Phenotypic and genotypic eligible methods for *Salmonella* Typhimurium source tracking. Front Microbiol 2017;8:2587; doi: 10.3389/fmicb.2017.02587

Ferrari RG, Rosario DKA, Cunha-Neto A, et al. Worldwide epidemiology of *Salmonella* serovars in animal-based foods: A meta-analysis. Appl Environ Microbiol 2019;85(14): e00591-19; doi: 10.1128/AEM.00591-19

Glass-Kaastra S, Dougherty B, Nesbitt A, et al. Estimated reduction in the burden of nontyphoidal *Salmonella* illness in Canada Circa 2019. Foodborne Pathog Dis 2022;19:744–749; doi: 10.1089/fpd.2022.0045

Government of Canada. National Enteric Surveillance Program (NESP) Annual Summary 2019. Public Health Agency of Canada: Guelph; 2020.

Guillier L, Gourmelon M, Lozach S, et al. AB_SA: Accessory genes-based source attribution—Tracing the source of *Salmonella enterica* Typhimurium environmental strains. Microb Genom 2020;6:mgen000366; doi: 10.1099/mgen.0.000366

Guo C, Hoekstra RM, Schroeder CM, et al. Application of Bayesian techniques to model the burden of human salmonellosis attributable to U.S. food commodities at the point of processing: Adaptation of a Danish model. 2011;8(4):509–516; doi: 10.1089/fpd.2010.0714

Hald T, Lo Fo Wong DMA, Aarestrup FM. The attribution of human infections with antimicrobial resistant *Salmonella* bacteria in Denmark to sources of animal origin. Foodborne Pathog Dis 2007;4:313–326; doi: 10.1089/fpd.2007.0002

Hald T, Vose D, Wegener HC, et al. A Bayesian approach to quantify the contribution of animal-food sources to human salmonellosis. Risk Anal 2004;24(1):255–269; doi: 10.1111/j.0272-4332.2004.00427.x

Hoelzer K, Switt AIM, Wiedmann M. Animal contact as a source of human non-typhoidal salmonellosis. Vet Res 2011; 42:1–28; doi: 10.1186/1297-9716-42-34

Holden N, Pritchard L, Toth I. Colonization outwith the colon: Plants as an alternative environmental reservoir for human pathogenic enterobacteria. FEMS Microbiol Rev 2009;33(4): 689–703; doi: 10.1111/j.1574-6976.2008.00153.x

Holschbach CL, Peek SF. *Salmonella* in dairy cattle. Vet Clin North Am Food Anim Pract 2018;34(1):133–154; doi: 10.1016/j.cvfa.2017.10.005

Hurst M, Nesbitt A, Kadykalo S, et al. Attributing salmonellosis cases to foodborne, animal contact and waterborne routes using the microbial subtyping approach and exposure weights. Food Control 2023;148:109636; doi: 10.1016/j.foodcont.2023.109636

James G, Witten D, Hastie T, et al. An introduction to statistical learning with applications in R. Second Edition. Springer publication: New York; 2022.

Jechalke S, Schierstaedt J, Becker M, et al. *Salmonella* establishment in agricultural soil and colonization of crop plants depend on soil type and plant species. Front Microbiol 2019; 10:967; doi: 10.3389/fmicb.2019.00967

Jehanne Q, Pascoe B, Bénéjat L, et al. Genome-wide identification of host-segregating single-nucleotide polymorphisms for source attribution of clinical *Campylobacter coli* isolates. Appl Environ Microbiol 2020;86:e01787-20; doi: 10.1128/aem.01787-20

Kirk MD, Pires SM, Black RE, et al. World Health Organization estimates of the global and regional disease burden of 22 foodborne bacterial, protozoal, and viral diseases, 2010: A data synthesis. PLoS Med 2015;12:e1001921; doi: 10.1371/journal.pmed.1001921

Laughlin M, Bottichio L, Weiss J, et al. Multistate outbreak of *Salmonella* Poona infections associated with imported cucumbers, 2015–2016. Epidemiol Infect 2019;147:e270; doi: 10.1017/s0950268819001596

Lupolova N, Dallman TJ, Holden NJ, et al. Patchy promiscuity: Machine learning applied to predict the host specificity of *Salmonella enterica* and *Escherichia coli.* Microb Genom 2017;3(10):e000135; doi: 10.1099/mgen.0.000135

Lupolova N, Lycett SJ, Gally DL. A guide to machine learning for bacterial host attribution using genome sequence data. Microb Genom 2019;5(12):e000317; doi: 10.1099/mgen.0.000317

Manel S, Berthier P, Luikart G. Detecting wildlife poaching: Identifying the origin of individuals with Bayesian assignment tests and multilocus genotypes. Conserv Biol 2002;16: 650; doi: 10.1046/j.1523-1739.2002.00576.x

Mather AE, Vaughan TG, French NP. Molecular approaches to understanding transmission and source attribution in nontyphoidal *Salmonella* and their application in Africa. Clin Infect Dis 2015;61(Suppl 4):S259–S265; doi: 10.1093/cid/civ727

McLure A, Shadbolt C, Desmarchelier PM, et al. Source attribution of salmonellosis by time and geography in New South Wales, Australia. BMC Infect Dis 2022;22:1–13; doi: 10.1186/s12879-021-06950-7

Merlotti A, Manfreda G, Munck N, et al. Network approach to source attribution of *Salmonella enterica* serovar Typhimurium and its monophasic variant. Front Microbiol 2020;11: 1205; doi: 10.3389/fmicb.2020.01205

Miller P, Marshall J, French N, et al. SourceR: Classification and source attribution of infectious agents among heterogeneous populations. PLoS Comput Biol 2017;13(5):e1005564; doi: 10.1371/journal.pcbi.1005564

Morton VK, Kearney A, Coleman S, et al. Outbreaks of *Salmonella* illness associated with frozen raw breaded chicken products in Canada, 2015–2019. Epidemiol Infect 2019;147: e254; doi: 10.1017/s0950268819001432

Mughini-Gras L, Benincà E, McDonald SA, et al. A statistical modelling approach for source attribution meta-analysis of sporadic infection with foodborne pathogens. Zoonoses Public Health 2022;69(5):475–486; doi: 10.1111/zph.12937

Mughini-Gras L, Kooh P, Augustin J-C, et al. Source attribution of foodborne diseases: Potentialities, hurdles, and future expectations. Front Microbiol 2018;9:1983; doi: 10.3389/fmicb.2018.01983

Mughini-Gras L, Kooh P, Fravalo P, et al. Critical orientation in the jungle of currently available methods and types of data for source attribution of foodborne diseases. Front Microbiol 2019;10:2578; doi: 10.3389/fmicb.2019.02578

Mughini-Gras L, Pijnacker R, Coipan C, et al. Sources and transmission routes of campylobacteriosis: A combined analysis of genome and exposure data. J Infect 2021;82:216–226; doi: 10.1016/j.jinf.2020.09.039

Mughini-Gras L, Smid J, Enserink R, et al. Tracing the sources of human salmonellosis: A multi-model comparison of phenotyping and genotyping methods. Infect Genet Evol 2014; 28:251–260; doi: 10.1016/j.meegid.2014.10.003

Mughini-Gras L, van Pelt W. *Salmonella* source attribution based on microbial subtyping: Does including data on food consumption matter? Int J Food Microbiol 2014;191:109–115; doi: 10.1016/j.ijfoodmicro.2014.09.010

Mulder AC, Franz E, de Rijk S, et al. Tracing the animal sources of surface water contamination with *Campylobacter Jejuni* and *Campylobacter coli.* Water Res 2020;187:116421; doi: 10.1016/j.watres.2020.116421

Mullner P, Jones G, Noble A, et al. Source attribution of foodborne zoonoses in New Zealand: A modified Hald model. Risk Anal 2009;29:970–984; doi: 10.1111/j.1539-6924.2009.01224.x

Munck N, Njage PMK, Leekitcharoenphon P, et al. Application of whole-genome sequences and machine learning in source attribution of *Salmonella* Typhimurium. Risk Anal 2020;40: 1693–1705; doi: 10.1111/risa.13510

Murray CE, Varga C, Ouckama R, et al. Temporal Study of *Salmonella enterica* serovars isolated from environmental samples from Ontario poultry breeder flocks between 2009 and 2018. Pathogens 2023;12:278; doi: 10.3390/pathogens12020278

Novembre J. Pritchard, Stephens, and Donnelly on population structure. Genetics 2016;204(2):391–393; doi: 10.1534/genetics.116.195164

Pérez-Reche FJ, Rotariu O, Lopes BS, et al. Mining whole genome sequence data to efficiently attribute individuals to source populations. Sci Rep 2020;10(1):12124; doi: 10.1038/s41598-020-68740-6

Pintar KD, Thomas KM, Christidis T, et al. A comparative exposure assessment of *Campylobacter* in Ontario, Canada. Risk Anal 2017;37(4):677–715; doi: 10.1111/risa.12653

Pires SM, Evers EG, van Pelt W, et al. Attributing the human disease burden of foodborne infections to specific sources. Foodborne Pathog Dis 2009;6(4):417–424; doi: 10.1089/fpd.2008.0208

Pires SM, Vieira AR, Hald T, et al. Source attribution of human salmonellosis: An overview of methods and estimates. Foodborne Pathog Dis 2014;11(9):667–676; doi: 10.1089/fpd.2014.1744

Popa GL, Papa MI. *Salmonella* spp. infection—A continuous threat worldwide. Germs 2021;11(1):88–96; doi: 10.18683/germs.2021.1244

Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics 2000; 155(2):945–959; doi: 10.1093/genetics/155.2.945

Ravel A, Hurst M, Petrica N, et al. Source attribution of human campylobacteriosis at the point of exposure by combining comparative exposure assessment and subtype comparison based on comparative genomic fingerprinting. PLoS One 2017;12:e0183790; doi: 10.1371/journal.pone.0183790

Rodrigues da Costa M, Pessoa J, Meemken D, et al. A systematic review on the effectiveness of pre-harvest meat safety interventions in pig herds to control *Salmonella* and other foodborne pathogens. Microorganisms 2021;9(9):1825; doi: 10.3390/microorganisms9091825

Saif NA, Cobo-Díaz JF, Elserafy M, et al. A pilot study revealing host-associated genetic signatures for source attribution of sporadic *Campylobacter jejuni* infection in Egypt. Transbound Emerg Dis 2022;69(4):1847–1861; doi: 10.1111/tbed.14165

Silva C, Calva E, Maloy S. One Health and food-borne disease: *Salmonella* transmission between humans, animals, and plants. Microbiol Spectr 2014;2:1–2; doi: 10.1128/microbiolspec.oh-0020-2013

Storer CG, Pascal CE, Roberts SB, et al. Rank and order: Evaluating the performance of SNPs for individual assignment in a non-model organism. PLoS One 2012;7(11): e49018; doi: 10.1371/journal.pone.0049018

Thames HT, Theradiyil Sukumaran A. A review of *Salmonella* and *Campylobacter* in broiler meat: Emerging challenges and food safety measures. Foods 2020;9(6):776; doi: 10.3390/foods9060776

Trampel DW, Holder TG, Gast RK. Integrated farm management to prevent *Salmonella enteritidis* contamination of eggs. J Appl Poult Res 2014;23:353–365; doi: 10.3382/japr.2014-00944

Totton SC, Farrar AM, Wilkins W, et al. A systematic review and meta-analysis of the effectiveness of biosecurity and vaccination in reducing *Salmonella* spp. in broiler chickens. Food Res Int 2012;45:617–627; doi: 10.1016/j.foodres.2011.09.005

Taylor M, Cox W, Otterstatter M, et al. Evaluation of agricultural interventions on human and poultry-related *Salmonella enteritidis* in British Columbia. Foodborne Pathog Dis 2018;15; doi: 10.1089/fpd.2017.2302

Vieira AR, Grass J, Fedorka-Cray PJ, et al. Attribution of *Salmonella enterica* serotype Hadar infections using antimicrobial resistance data from two points in the food supply system. Epidemiol Infect 2016;144:1983–1990; doi: 10.1017/s0950268816000066

Wagenaar JA, French NP, Havelaar AH. Preventing *Campylobacter* at the source: Why is it so difficult? Clin Infect Dis 2013;57(11):1600–1606; doi: 10.1093/cid/cit555

Wainaina L, Merlotti A, Remondini D, et al. Source attribution of human campylobacteriosis using whole-genome sequencing data and network analysis. Pathogens 2022;11:645; doi: 10.3390/pathogens11060645

Wessels K, Rip D, Gouws P. *Salmonella* in chicken meat: Consumption, outbreaks, characteristics, current control methods and the potential of bacteriophage use. Foods 2021; 10(8):1742; doi: 10.3390/foods10081742

Wheeler NE. Tracing outbreaks with machine learning. Nat Rev Microbiol 2019;17(5):269; doi: 10.1038/s41579-019-0153-1

Wiedmann M. Subtyping of bacterial foodborne pathogens. Nutr Rev 2002;60(7 Pt 1):201–208; doi: 10.1301/00296640260184273

Wilson DJ, Gabriel E, Leatherbarrow AJ, et al. Tracing the source of campylobacteriosis. PLoS Genet 2008;4:e1000203; doi: 10.1371/journal.pgen.1000203

World Health Organization. *Salmonella* (Non-Typhoidal). 2018. Available from: https://www.who.int/news-room/fact-sheets/detail/salmonella-(non-typhoidal) [Last accessed: May 27, 2023].

Yan SS, Pendrak ML, Abela-Ridder B, et al. An overview of *Salmonella* typing: Public health perspectives. Clin Appl Immunol Rev 2004;4:189–204; doi: 10.1016/j.cair.2003.11.002

Zhang S, Li S, Gu W, et al. Zoonotic source attribution of *Salmonella enterica* serotype Typhimurium using genomic surveillance data, United States. Emerg Infect Dis 2019; 25(1):82–91; doi: 10.3201/eid2501.180835

Address correspondence to:
Rebeca Cardim Falcao, PhD
British Columbia Centre for Disease Control
655 West 12th Avenue
Vancouver BC V5Z 4R4
Canada

E-mail: rebeca.falcao@bccdc.ca