International Journal of Food Microbiology 427 (2025) 110953

Contents lists available at ScienceDirect

International Journal of Food Microbiology

journal homepage: www.elsevier.com/locate/ijfoodmicro

Source attribution of *Listeria monocytogenes* in the Netherlands

Lapo Mughini-Gras [a,b,*], Julian A. Paganini [b], Ruoshui Guo [b], Claudia E. Coipan [a], Ingrid H.M. Friesema [a], Angela H.A.M. van Hoek [a], Maaike van den Beld [a], Sjoerd Kuiling [a], Indra Bergval [a], Bart Wullings [c], Menno van der Voort [c], Eelco Franz [a], Timothy J. Dallman [b]

[a] Centre for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM), Bilthoven, Netherlands
[b] Institute for Risk Assessment Sciences (IRAS), Utrecht University, Utrecht, Netherlands
[c] Wageningen Food Safety Research (WFSR), Wageningen, Netherlands

ARTICLE INFO

Keywords:
Human listeriosis
Source tracing
Core-genome MLST
Transmission routes
Risk factors

ABSTRACT

The aim of this study was to determine the relative contributions of various potential food sources of human listeriosis and to identify source-specific risk factors, at exposure level, for human *Listeria monocytogenes* (*Lm*) infection. To achieve this, available *Lm* isolates from human cases ($n = 756$) and food/animal sources ($n = 950$) from national surveillance systems in the Netherlands (2010−2020) were whole genome sequenced. Additionally, questionnaire-based exposure data for human cases was collected. Source attribution analysis was performed using a Random Forest model based on core-genome multilocus sequence typing (cgMLST). Risk factors for human *Lm* infection of cattle, chicken and seafood origin were determined using beta regression analysis on the cgMLST-based attribution estimates. Results indicated that the 756 human *Lm* isolates were mainly attributed to cattle (62.3 %), chicken (19.4 %), and seafood (16.9 %). Specifically, fresh meat (86.2 %), including fresh bovine meat (43.7 %) and fresh chicken meat (39.3 %), accounted for most cases. These attributions stemmed from *Lm* contamination of either the food products or their production environments. Consumption of steak tartare and smoked salmon was associated with an increased risk of human *Lm* infections attributed to cattle and seafood, respectively, while no specific risk factors for chicken-borne listeriosis were identified. This study indicated that *Lm* isolates of cattle origin, particularly those from fresh bovine meat and associated production environments, are estimated to be the primary cause of human listeriosis in the Netherlands. This aligns with several other European source attribution studies on *Lm*. Moreover, the identified risk factors for human *Lm* infection from cattle (i.e. steak tartare) and seafood (i.e. smoked salmon) clearly indicated their attributable sources. This joint analysis of core genome and epidemiological data provided novel insights into the origins and transmission pathways of human listeriosis.

1. Introduction

Listeria monocytogenes (*Lm*), the causative agent of listeriosis, is widespread in the environment, including food-processing facilities where it can contaminate various food products, leading to foodborne disease (Allerberger and Wagner, 2010; Buchanan et al., 2017; Ferreira et al., 2014). Besides gastroenteritis, listeriosis may lead to sepsis, meningoencephalitis, abortion, still-birth and perinatal infection (Allerberger and Wagner, 2010; Koopmans et al., 2013). *Lm* most commonly causes (severe) illness among pregnant women and newborns, elderly and immunocompromised people (Friesema et al., 2015; Pohl et al., 2019). Listeriosis incidence shows a stable trend in the European Union (EU), except during the Coronavirus Disease 2019 (COVID-19) pandemic years. Moreover, *Lm* prevalence is generally low in food samples taken at manufacturing and distribution stage for verification of food safety microbiological criteria according to Commission Regulation (EC) 2073/2005 (EFSA and ECDC, 2023). While contamination with *Lm* can occur at any point in the food supply chain, its ability to replicate in refrigerated conditions and to form biofilms on food-processing surfaces makes *Lm* a highly persistent pathogen in the food industry. This is particularly problematic for pre-packed, ready-made and ready-to-eat (RTE) food products (Hurley et al., 2019), as these products are prepared in advance, with no further cooking or preparation step required to kill *Lm* before being consumed.

* Corresponding author at: Centre for Infectious Disease Control, National Institute for Public Health and the Environment (RIVM), Bilthoven, Netherlands.
E-mail addresses: L.MughiniGras@uu.nl, lapo.mughini.gras@rivm.nl (L. Mughini-Gras).

https://doi.org/10.1016/j.ijfoodmicro.2024.110953
Received 21 June 2024; Received in revised form 18 October 2024; Accepted 19 October 2024
Available online 29 October 2024

Whole genome sequencing (WGS) can help elucidate bacterial population structures and identify sources of zoonotic infections by comparing bacterial genomes from human patients with those from potential sources of infection, such as animals, food, and the environment (Franz et al., 2016; Lupolova et al., 2019; Mughini-Gras et al., 2021). In particular, core-genome Multi-Locus Sequence Typing (cgMLST) has been used for source attribution of foodborne pathogens like *Campylobacter* (Mughini-Gras et al., 2021), allowing for fine-scale differentiation of closely related strains as compared to conventional Multi-Locus Sequence Typing (MLST) (Cody et al., 2017; Maiden et al., 2013; Sheppard et al., 2012). Recently, machine-learning (ML) methods have been used on sequence data of *Lm* to perform source attribution of human listeriosis (Castelli et al., 2023; Gu et al., 2023; Tanui et al., 2022). This follows the successful application of ML methods based on sequence data for source attribution of other pathogens, such as *Salmonella* (Lupolova et al., 2017; Munck et al., 2020), *Campylobacter* (Arning et al., 2021) and Shiga toxin-producing *E. coli* (STEC) (Lupolova et al., 2021).

In the Netherlands, WGS has been used as a routine typing method for human and food *Lm* isolates since 2017. A retrospective genomic analysis of human (clinical) and food *Lm* isolates from the Netherlands in 2010–2020 has recently been performed (Coipan et al., 2023). This analysis revealed a high degree of temporal persistence of food, human and mixed (food-human) isolate clusters based on cgMLST, with more than half of the clusters spanning over more than one year and up to 10 years (Coipan et al., 2023). Of the 583 human *Lm* sequences that could be attributed to food sources using the Hamming distance, the largest proportion was closest to *Lm* sequences from bovine products (25 %), followed by fish/shellfish (24 %), chicken (21 %) and small ruminants (12 %) (Coipan et al., 2023). The associations between human and food *Lm* isolates seemed less strong than those between human and specific food producers, suggesting that *Lm* is predominantly an environmental microorganism (Coipan et al., 2023). While the study of Coipan et al. (2023) examined *Lm* genetic clustering, including temporal distribution of clusters and their links to food sources, it primarily provided a descriptive analysis of the molecular epidemiology of *Lm* in the Netherlands, and did not include model-based source attribution analyses.

In the present study, a source attribution analysis of *Lm* was performed using a ML modelling approach based on the same cgMLST data set from the Netherlands between 2010 and 2020 as in Coipan et al. (2023). Moreover, the cgMLST-based attribution results were combined with available exposure (risk factor) data for the attributed human cases. This approach helped identify possible source-specific risk factors for *Lm* infection, providing insights into the potential transmission routes involved. Therefore, this study presents findings from both source attribution and risk factor analyses, complementing the work of Coipan et al. (2023), to which we refer for additional information on the molecular epidemiology of *Lm* in the Netherlands.

2. Methods

2.1. Data sources

In the Netherlands, since 2008, laboratory-confirmed cases of listeriosis have been mandatorily reported for regional public health services, medical practitioners and/or diagnostic laboratories. Notified cases originate from routine diagnostic activities of people seeking medical attention with mainly invasive disease, i.e., the most severe infections occurring in the population. The public health service contacts the patient or their relatives and enquires, using a questionnaire, about underlying health conditions and medicine use, clinical course of listeriosis and exposure to possible risk factors in the 30 days before disease onset. This information is then registered in a national database for infectious disease surveillance. Mothers and their newborns with listeriosis are notified separately, but they are linked to each other in the database and are considered as one case in the analyses. In parallel to the notification, diagnostic laboratories send the respective *Lm* isolates from invasive listeriosis cases to the Netherlands Reference Laboratory for Bacterial Meningitis (NRLBM) on a voluntary basis, which forwards them to the National Institute for Public Health and the Environment (RIVM) for WGS analysis as part of national surveillance activities of human *Lm* infections.

Sampling of animals and food products thereof for *Lm* detection is routinely performed, both randomly and risk-based throughout the food chain, by the Netherlands Food and Consumer Product Safety Authority (NVWA). Food samples collected by the NVWA are tested by Wageningen Food Safety Research (WFSR) using methods equivalent to ISO 11290-1 or 11290–2. For WGS, one colony per food *Lm* isolate, or a smear of colonies in case of a human *Lm* isolate, are used.

In this study, we used the same WGS data set used previously (Coipan et al., 2023). It includes all available *Lm* isolates obtained from human patients in 2010–2020 (n = 756) and from non-human isolates in 2015–2020 (n = 950). The human isolates originated from patients with sepsis (24 %), meningitis (21 %), gastroenteritis (15 %), pneumonia (8 %), neonatal infection (5 %), encephalitis (3 %) and endocarditis (2 %), whereas no clinical data was available for 34 % of patients. Note that the aforementioned percentages do not add up to 100 % because 91 patients had more than one of these conditions. Moreover, 10 % of cases were fatal.

The non-human isolates were obtained mainly from foods of animal origin (92 %), but also plant-based food products (2 %), animal feces (2 %), and food products whose origin could not be determined unambiguously (4 %) (Table 1). Because most non-human isolates originated from food products, for simplicity we refer generically to “food isolates” throughout the manuscript.

2.2. Listeria isolate whole genome sequencing

Since 2017, WGS has been adopted as the standard typing method for all *Lm* isolates collected within the Dutch national surveillance of human listeriosis and monitoring of *Lm* in food products. *Lm* isolates from human cases collected in 2016 have been sequenced retrospectively by the RIVM, while those collected in 2010–2015 have been sequenced within the ELiTE study (European *Listeria* Typing Exercise Extension to Whole Genome Sequencing) led by the European Centre for Disease Prevention and Control (ECDC) (Van Walle et al., 2018). For both human and food *Lm* isolates from 2016 to 2019, WGS was performed by a commercial sequencing company using Illumina HiSeq (2 × 100 bp) or Illumina NovaSeq (2 × 150 bp), and from 2020 onwards by the RIVM using Illumina MiSeq or NextSeq 500/550 (2 × 150 bp). All sequences were subjected to quality control and de novo assembled using an in-house developed pipeline (https://github.com/Papos92/assembly_pipeline). Raw data with phred score >30 and draft genomes with a total length of 2700,00 to 3,230,000 bp, N50 > 10,000 bp, GC% of 37.6 to 38.2 %, and an average read coverage ≥10 were considered for further analysis. Yet, average read coverage was much higher: mean 153 (range 37–267).

Determination of cgMLST (Ruppitsch et al., 2015) profiles was performed in Ridom SeqSphere+ version 5.0.0 (Ridom GmbH, Münster, Germany). All assembled genomes with 98.1–100 % of loci identified (i.e., <33 loci missing), were considered for further analyses.

2.3. Source categorization

Through the labelling information available in the metadata of food samples, food *Lm* isolates were categorized in three ways: 1) Source 1 (S1) denoting the host in question, i.e. cattle, chicken, game (i.e., wild animals), pig, plants (vegetables, fruit, spices and herbs), small ruminants (goats and sheep), seafood (amphibians, fish and shellfish), and turkey; 2) Source 2 (S2) denoting the type of food product thereof, i.e. dairy, eel, fresh meat (any animal), frog, herring, lobster, mackerel,

Table 1
Number of *Lm* isolates from non-human samples categorized by host (columns) and food group therein (rows).

	Total	Chicken	Cattle	Seafood[1]	Small ruminants[2]	Turkey	Plants[3]	Pig	Game	Unknown[4]
Fresh meat[5]	449	248	149		13	23		11	5	
Processed meat[6]	199	68	110		2	11		7		1
Salmon	87			87						
Trout	51			51						
Herring	41			41						
Unknown[4]	36		28	2	4					2
Plants[3]	20						20			
Feces[7]	19	3			16					
Dairy	12		12							
Mackerel	12			12						
Eel	10			10						
Shrimp	10			10						
Frog	2			2						
Lobster	1			1						
Mussel	1			1						
Total	950	347	271	217	35	34	20	18	5	3

[1] Includes amphibians, fish and shellfish.
[2] Includes sheep and goats.
[3] Includes, fruit, herbs, spices and vegetables.
[4] Includes isolates that could not be classified unambiguously.
[5] Any meat that has been modified/transformed in order to improve its taste and/or extend its shelf life (through salting, curing, fermentation, heating or smoking). Processed meat did not include simple mechanical processes, such as cutting, grinding or mixing.
[6] Any meat that has not been modified/transformed in order to improve its taste and/or extend its shelf life as described above: only simple mechanical processes, such as cutting, grinding or mixing were allowed.
[7] Isolates from feces samples from chicken and small ruminants.

mussel, plants, processed meat (any animal), salmon, shrimp, and trout; 3) Source 3 (S3) denoting the food-processing establishments that produced and/or packed the products (categorized through anonymized EC identification and health marks). An additional analysis was performed by combining S1 and S2 (hereafter referred to as S1 + 2).

In S2, processed meat was defined as any meat that has been modified/transformed in order to improve its taste and/or extend its shelf life (through curing, fermenting, heating, salting or smoking). Processed meat did not include simple mechanical processes, such as cutting, grinding or mixing. Conversely, fresh meat was defined as any meat that has not been modified/transformed in order to improve its taste and/or extend its shelf life as described above: only simple mechanical processes, such as cutting, grinding or mixing were allowed.

The food isolates that could not be categorized unambiguously among the aforementioned food sources were excluded from the analyses. These were mainly isolates from samples with only generic indications of their origin, such as “meat” or “sausage”. Also food sources with less than five isolates were excluded. This occurred only in S2 for amphibians, lobsters, mussels, and several food-processing establishments. No additional information about the food samples was available to further categorize them. In total, 947, 910, 484 and 904 food *Lm* isolates were included in the analyses for S1, S2, S3 and S1 + 2, respectively.

2.4. Random forest model

Source attribution analysis aimed at predicting the origin of human *Lm* isolates (i.e., to classify them) according to the aforementioned categorizations of food *Lm* isolates (S1, S2, S1 + 2, and S3). For each category separately, a Random Forest (RF) model was built using the Scikit (v1.3.0) library (Pedregosa et al., 2011) in Python (v3.9). All 1701 loci of the cgMLST profiles were used as predictive features, excluding those with ≥5 % missing alleles; missing values were then imputed using K-Nearest Neighbors (KNN) algorithm. Each cgMLST allele was one-hot encoded using the OneHotEncoder function implemented in scikit-learn, yielding a total of 46,000 binary features. Features with the highest predictive power were selected by applying a multivariate unbiased variable selection method (Shi et al., 2018), as implemented in the py-MUVR package (v1.0.1), available at https://github.com/datarevenue-berlin/py-MUVR. A total of 1270, 833, 2619 and 492 binary features were selected for inclusion in the RF models for S1, S2, S1 + 2 and S3, respectively.

RF hyperparameters were optimized using a random search in a predefined search space. We performed 10-fold stratified cross-validation to assess the quality of hyperparameters combination by using the RandomizedSearchCV function from the Scikit-Learn library; optimum values were chosen based on model accuracy. Given the imbalanced data set, training sets in each fold were oversampled using the RandomOversampler function from the imblearn library (v0.11.0) (Lemaître et al., 2017). To evaluate the impact of this step, each model was also optimized without up-sampling the training data. Parameters optimized included the number of decision trees, the maximum depth of the tree and the number of features to consider when searching for the best split. The values of optimized hyperparameters for each model can be found in Supplementary Table 11. The overall performance of each optimized model was evaluated using precision, recall, F1 scores and a confusion matrix. All metrics were computed by aggregating the results from all cross-validation folds. For final source attribution of human isolates, RF models with optimized parameters were re-trained using all available food isolates. Final predictions obtained with RF models rely on the proportions of individual decision trees voting for each class (i.e., food source). The final classification of human isolates was determined by the most frequently voted class, considering all decision trees that compose the RF model.

Because most food *Lm* isolates were obtained from food samples (i.e., from production facilities or retail), rather than directly from animals on farms, contamination of these foods from their production environments cannot be excluded. Consequently, the attribution estimates refer broadly to either the foods or their production environments as potential sources.

2.5. Risk factor analysis

Using data collected with the questionnaires administered to human listeriosis cases, a risk factor analysis was performed to identify risk factors for infection with *Lm* isolates attributable to specific sources (based on the results from the RF model with the highest accuracy). This risk factor analysis was performed as described previously (Mughini-

Gras et al., 2021). In brief, the attribution estimates for each human *Lm* isolate, as estimated by the RF model (i.e., the proportion of trees voting for a specific source, or class probability), were used as outcome variable. This allowed us to identify source-specific risk factors for human listeriosis, i.e., factors associated with increased or decreased risk for human *Lm* isolates to originate from specific food sources. We first performed a preliminary significance testing of 43 candidate risk factors using beta regression models with logit link, which is appropriate where the variable of interest is continuous but restricted to 0–1 (Ferrari and Cribari-Neto, 2004). All analyses were adjusted for patients' age (categorical variable, ≤64, 65–79, ≥80 years), sex (male or female, this latter in interaction with pregnancy status), and foreign travel history. Factors with a *p*-value <0.10 for the association with the outcome in the univariable analysis were selected for inclusion in a multivariable model built in stepwise fashion to retain only variables with a p-value <0.05. Variables were dropped one by one only if their exclusion from the model did not change the coefficients of the other covariates by >10 %. Collinear variables were identified before multivariable analysis using the variance inflation factor (VIF) and selection between collinear variables (VIF > 5) was made based on improved model fit as revealed by the Akaike information criterion (AIC). Risk factor analysis was performed in Stata version 16.0 (StataCorp, College Station, USA).

3. Results

3.1. Random Forest model validation

The mean model accuracy for S1 was the highest (0.794, Standard Deviation [SD]: 0.028), while the accuracy of S2, S1 + 2 and S3 were 0.704 (SD: 0.043), 0.603 (SD: 0.043) and 0.617 (SD: 0.065), respectively. For all models, a decrease in mean accuracy was observed when up-sampling the training data (Table 2).

In S1, we found that chicken, seafood and cattle were classified with an F1 score higher than 0.8 (Table 3). For these three classes, recall values were higher than precision. The remaining classes showed lower F1 scores, ranging from 0.000 (game) to 0.689 (turkey), and precision surpassed recall in all cases. Game, plants and small ruminants were classified with low precision and recall, while swine and turkey had high precision values (0.778 in both classes), but low recall (0.389 and 0.618, respectively). Up-sampling led to an increase in recall only for the classification of turkey isolates, while recall for the other classes remained constant or decreased.

Confusion matrix for S1 (Table 4) revealed that 29.4 % (10/34) of turkey isolates were frequently misclassified as chicken, contributing to the observed decrease in precision and recall values for this class. Similarly, 48.6 % (17/35) of isolates from small ruminants were misclassified as cattle and 25.7 % (9/35) as chicken. Comparable results were obtained with the model trained with up-sampled data (Supplementary Table 1).

In S2, dairy, fresh meat, herring, salmon and trout showed high F1 scores, ranging from 0.779 (herring) to 0.869 (trout) (Supplementary Table 2). Processed meat showed low recall (0.442) and precision (0.568), despite being the second most numerous class (21.8 %, 199/910). Up-sampling led to moderate increases in F1 scores for processed meat and eel. For all other classes, up-sampling led to lower F1 scores. Precision and recall for mackerel were 0 for both models, indicating that the minority class was never predicted correctly. Classification metrics for all classes were negatively influenced by fresh meat, which was the majority class (Supplementary Table 3). Moreover, 50.25 % (100/199) of isolates from processed meat were misclassified as fresh meat. Conversely, 11.3 % (51/449) of fresh meat isolates were misclassified as processed meat. Similar results were observed in the up-sampled model (Supplementary Table 4). In S1 + 2, salmon, dairy from cattle and trout were classified with the highest F1 scores (~0.8 for all classes), followed by herring and fresh meat from chicken (~0.7) (Supplementary Table 5). The other classes were classified with F1 scores below 0.6. Models with up-sampled data achieved similar classification scores in most classes (Supplementary Table 5). Isolates of most classes were commonly misclassified as fresh meat from chicken or cattle, which negatively impacts precision and recall values. In the model without up-sampling, confusion with these classes reached 59 and 109 isolates, respectively (Supplementary Table 6). Comparable outcomes were observed in the up-sampled model, although confusion with chicken (fresh meat) decreased to 76 isolates (Supplementary Table 7). Notably, dairy from cattle consistently avoided confusion with these specific classes.

In S3 model, producers including bbb, dddd, gg, nnnn, s, v, wwww and x, achieved classification F1 scores exceeding 0.8 (Supplementary Table 8). Notably, the 78 isolates from the predominant class (producer jj) were classified with an F1 score of 0.73. The remaining classes showed F1 score values below 0.67. Similar results were obtained after up-sampling the data. Confusion with the majority class impacted the classification metrics of 50 % (15/30) and 43 % (13/30) classes before and after up-sampling, respectively. Although the number of classes was high, clearly some producers (e.g., ddd, hhh, iii, kkk, q) were mainly misclassified with others (e.g., ggg/hhhh, mmm, kkk/mmm, mmm, hhhh, respectively) in both models with and without up-sampling (Supplementary Tables 9 and 10). The optimized hyperparameter values for each model obtained through 10-fold cross validation are reported in Supplementary Table 11.

Table 2
Accuracy of Random Forest model validation with and without up-sampling.

	Number of classes	Without up-sampling		With up-sampling	
		Mean accuracy	SD	Mean accuracy	SD
Source 1 ("host level")	8	0.794	0.028	0.765	0.035
Source 2 ("food product level")	11	0.704	0.043	0.674	0.05
Source 1 + 2 ("host + food product level")	19	0.603	0.045	0.574	0.043
Source 3 ("producer level")	30	0.617	0.065	0.547	0.061

3.2. Attributable sources of human listeriosis cases

RF models without up-sampling (S1, S2, S1 + 2 and S3) were employed to predict the source of human *Lm* isolates. A summary of the attributed sources is presented in Table 5. Human *Lm* isolates were attributed to five classes based on S1: most human *Lm* isolates were attributed to cattle (62.3 %, 471/756), chicken (19.4 %, 147/756) and seafood (16.9 %, 128/756). In S2, human *Lm* isolates were attributed to 9 classes: fresh meat predominated (86.2 %, 652/756), followed by processed meat (7.4 %, 56/756) and salmon (3.8 %, 29/756). In S1 + 2, most human *Lm* isolates were attributed to fresh meat from cattle (43.7 %, 331/756), fresh chicken meat (39.3 %, 297/756), salmon (6.2 %, 47/756) and processed meat from cattle (5.8 %, 44/756). In S3, three producers (jj, dddd and rr) accounted together for most human *Lm* isolates (60.6 %, 460/756).

The RF class probability predictions of each human *Lm* isolate based on S1 are illustrated in Fig. 1. For the human *Lm* isolates classified as originating from cattle, the second most likely source was chicken (63.7 %, 300/471). Conversely, for the majority of human *Lm* isolates attributed to chicken, cattle was the second most likely source (92.5 %, 136/147). The second most probable sources for human *Lm* isolates attributed to seafood, plants or small ruminants were cattle and chicken as well (94.9 %, 131/138, for cattle and chicken combined).

The RF class probability predictions of human *Lm* isolates based on S2, S1 + 2 and S3 models are reported in Supplementary Figs. 1–3. In S2,

Table 3
Random Forest model validation statistics for host level classes (S1), with and without up-sampling.

Source	Number of isolates	Without up-sampling			With up-sampling		
		Precision	Recall	F1	Precision	Recall	F1
Seafood[1]	217	0.813	0.880	0.845	0.803	0.866	0.834
Cattle	271	0.783	0.827	0.804	0.802	0.808	0.805
Chicken	347	0.839	0.859	0.849	0.862	0.793	0.826
Game	5	0.000	0.000	0.000	0.000	0.000	0.000
Plants[3]	20	0.417	0.250	0.313	0.294	0.250	0.270
Small ruminants[2]	35	0.300	0.171	0.218	0.171	0.171	0.171
Pig	18	0.778	0.389	0.519	0.538	0.389	0.452
Turkey	34	0.778	0.618	0.689	0.481	0.735	0.581
Average		0.588	0.499	0.530	0.494	0.502	0.492
Weighted Average		0.781	0.794	0.784	0.770	0.766	0.766

[1] Includes fish, shellfish and amphibians.
[2] Includes sheep and goats.
[3] Includes vegetables, fruit, spices and herbs.

Table 4
Confusion matrix of the Random Forest model for host level classes (S1) without up-sampling.

	Seafood[1]	Cattle	Chicken	Game	Plants[3]	Small ruminants[2]	Pig	Turkey
Seafood[1]	191	13	7	1	3	2	0	0
Cattle	14	224	22	2	1	5	2	1
Chicken	17	21	298	0	2	5	0	4
Game	0	4	1	0	0	0	0	0
Plants[3]	6	2	7	0	5	0	0	0
Small ruminants[2]	2	17	9	0	0	6	0	1
Pig	4	4	1	0	1	1	7	0
Turkey	1	1	10	0	0	1	0	21

[1] Includes amphibians, fish and shellfish.
[2] Includes goats and sheep.
[3] Includes fruit, herbs, spices and vegetables.

human *Lm* isolates classified as originating from fresh meat had processed meat as the second most likely source (84.35 %, 550/652). For most other isolates (89.5 %, 43/48), fresh meat predominated as the second most likely source. In S1 + 2, human *Lm* isolates classified as originating from fresh cattle meat had fresh chicken meat as the second most likely source (44.1 %, 146/331), followed by processed cattle meat, which appeared as the second most likely source in 29.39 % (99/337) of isolates. Fresh cattle meat was the second most likely source in 86.9 % (258/297) of isolates classified as originating from fresh chicken meat. For isolates classified as originating from processed cattle meat, fresh chicken meat was the second most likely source in 54.5 % of cases, thereby surpassing the number of cases in which fresh cattle meat was the second most probable class (36.4 %, 16/44). In S3, human *Lm* isolates estimated to originate from the main producers dddd, jj and rr had a generally homogeneous distribution of class probabilities across other producers.

3.3. Source specific risk factors

Model validation showed that S1 had the highest accuracy; thus, further exploration of potential risk factors for human listeriosis originating from specific sources was conducted using class probabilities at host level. All 756 human listeriosis cases were included in the risk factor analysis. Mean class probabilities from the RF model for these cases were as follows: cattle 37.0 %, chicken 26.2 %, seafood 21.5 %, small ruminants 6.9 %, plants 2.6 %, pig 2.5 %, turkey 2.5 % and game 0.7 %. Given the relatively low probabilities for sources other than cattle, chicken and seafood, risk factors were presented only for these three main sources.

As shown in Table 6, the only factor significantly associated with increased risk for a human *Lm* isolate to originate from cattle, was consumption of steak tartare, whereas consumption of smoked salmon was significantly associated with decreased risk for a human *Lm* isolate to originate from cattle. However, smoked salmon consumption was significantly associated with increased risk for a human *Lm* isolate to be attributed to seafood, whereas consumption of roast beef was significantly associated with decreased risk to originate from seafood. Finally, consuming steak tartare and soft cheese were significantly associated with decreased risk for a human *Lm* isolate to originate from chicken. No other significant factors were identified. Also the a-priori confounders (age, sex and travel history), as well as several underlying conditions and medicines, were not significant.

4. Discussion

In this study, RF-based modelling of cgMLST data was used to attribute human *Lm* isolates to various food sources in the Netherlands. In total, 756 human *Lm* isolates were attributed, with S1 (host level) source categorization providing the most accurate predictions. Yet, results were consistent across attribution levels, with cattle accounting for most human *Lm* isolates (62.3 %) at host level (S1), fresh meat at food level (86.2 %, S2) and fresh bovine meat at host and food levels combined (43.8 %, S1 + 2). Other important sources were chicken and seafood in S1 (19.4 % and 16.9 %). These results aligned well also with those from the S2 and S1 + 2 models, as the sources ranking second and third after fresh meat in S2 were processed meat (7.4 %) and salmon (3.8 %), and those in S1 + 2 were fresh chicken meat and salmon (39.2 % and 6.2 %), respectively. These results largely agree with those of Coipan et al. (2023) based on the same data where another approach was used based on Hamming distance clustering of *Lm* sequences, showing that most human *Lm* isolates might originate in almost equal proportions from cattle, chicken and seafood, and about 10 % from small ruminants.

In the present study, a picture emerged in which cattle, and particularly fresh meat of bovine origin, appeared to be the predominant source of human listeriosis cases. In the Netherlands, a meta-analysis of

Table 5
Attribution estimates for human *Lm* isolates from the four Random Forest models fitted with different source categorizations.

	Attributed human *Listeria* isolates	
	Number	Percentage
Source categorization 1 ("host level")		
Cattle	471	62.3
Chicken	147	19.44
Seafood[1]	128	16.93
Small ruminants[2]	6	0.79
Plants[3]	4	0.53
Source categorization 2 ("food level")		
Fresh meat[5] (any animal)	652	86.24
Processed meat[4] (any animal)	56	7.41
Salmon	29	3.84
Trout	11	1.46
Feces (small ruminants)[2]	5	0.66
Dairy	2	0.26
Herring	1	0.13
Source categorization 1 + 2 ("host-food level")		
Fresh meat[5] from cattle	331	43.78
Fresh meat[5] from chicken	297	39.29
Salmon	47	6.22
Processed meat[4] from cattle	44	5.82
Feces from small ruminants[2]	20	2.65
Trout	11	1.46
Plants	4	0.53
Herring	1	0.13
Dairy from cattle	1	0.13
Source categorization 3 ("producer level")		
jj	255	33.73
dddd	110	14.55
rr	95	12.57
ggg	67	8.86
ddd	56	7.41
bbbbb	54	7.14
ccc	21	2.78
uuuu	18	2.38
hhh	16	2.12
aaaa	15	1.98
hhhh	11	1.46
qq	11	1.46
gg	8	1.06
tt	5	0.66
mmm	4	0.53
q	3	0.4
xxxx	2	0.26
nnnn	2	0.26
iiii	1	0.13
j	1	0.13
kk	1	0.13

[1] Includes amphibians, fish and shellfish.
[2] Includes goats and sheep.
[3] Includes fruit, herbs, spices and vegetables.
[4] Any meat that has been modified/transformed in order to improve its taste and/or extend its shelf life (through curing, fermentation, heating, salting or smoking). Processed meat did not include simple mechanical processes, such as cutting, grinding or mixing.
[5] Any meat that has not been modified/transformed in order to improve its taste and/or extend its shelf life as described above: only simple mechanical processes, such as cutting, grinding or mixing were allowed.

source attribution studies based on subtyping and epidemiological data (Mughini-Gras et al., 2022), including an expert elicitation (Havelaar et al., 2008), also pointed to cattle as the main source of human listeriosis. This is also in agreement with findings reported elsewhere. For instance, a study in eleven European countries in 2010–2013 using five different source attribution models based on different MLST and cgMLST schemes also identified cattle as the main source of human listeriosis (38–64 %) (Nielsen et al., 2017). In the UK, a study based on *Lm* serotyping and amplified fragment length polymorphism (AFLP) data for 2004–2007 also identified cattle as an important source, particularly RTE beef products (15 %), which were preceded by plant-based foods, including sandwiches and prepacked mixed vegetable salads (23 %), and finfish (17 %) (Little et al., 2010). Yet, for pregnancy-associated listeriosis, beef (12 %), dairy (12 %) and finfish (11 %) were the most important sources (Little et al., 2010).

A source attribution study from Northern Italy based on MLST and Multi-Virulence Locus Sequence Typing (MVLST) data for 2005–2016 attributed 50 % of listeriosis cases to dairy products, followed by poultry and pork (15 % each), and mixed foods (15 %) (Filipello et al., 2020). Moreover, dairy products and plants have been found to be important sources of human listeriosis in a more recent study in the United States of America (USA) that also used a ML model (logit boost) based on cgMLST data (Tanui et al., 2022). However, the training data set in this USA study had a higher number of *Lm* isolates from plant and dairy sources, which probably reflect the importance of these sources in certain countries and therefore the differences in attribution estimates between studies. Dairy products have also been emphasized as a primary source of human listeriosis in other studies (Greig and Ravel, 2009), and vegetables are a known source of *Lm*, whose presence in, e.g., vegetable salads was confirmed as the cause of several documented *Lm* outbreaks (Ajayeoba et al., 2016; Ponniah et al., 2010; Stephan et al., 2015; Truong et al., 2021). Especially in the USA, foodborne outbreaks, including those caused by *Lm*, are often liked to produce items, including cantaloupes, celery, sprouts, and leafy greens among others (Cartwright et al., 2013; Garner and Kathariou, 2016; Pomeroy et al., 2021), as well as frozen vegetables and fruit (Madad et al., 2023).

Cross-contamination of vegetables from fresh meat might also explain the observed predominance of fresh meat found in this study, as our analysis is not able to discern whether meat processing or preparation was responsible for the *Lm* isolates of plant origin. This could have happened even before at primary production through spread/run-off of bovine manure in the environment causing *Lm* contamination of vegetables (Martín et al., 2014; Muhterem-Uyar et al., 2015; Zwirzitz et al., 2021). Cattle farm environments usually show high prevalence of *Lm* (Rocha et al., 2013), and ruminants do appear to contribute to amplification and spread of *Lm* in the farm environment (Nightingale et al., 2004). Poultry is also a recognized reservoir of *Lm*, and raw chicken meat can pose a risk to consumers when handled unhygienically or consumed insufficiently cooked (Dhama et al., 2013). Several studies have found high *Lm* prevalence in seafood and related environments, with RTE seafood in particular being most often involved as the source of outbreaks, although the concentrations of *Lm* in seafood are usually low (<10 colony-forming units [CFU]/g) (Jami et al., 2014).

Previous source attribution studies on listeriosis have included predominantly RTE products (Filipello et al., 2020; Little et al., 2010; Nielsen et al., 2017), as these foods are usually consumed without additional processing by the consumer, and their properties and refrigerated storage conditions often makes bacterial growth likely. This is less common for, e.g., campylobacteriosis or salmonellosis, which are often attributed at animal reservoir or primary production levels (Mossong et al., 2016; Mughini-Gras et al., 2014a; Mughini-Gras et al., 2014b; Mughini-Gras et al., 2016; Mughini-Gras et al., 2021; Pires et al., 2009; Pires and Hald, 2010). Ideally, control activities for foodborne pathogens in general should target reservoirs at primary production to prevent pathogen spread as soon as possible in the food supply chain. Indeed, effective management of cross-contamination at food-processing (e.g., sliced cold cut manufacturing), distribution (e.g., deli meat counters) and household level (e.g., food preparation/storage) is more achievable when contamination is minimized already at reservoir level (Filipello et al., 2020). This is particularly challenging for *Lm*, as it is still largely unclear how *Lm* is transmitted among animals, humans and the environment (Walland et al., 2015). Moreover, *Lm* is ubiquitous and can become established in food production facilities, resulting in (cross-)contamination of a variety of food products (Gupta and Adhikari, 2022).

Here we performed attribution analyses at different levels (host, food product, and producer), but because most *Lm* isolates originated from

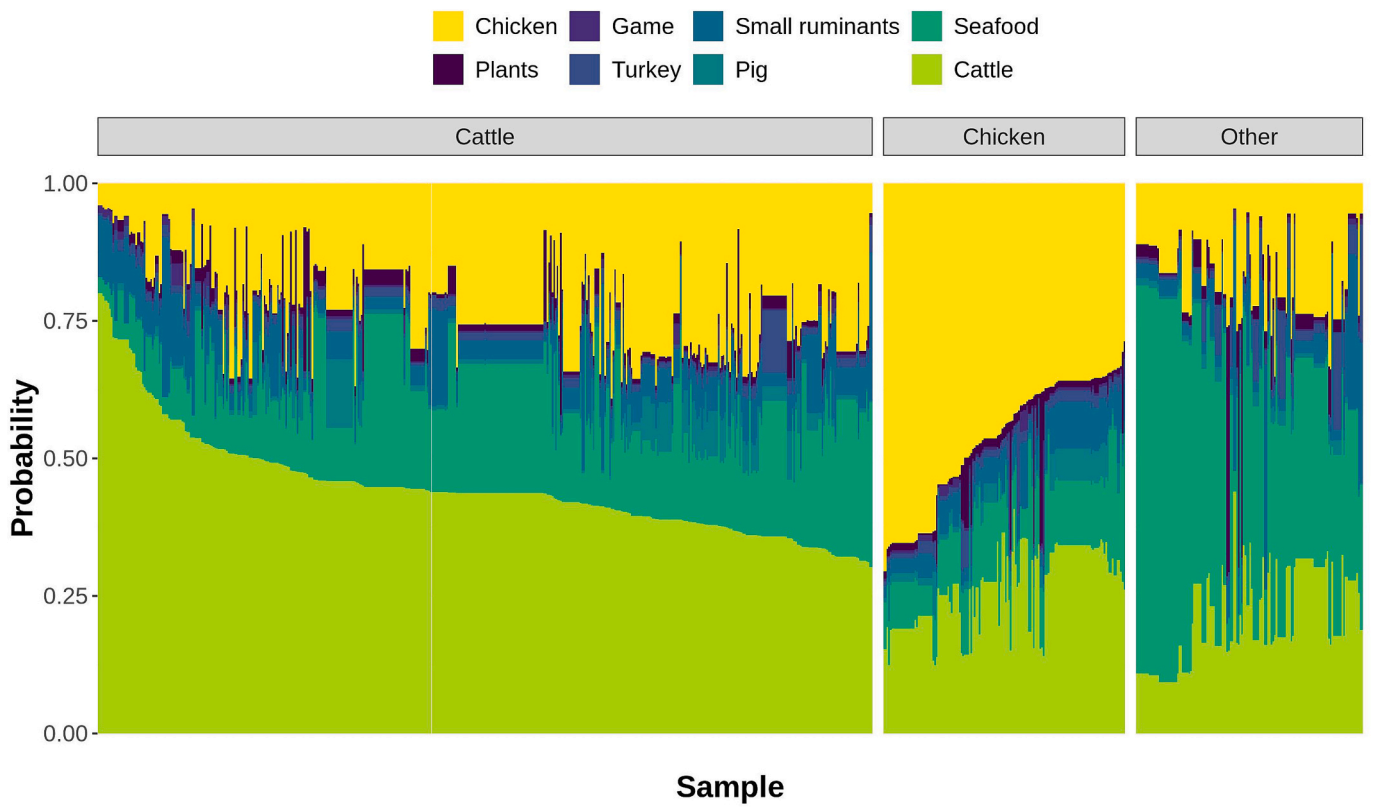

Fig. 1. Random Forest model class probability of each human *Lm* isolate (x-axis) attributed to sources (y-axis) according to host level source categorization (S1).

food samples (from production facility or retail), we cannot exclude that they were the result of contamination from the production environment. Therefore, we cannot talk here about animal reservoirs of *Lm*. Consistent with this hypothesis, we found that our food product-level model (S2) struggled to classify isolates from processed meat, despite this being the second most abundant class. This finding is significant, as it indicates that the signal for source attribution diminishes when meat is processed. As contamination of production environments must have come from somewhere, it is conceivable that a given food production facility is (mainly) contaminated with *Lm* strains originally coming from the main raw materials processed. Our producer-based model (S3) was able to identify a few food producers processing different food categories that accounted altogether for most human *Lm* isolates. However, these results should be interpreted with caution, as the S3 model exhibited the lowest accuracy among all models. Moreover, there are of course many other producers that were not in the sample. Previous analyses based on *Lm* genetic clustering based on cgMLST demonstrated stronger signals and a higher degree of temporal persistence of clusters at the food producer level compared to the food source level (Coipan et al., 2023). Moreover, some genetic clusters involved food categories from multiple food producers, which is compatible with the hypothesis of contaminated raw materials, but some clusters also involved different food categories from the same food producers, which is compatible with cross-contamination (Coipan et al., 2023). Overall, our findings indicate that attributions at the food level can also be accurate and offer valuable insights, similar to those from previous studies (Coipan et al., 2023).

We also performed a risk factor analysis combining the attribution estimates from S1 with case exposure data to identify source-specific risk factors for human listeriosis of cattle, chicken and seafood origin. This approach has been applied previously for salmonellosis and campylobacteriosis, using either a source-assigned case-control (Mossong et al., 2016; Mughini-Gras et al., 2014b; Mughini-Gras et al., 2012; Mughini-Gras et al., 2018; Rosner et al., 2017) or case-case (Bessell et al., 2012; Levesque et al., 2013; Mullner et al., 2010) study design. In these studies, groups of cases were first assigned to specific sources based on their majority attributable source and then the exposures of these groups of cases were compared with one another or with a control group. Here, instead, we modelled the attributions directly as an outcome variable with the corresponding exposure data for cases only, as described in a recent study (Mughini-Gras et al., 2021). This approach has the advantage to capture more fine-grain differences in associations and to optimize data use, as no observation is excluded due to human case grouping by majority attributable source. The combined source attribution and risk factor analysis allowed for the identification of possible pathways (in this case foodborne) by which the infecting *Lm* isolates might have reached humans, including specific food products involved in the transmission of *Lm* strains originating from specific animal sources.

We found that consuming steak tartare increased the risk for a human *Lm* isolate to originate from cattle, and consuming smoked salmon increased the risk for those *Lm* isolates to originate from seafood. Clearly, these associations are highly plausible. Steak tartare is a dish which main ingredient is raw ground (minced) beef, and smoked salmon is a fish preparation, typically a salmon fillet that has been cured and (hot or cold) smoked. Both these products have a high likelihood of being contaminated with *Lm* isolates originating from cattle and seafood (i.e., a source category including fish and shellfish here), respectively. In general, red meats and particularly beef, is more likely to be (intentionally) consumed raw or undercooked, and thus more likely to harbor viable *Lm* isolates due to incomplete cooking, than other types of meats. For instance, among 3165 control participants in a case-control study of human salmonellosis in the Netherlands, those reporting to have consumed raw or undercooked meat were 24.6 %, 8.8 % and 7.7 % for beef, pork, and chicken, respectively (Mughini-Gras et al., 2014b). Indeed, several popular dishes containing raw or undercooked meat are based on beef (e.g., carpaccio, steak tartare and similar dishes like yukhoe, filet Americain or other raw meat spreads like ossenworst and ciğ köfte, etc.). This is different for, e.g., chicken meat, which is most

Table 6
Factors associated with increased or decreased risk for human *Lm* isolates to originate from cattle, chicken or seafood.

	N	Attributable source probability Exponentiated β-coefficient (95 % Confidence Interval)		
		Cattle	Chicken	Seafood
Age group				
≤64 years	196	Reference	Reference	Reference
65–79 years	293	0.911 (0.813–1.020)	1.046 (0.919–1.191)	1.041 (0.899–1.205)
≥80 years	195	0.949 (0.839–1.073)	0.932 (0.809–1.074)	1.010 (0.861–1.185)
Unknown	72	0.628 (0.357–1.107)	1.133 (0.653–1.965)	1.603 (0.816–3.152)
Sex				
Male	391	Reference	Reference	Reference
Female (non-pregnant)	251	1.006 (0.917–1.103)	1.062 (0.956–1.180)	0.996 (0.884–1.123)
Female (pregnant)	41	1.099 (0.894–1.351)	1.015 (0.800–1.288)	0.812 (0.615–1.073)
Female (unknown pregnancy)	2	1.850 (0.840–4.073)	0.706 (0.261–1.910)	0.607 (0.193–1.913)
Unknown sex	71	1.172 (0.659–2.084)	0.854 (0.486–1.502)	0.823 (0.412–1.644)
Travel history				
No	635	Reference	Reference	Reference
Yes	28	0.990 (0.792–1.237)	0.883 (0.679–1.147)	1.075 (0.804–1.437)
Unknown	93	1.011 (0.869–1.176)	1.018 (0.853–1.214)	0.983 (0.809–1.195)
Ate smoked salmon				
No	362	Reference	Reference	Reference
Yes	164	0.842 (0.754–0.939)**	ns	1.276 (1.111–1.465)***
Unknown	230	1.065 (0.865–1.312)	ns	0.937 (0.705–1.247)
Ate steak tartare				
No	448	Reference	Reference	ns
Yes	60	1.314 (1.125–1.534)***	0.831 (0.691–0.998)*	ns
Unknown	248	0.931 (0.762–1.137)	0.944 (0.763–1.168)	ns
Ate roast beef				
No	399	ns	ns	Reference
Yes	109	ns	ns	0.795 (0.675–0.936)**
Unknown	248	ns	ns	1.118 (0.850–1.472)
Ate soft cheese				
No	406	ns	Reference	ns
Yes	127	ns	0.871 (0.761–0.997)*	ns
Unknown	223	ns	0.962 (0.766–1.209)	ns

ns = not significant (p-value>0.05).
* p-value <0.05.
** p-value <0.01.
*** p-value <0.001.

often accidentally (rather than intentionally) consumed undercooked, although it can contribute substantially to human infections through cross-contamination and consumption of contaminated RTE products (Lomonaco et al., 2013). This might be also be a reason as to why no significant risk factors were identified for human infection with chicken-borne *Lm* isolates.

Only two protective factors were identified for human infection with chicken-borne *Lm* isolates: consumption of steak tartare and consumption of soft cheese. Protective factors were also found for human infection with *Lm* isolates of seafood origin (consumption of roast beef) and of cattle origin (consumption of smoked salmon). These associations mean that having consumed these products was associated with decreased risk for human *Lm* isolates to originate from these sources. Also these negative associations are highly plausible, as people consuming these products would be less at risk of infection with *Lm* isolates originating from other, different sources, as observed previously as well (Mossong et al., 2016; Mughini-Gras et al., 2014b; Mughini-Gras et al., 2021; Mughini-Gras et al., 2012; Mughini-Gras et al., 2018). Overall, these results indicate that the exposure matches the source in question. Because the exposure data on risk factors were independently generated (i.e., independent from the sequence data used for attribution), the fact that they point to the most logical outcomes of the RF predictions can be seen as a validation of the attributions themselves. Moreover, the a-priori confounders included in the models (age, gender and travel history) were always non-significant, and the same was true for underlying diseases and medicine use, indicating that these are common risk factors to all *Lm* infections, regardless of the origin of the isolates.

Similar to other studies of this kind, this study has some limitations. Firstly, the risk factor analysis included only case exposure data. Although this eliminated issues related to, e.g. differential recall bias, selection bias, misclassification, etc. between cases and controls, it might be less sensitive in identifying common risk factors for listeriosis in general, such as underlying diseases among others. Other limitations were related to different isolation media and multiple hypothesis testing. However, this study was explorative in nature and meant to generate hypotheses rather than test them conclusively. As cases originated from routine diagnostic activities of people seeking medical care, including isolates from invasive cases, they represent the most severe, symptomatic infections occurring in the population. Thus, the attributions and source-specific risk factors identified here pertained to severe listeriosis cases. As data was limited to one country, the generalizability of findings is constrained by factors such as local eating habits, production practices and regulations. While there is substantial uniformity within the European Union (EU) regarding production practices and regulations thanks to European common market policies, EU Member State retain the flexibility to implement regulations in ways that suit their local contexts (e.g., varying testing frequencies for high-risk foods), and variations in production practices might require tailored risk management approaches. Accordingly, the results from the present study agree to a major extent with those of studies in other European countries, suggesting that the Dutch situation is unlikely to be unique.

There were also limitations inherent to ML methods. The training data set for S1 model exhibited class imbalance, with under-representation of plants, small ruminants, pig, turkey and game-derived isolates. This imbalance challenges the ability of the algorithm to correctly classify the minority classes, probably due to lack of information, class overlapping or small disjunct in the minority classes (Ali et al., 2015). Despite employing random up-sampling to artificially balance the training data set, no improvements in classification metrics were observed in most cases. Furthermore, optimal validation of all RF models would ideally involve testing on an external data set (Tanui et al., 2022). Finally, a previous study (Castelli et al., 2023) demonstrated superior performance in *Lm* source attribution using accessory genome content as input for a RF model compared to a cgMLST-based model. Although our S1 model outperformed the cgMLST-based RF model from Castelli et al. (2023), the potential benefits of incorporating accessory genome content into future studies warrant further evaluation.

In conclusion, cattle, and specifically fresh bovine meat, appeared to be the main attributable source of human listeriosis in the Netherlands, followed by (fresh meat from) chicken and seafood, particularly (smoked) salmon. These sources were identified using a RF-based source attribution analysis based on cgMLST data, suggesting that *Lm* isolates from either these foods or their production environments are likely main

contributors to human *Lm* infection. Moreover, risk factors at exposure level for human *Lm* infection clearly pointed towards those food sources that were most likely involved as the origin of the *Lm* isolates in question. Indeed, these factors pertained to consumption of food products derived from the same hosts from which the human *Lm* isolates were likely to originate based on cgMLST. This further confirmed the reliability of the attribution estimates. Attributions were also generally consistent with foodborne transmission, as suggested by the risk factor analysis as well. Overall, we showed that risk factors for human *Lm* infection may differ depending on the attributable source and that a joint analysis of core genome and epidemiological data may provide insights into the origins and transmission pathways of human listeriosis.

CRediT authorship contribution statement

Lapo Mughini-Gras: Writing – review & editing, Writing – original draft, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Julian A. Paganini:** Writing – review & editing, Validation, Software, Methodology, Formal analysis. **Ruoshui Guo:** Writing – review & editing, Writing – original draft, Investigation, Formal analysis, Data curation. **Claudia E. Coipan:** Writing – review & editing, Investigation, Data curation, Conceptualization. **Ingrid H.M. Friesema:** Writing – review & editing, Investigation, Data curation. **Angela H.A.M. van Hoek:** Writing – review & editing, Investigation, Data curation. **Maaike van den Beld:** Writing – review & editing, Investigation, Data curation. **Sjoerd Kuiling:** Writing – review & editing, Investigation, Data curation. **Indra Bergval:** Writing – review & editing, Investigation. **Bart Wullings:** Writing – review & editing, Investigation, Data curation. **Menno van der Voort:** Writing – review & editing, Investigation, Data curation. **Eelco Franz:** Writing – review & editing, Project administration, Investigation, Funding acquisition, Conceptualization. **Timothy J. Dallman:** Writing – review & editing, Supervision, Methodology, Investigation, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was supported by the Dutch National Food and Consumer Products Authority and Dutch Ministry of Public Health, Welfare and Sports. The authors are thankful to Aarieke de Jong and Coen van der Weijden for their thoughtful comments on this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijfoodmicro.2024.110953.

Data availability

All sequences included in this study are publicly available at the European Nucleotide Archive (ENA, accession numbers PRJEB26058, PRJEB58647 and PRJEB54813) and in Coipan et al. (2023). Scripts can be accessed at https://github.com/jpaganini/listeria_source_attribution.

References

Ajayeoba, T.A., Atanda, O.O., Obadina, A.O., Bankole, M.O., Adelowo, O.O., 2016. The incidence and distribution of Listeria monocytogenes in ready-to-eat vegetables in South-Western Nigeria. Food Sci. Nutr. 4, 59–66.

Ali, A., Shamsuddin, S.M.H., Ralescu, A.L., 2015. Classification with class imbalance problem: a review. Soft Comput. Models Ind. Environ. 7, 176–204.

Allerberger, F., Wagner, M., 2010. Listeriosis: a resurgent foodborne infection. Clin. Microbiol. Infect. 16, 16–23.

Arning, N., Sheppard, S.K., Bayliss, S., Clifton, D.A., Wilson, D.J., 2021. Machine learning to predict the source of campylobacteriosis using whole genome data. PLoS Genet. 17, e1009436.

Bessell, P.R., Rotariu, O., Innocent, G.T., Smith-Palmer, A., Strachan, N.J., Forbes, K.J., Cowden, J.M., Reid, S.W., Matthews, L., 2012. Using sequence data to identify alternative routes and risk of infection: a case-study of campylobacter in Scotland. BMC Infect. Dis. 12, 80.

Buchanan, C.J., Webb, A.L., Mutschall, S.K., Kruczkiewicz, P., Barker, D.O.R., Hetman, B. M., Gannon, V.P.J., Abbott, D.W., Thomas, J.E., Inglis, G.D., Taboada, E.N., 2017. A genome-wide association study to identify diagnostic markers for human pathogenic Campylobacter jejuni strains. Front. Microbiol. 8, 1224.

Cartwright, E.J., Jackson, K.A., Johnson, S.D., Graves, L.M., Silk, B.J., Mahon, B.E., 2013. Listeriosis outbreaks and associated food vehicles, United States, 1998-2008. Emerg. Infect. Dis. 19, 1–9 quiz 184.

Castelli, P., De Ruvo, A., Bucciacchio, A., D'Alterio, N., Cammà, C., Di Pasquale, A., Radomski, N., 2023. Harmonization of supervised machine learning practices for efficient source attribution of Listeria monocytogenes based on genomic data. BMC Genomics 24, 560.

Cody, A.J., Bray, J.E., Jolley, K.A., McCarthy, N.D., Maiden, M.C.J., 2017. Core genome multilocus sequence typing scheme for stable, comparative analyses of Campylobacter jejuni and C. coli human disease isolates. J. Clin. Microbiol. 55, 2086–2097.

Coipan, C.E., Friesema, I.H.M., van Hoek, A., van den Bosch, T., van den Beld, M., Kuiling, S., Gras, L.M., Bergval, I., Bosch, T., Wullings, B., van der Voort, M., Franz, E., 2023. New insights into the epidemiology of Listeria monocytogenes - a cross-sectoral retrospective genomic analysis in the Netherlands (2010–2020). Front. Microbiol. 14, 1147137.

Dhama, K., Verma, A.K., Rajagunalan, S., Kumar, A., Tiwari, R., Chakraborty, S., Kumar, R., 2013. Listeria monocytogenes infection in poultry and its public health importance with special reference to food borne zoonoses. Pak. J. Biol. Sci. 16, 301–308.

EFSA, ECDC, 2023. The European Union one health 2022 zoonoses report. EFSA J. 21, e8442.

Ferrari, S., Cribari-Neto, F., 2004. Beta regression for modelling rates and proportions. J. Appl. Stat. 31, 799–815.

Ferreira, V., Wiedmann, M., Teixeira, P., Stasiewicz, M.J., 2014. Listeria monocytogenes persistence in food-associated environments: epidemiology, strain characteristics, and implications for public health. J. Food Prot. 77, 150–170.

Filipello, V., Mughini-Gras, L., Gallina, S., Vitale, N., Mannelli, A., Pontello, M., Decastelli, L., Allard, M.W., Brown, E.W., Lomonaco, S., 2020. Attribution of Listeria monocytogenes human infections to food and animal sources in Northern Italy. Food Microbiol. 89, 103433.

Franz, E., Gras, L.M., Dallman, T., 2016. Significance of whole genome sequencing for surveillance, source attribution and microbial risk assessment of foodborne pathogens. Curr. Opin. Food Sci. 8, 74–79.

Friesema, I.H., Kuiling, S., van der Ende, A., Heck, M.E., Spanjaard, L., van Pelt, W., 2015. Risk factors for sporadic listeriosis in the Netherlands, 2008 to 2013. Euro Surveill 20.

Garner, D., Kathariou, S., 2016. Fresh produce-associated listeriosis outbreaks, sources of concern, teachable moments, and insights. J. Food Prot. 79, 337–344.

Greig, J.D., Ravel, A., 2009. Analysis of foodborne outbreak data reported internationally for source attribution. Int. J. Food Microbiol. 130, 77–87.

Gu, W., Cui, Z., Stroika, S., Carleton, H.A., Conrad, A., Katz, L.S., Richardson, L.C., Hunter, J., Click, E.S., Bruce, B.B., 2023. Predicting food sources of listeria monocytogenes based on genomic profiling using random forest model. Foodborne Pathog. 20, 579–586.

Gupta, P., Adhikari, A., 2022. Novel approaches to environmental monitoring and control of listeria monocytogenes in food production facilities. Foods 11.

Havelaar, A.H., Galindo, A.V., Kurowicka, D., Cooke, R.M., 2008. Attribution of foodborne pathogens using structured expert elicitation. Foodborne Pathog. Dis. 5, 649–659.

Hurley, D., Luque-Sastre, L., Parker, C.T., Huynh, S., Eshwar, A.K., Nguyen, S.V., Andrews, N., Moura, A., Fox, E.M., Jordan, K., Lehner, A., Stephan, R., Fanning, S., 2019. Whole-genome sequencing-based characterization of 100 *Listeria monocytogenes* isolates collected from food processing environments over a four-year period. mSphere 4.

Jami, M., Ghanbari, M., Zunabovic, M., Domig, K.J., Kneifel, W., 2014. Listeria monocytogenes in aquatic food products—a review. Compr. Rev. Food Sci. Food Saf. 13, 798–813.

Koopmans, M.M., Brouwer, M.C., Bijlsma, M.W., Bovenkerk, S., Keijzers, W., van der Ende, A., van de Beek, D., 2013. Listeria monocytogenes sequence type 6 and increased rate of unfavorable outcome in meningitis: epidemiologic cohort study. Clin. Infect. Dis. 57, 247–253.

Lemaître, G., Nogueira, F., Aridas, C.K., 2017. Imbalanced-learn: a python toolbox to tackle the curse of imbalanced datasets in machine learning. J. Mach. Learn. Res. 18, 559–563.

Levesque, S., Fournier, E., Carrier, N., Frost, E., Arbeit, R.D., Michaud, S., 2013. Campylobacteriosis in urban versus rural areas: a case-case study integrated with molecular typing to validate risk factors and to attribute sources of infection. PloS One 8, e83731.

Little, C.L., Pires, S.M., Gillespie, I.A., Grant, K., Nichols, G.L., 2010. Attribution of human Listeria monocytogenes infections in England and Wales to ready-to-eat food

sources placed on the market: adaptation of the Hald Salmonella source attribution model. Foodborne Pathog. Dis. 7, 749–756.
Lomonaco, S., Verghese, B., Gerner-Smidt, P., Tarr, C., Gladney, L., Joseph, L., Katz, L., Turnsek, M., Frace, M., Chen, Y., Brown, E., Meinersmann, R., Berrang, M., Knabel, S., 2013. Novel epidemic clones of Listeria monocytogenes, United States, 2011. Emerg. Infect. Dis. 19, 147–150.
Lupolova, N., Dallman, T.J., Holden, N.J., Gally, D.L., 2017. Patchy promiscuity: machine learning applied to predict the host specificity of Salmonella enterica and Escherichia coli. Microb. Genom. 3, e000135.
Lupolova, N., Lycett, S.J., Gally, D.L., 2019. A guide to machine learning for bacterial host attribution using genome sequence data. Microb. Genom. 5.
Lupolova, N., Chalka, A., Gally, D.L., 2021. Predicting host association for Shiga toxin-producing E. coli serogroups by machine learning. Methods Mol. Biol. 2291, 99–117.
Madad, A., Marshall, K.E., Blessington, T., Hardy, C., Salter, M., Basler, C., Conrad, A., Stroika, S., Luo, Y., Dwarka, A., Gerhardt, T., Rosa, Y., Cibulskas, K., Rosen, H.E., Adcock, B., Kiang, D., Hutton, S., Parish, M., Podoski, B., Patel, B., Viazis, S., 2023. Investigation of a multistate outbreak of listeria monocytogenes infections linked to frozen vegetables produced at individually quick-frozen vegetable manufacturing facilities. J. Food Prot. 86, 100117.
Maiden, M.C.J., Jansen van Rensburg, M.J., Bray, J.E., Earle, S.G., Ford, S.A., Jolley, K. A., McCarthy, N.D., 2013. MLST revisited: the gene-by-gene approach to bacterial genomics. Nat. Rev. Microbiol. 11, 728–736.
Martín, B., Perich, A., Gómez, D., Yangüela, J., Rodríguez, A., Garriga, M., Aymerich, T., 2014. Diversity and distribution of Listeria monocytogenes in meat processing plants. Food Microbiol. 44, 119–127.
Mossong, J., Mughini-Gras, L., Penny, C., Devaux, A., Olinger, C., Losch, S., Cauchie, H. M., van Pelt, W., Ragimbeau, C., 2016. Human Campylobacteriosis in Luxembourg, 2010-2013: a case-control study combined with multilocus sequence typing for source attribution and risk factor analysis. Sci. Rep. 6, 20939.
Mughini-Gras, L., Smid, J.H., Wagenaar, J.A., de Boer, A.G., Havelaar, A.H., Friesema, I. H., French, N.P., Busani, L., van Pelt, W., 2012. Risk factors for campylobacteriosis of chicken, ruminant, and environmental origin: a combined case-control and source attribution analysis. PloS One 7, e42599.
Mughini-Gras, L., Barrucci, F., Smid, J.H., Graziani, C., Luzzi, I., Ricci, A., Barco, L., Rosmini, R., Havelaar, A.H., W, V.A.N.P., Busani, L., 2014a. Attribution of human Salmonella infections to animal and food sources in Italy (2002–2010): adaptations of the Dutch and modified Hald source attribution models. Epidemiol. Infect. 142, 1070–1082.
Mughini-Gras, L., Enserink, R., Friesema, I., Heck, M., van Duynhoven, Y., van Pelt, W., 2014b. Risk factors for human salmonellosis originating from pigs, cattle, broiler chickens and egg laying hens: a combined case-control and source attribution analysis. PloS One 9, e87933.
Mughini-Gras, L., Heck, M., van Pelt, W., 2016. Increase in reptile-associated human salmonellosis and shift toward adulthood in the age groups at risk, the Netherlands, 1985 to 2014. Euro Surveill 21.
Mughini-Gras, L., van Pelt, W., van der Voort, M., Heck, M., Friesema, I., Franz, E., 2018. Attribution of human infections with Shiga toxin-producing Escherichia coli (STEC) to livestock sources and identification of source-specific risk factors, the Netherlands (2010-2014). Zoonoses Public Health 65, e8–e22.
Mughini-Gras, L., Pijnacker, R., Coipan, C., Mulder, A.C., Fernandes Veludo, A., de Rijk, S., van Hoek, A., Buij, R., Muskens, G., Koene, M., Veldman, K., Duim, B., van der Graaf-van Bloois, L., van der Weijden, C., Kuiling, S., Verbruggen, A., van der Giessen, J., Opsteegh, M., van der Voort, M., Castelijn, G.A.A., Schets, F.M., Blaak, H., Wagenaar, J.A., Zomer, A.L., Franz, E., 2021. Sources and transmission routes of campylobacteriosis: a combined analysis of genome and exposure data. J. Infect. 82, 216–226.
Mughini-Gras, L., Benincà, E., McDonald, S.A., de Jong, A., Chardon, J., Evers, E., Bonačić Marinović, A.A., 2022. A statistical modelling approach for source attribution meta-analysis of sporadic infection with foodborne pathogens. Zoonoses Public Health 69, 475–486.
Muhterem-Uyar, M., Dalmasso, M., Bolocan, A.S., Hernandez, M., Kapetanakou, A.E., Kuchta, T., Manios, S.G., Melero, B., Minarovičová, J., Nicolau, A.I., Rovira, J., Skandamis, P.N., Jordan, K., Rodríguez-Lázaro, D., Stessl, B., Wagner, M., 2015. Environmental sampling for Listeria monocytogenes control in food processing facilities reveals three contamination scenarios. Food Control 51, 94–107.
Mullner, P., Shadbolt, T., Collins-Emerson, J.M., Midwinter, A.C., Spencer, S.E., Marshall, J., Carter, P.E., Campbell, D.M., Wilson, D.J., Hathaway, S., Pirie, R., French, N.P., 2010. Molecular and spatial epidemiology of human campylobacteriosis: source association and genotype-related risk factors. Epidemiol. Infect. 138, 1372–1383.
Munck, N., Njage, P.M.K., Leekitcharoenphon, P., Litrup, E., Hald, T., 2020. Application of whole-genome sequences and machine learning in source attribution of Salmonella typhimurium. Risk Anal. 40, 1693–1705.
Nielsen, E.M., Björkman, J.T., Kiil, K., Grant, K., Dallman, T., Painset, A., Amar, C., Roussel, S., Guillier, L., Félix, B., Rotariu, O., Perez-Reche, F., Forbes, K., Strachan, N., 2017. Closing gaps for performing a risk assessment on Listeria monocytogenes in ready-to-eat (RTE) foods: activity 3, the comparison of isolates from different compartments along the food chain, and from humans using whole genome sequencing (WGS) analysis. EFSA Supporting Publ. 14, 1151E.
Nightingale, K.K., Schukken, Y.H., Nightingale, C.R., Fortes, E.D., Ho, A.J., Her, Z., Grohn, Y.T., McDonough, P.L., Wiedmann, M., 2004. Ecology and transmission of Listeria monocytogenes infecting ruminants and in the farm environment. Appl. Environ. Microbiol. 70, 4458–4467.
Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.
Pires, S.M., Hald, T., 2010. Assessing the differences in public health impact of salmonella subtypes using a bayesian microbial subtyping approach for source attribution. Foodborne Pathog. Dis. 7, 143–151.
Pires, S.M., Evers, E.G., van Pelt, W., Ayers, T., Scallan, E., Angulo, F.J., Havelaar, A., Hald, T., 2009. Attributing the human disease burden of foodborne infections to specific sources. Foodborne Pathog. Dis. 6, 417–424.
Pohl, A.M., Pouillot, R., Bazaco, M.C., Wolpert, B.J., Healy, J.M., Bruce, B.B., Laughlin, M.E., Hunter, J.C., Dunn, J.R., Hurd, S., Rowlands, J.V., Saupe, A., Vugia, D.J., Van Doren, J.M., 2019. Differences among incidence rates of invasive listeriosis in the U.S. FoodNet population by age, sex, race/ethnicity, and pregnancy status, 2008-2016. Foodborne Pathog. Dis. 16, 290–297.
Pomeroy, M., Conrad, A., Pettengill, J.B., Mc, C.M., Wellman, A.A., Marus, J., Huffman, J., Wise, M., 2021. Evaluation of avocados as a possible source of listeria monocytogenes infections, United States, 2016 to 2019. J. Food Prot. 84, 1122–1126.
Ponniah, J., Robin, T., Paie, M.S., Radu, S., Ghazali, F.M., Kqueen, C.Y., Nishibuchi, M., Nakaguchi, Y., Malakar, P.K., 2010. Listeria monocytogenes in raw salad vegetables sold at retail level in Malaysia. Food Control 21, 774–778.
Rocha, P.R., Lomonaco, S., Bottero, M.T., Dalmasso, A., Dondo, A., Grattarola, C., Zuccon, F., Iulini, B., Knabel, S.J., Capucchio, M.T., Casalone, C., 2013. Ruminant rhombencephalitis-associated Listeria monocytogenes strains constitute a genetically homogeneous group related to human outbreak strains. Appl. Environ. Microbiol. 79, 3059–3066.
Rosner, B.M., Schielke, A., Didelot, X., Kops, F., Breidenbach, J., Willrich, N., Golz, G., Alter, T., Stingl, K., Josenhans, C., Suerbaum, S., Stark, K., 2017. A combined case-control and molecular source attribution study of human Campylobacter infections in Germany, 2011-2014. Sci. Rep. 7, 5139.
Ruppitsch, W., Pietzka, A., Prior, K., Bletz, S., Fernandez, H.L., Allerberger, F., Harmsen, D., Mellmann, A., 2015. Defining and evaluating a core genome multilocus sequence typing scheme for whole-genome sequence-based typing of Listeria monocytogenes. J. Clin. Microbiol. 53, 2869–2876.
Sheppard, S.K., Jolley, K.A., Maiden, M.C., 2012. A gene-by-gene approach to bacterial population genomics: whole genome MLST of campylobacter. Genes (Basel) 3, 261–277.
Shi, L., Westerhuis, J.A., Rosén, J., Landberg, R., Brunius, C., 2018. Variable selection and validation in multivariate modelling. Bioinformatics 35, 972–980.
Stephan, R., Althaus, D., Kiefer, S., Lehner, A., Hatz, C., Schmutz, C., Jost, M., Gerber, N., Baumgartner, A., Hächler, H., Mäusezahl-Feuz, M., 2015. Foodborne transmission of Listeria monocytogenes via ready-to-eat salad: a nationwide outbreak in Switzerland, 2013–2014. Food Control 57, 14–17.
Tanui, C.K., Benefo, E.O., Karanth, S., Pradhan, A.K., 2022. A machine learning model for food source attribution of Listeria monocytogenes. Pathogens 11.
Truong, H.-N., Garmyn, D., Gal, L., Fournier, C., Sevellec, Y., Jeandroz, S., Piveteau, P., 2021. Plants as a realized niche for Listeria monocytogenes. MicrobiologyOpen 10, e1255.
Van Walle, I., Björkman, J.T., Cormican, M., Dallman, T., Mossong, J., Moura, A., Pietzka, A., Ruppitsch, W., Takkinen, J., 2018. Retrospective validation of whole genome sequencing-enhanced surveillance of listeriosis in Europe, 2010 to 2015. Euro Surveill 23.
Walland, J., Lauper, J., Frey, J., Imhof, R., Stephan, R., Seuberlich, T., Oevermann, A., 2015. Listeria monocytogenes infection in ruminants: is there a link to the environment, food and human health? A review. Schweiz. Arch. Tierheilkd. 157, 319–328.
Zwirzitz, B., Wetzels, S.U., Dixon, E.D., Fleischmann, S., Selberherr, E., Thalguter, S., Quijada, N.M., Dzieciol, M., Wagner, M., Stessl, B., 2021. Co-occurrence of listeria spp. and spoilage associated microbiota during meat processing due to cross-contamination events. Front. Microbiol. 12.