Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world-wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Bonnie Gale Date

Campylobacter jejuni Attribution with Multilocus Sequence Typing (MLST) and the Asymmetric Island Model

By

Bonnie Gale

Master of Public Health

Epidemiology

_____

Samuel Jenness, PhD MPH

Committee Chair

_____

Beau B. Bruce, MD PhD

Committee Member

Campylobacter jejuni Attribution with Multilocus Sequence Typing (MLST) and the Asymmetric Island Model

By

Bonnie Gale

B.A.

Macalester College

2015

Faculty Thesis Advisor: Samuel Jenness, PhD MPH

An abstract of

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of
Master of Public Health of Emory University
in Epidemiology
2018

Abstract

Campylobacter jejuni Attribution with Multilocus Sequence Typing (MLST) and the Asymmetric Island Model

By Bonnie Gale

Campylobacter is one of the leading causes of gastroenteritis in the United States, with *C. jejuni* accounting for 90% of those infections. *C. jejuni* is a widespread bacterial pathogen that is found in food, animal, and environmental sources. Using iSource software and MLST data, we can estimate the proportion and distribution of human isolates attributed to non-human sources. This was a retrospective study using 2,040 non-human *C. jejuni* isolates in North America from PubMLST and 789 human isolates from ten Foodborne Diseases Active Surveillance Network sites. We used seven constitutive genes for MLST schemes and linked them to isolate records to attribute human cases to our five categorized sources: poultry, meat, dairy environmental/animal contact, and water. The largest average proportion of human cases was attributed to poultry, accounting for 70% (95% confidence interval [CI]: 65-75%) while water had the smallest proportion (3%; 95% CI: 0–6%). The most likely order of the non-human sources that explained the most to the least number of human cases was: poultry, dairy, meat, environmental/animal contact, and water. Among the 228 isolates in our dairy source, 215 (94%) had a known pasteurization status, and all of these were raw or unpasteurized milk products from cows (215/228). Our work helps our understanding of the food and non-food sources of *Campylobacter* in North America using the genetic information of food, animal, and environmental isolates. Future attribution using just genetic information could help us refine the tools to link genetic human isolates to genetic source isolates via modeling techniques.

Campylobacter jejuni Attribution with Multilocus Sequence Typing (MLST) and the Asymmetric Island Model

By

Bonnie Gale

B.A.

Macalester College

2015

Faculty Thesis Advisor: Samuel Jenness, PhD MPH

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology

2018

Acknowledgements

I would like to thank Beau B. Bruce, my Thesis Field Advisor, for letting me take on this project and mentoring me throughout my position at the Centers for Disease Control and Prevention (CDC). I appreciate the time you took out of your busy day to answer my questions and give me feedback throughout this learning process. Another thank you to Zhaohui Cui, who answered my endless R code questions with patience. A thank you to Samuel Jenness for being my Faculty Thesis Advisor and taking on my project even though it was not in his research interests. And a final, huge thank you to all my friends and family that supported me during my two years at Rollins School of Public Health, guiding me towards my passion for public health and epidemiology.

TABLE OF CONTENTS

BACKGROUND

Campylobacter is one of the five most common foodborne illnesses in the United States, causing about 2–3 million cases per year of *Campylobacter jejuni* (*C. jejuni*) that account for 90% of all infections (1). *Campylobacter* infections typically present with sudden onset of fever, abdominal cramps, and bloody diarrhea with some cases later leading to Guillain-Barré syndrome, which is a severe, often life-threatening neurological disorder (1). *C. jejuni* infections come from a variety of sources that range from food and water to animal contact (2). Determining the distribution of human *C. jejuni* infections remains challenging; estimating the proportion of non-human isolates attributed to human illness can help us better understand the food and non-food sources of *Campylobacter*.

Few studies have explored relationships between cases of *Campylobacter jejuni* and their sources using genetic information due to a lack of standardized genotyping methods, reagents for serotyping, and a limited universal nomenclature system for *C. jejuni* (1). However, multilocus sequence typing (MLST) has been effective at characterizing several other bacteria and can be used to attribute *Campylobacter jejuni* human cases to source data (1). Using nucleotide sequences from seven loci, MLST data are curated in a standardized fashion with established protocols for shared use among international laboratories, facilitated through PubMLST (http://mlst.zoo.ox.ac.uk) (1).

Previous studies, such as one in Lancashire, England, have used MLST for *C. jejuni* source attribution (1, 2, 3). We undertook a retrospective study, compiling human *C. jejuni* case isolates throughout the United States from ten Foodborne Diseases Active Surveillance Network sites with unknown sources, to help identify what food, animal, or environmental factors are the most prevalent sources of *C. jejuni* illnesses. We

accomplished this by modeling *Campylobacter* seven gene MLST data using iSource software, which allows the prediction of the (unknown) source of human *C. jejuni* isolates with associated uncertainty (2).

METHODS

Study Design. This was a retrospective study using *Campylobacter jejuni* isolates from North America conducted with publicly-available non-human isolate data collected from PubMLST and human isolate data from ten Foodborne Diseases Active Surveillance Network sites (1). Academic research users registered with PubMLST have submitted 53,444 *Campylobacter jejuni* and *C. coli* global isolates, from cases, general outbreaks, environmental isolates, and carriers from 1980 to 2016. Isolate DNA sequencing was accomplished by PCR using standard procedures used by submitters. Those isolates outside of North America and those without labeled sources were excluded from analysis. Isolates with seven genetic loci MLST were included in the study. The final dataset included 2,829 isolates of human and non-human *C. jejuni* isolates from North America (2). There were 789 human isolates from ten Foodborne Diseases Active Surveillance Network sites with unknown sources and 2,040 non-human isolates from PubMLST with known sources.

Measures. Campylobacter isolate data were sequenced by the submitter, designating the specific strains of *Campylobacter* and providing new trimmed allele sequences, sequence typing (ST) assignment, or whole genome data on PubMLST (1). An automated submission system assessed the quality of the data submissions after a user made an

account, although their actions are tracked and they may only view, but not modify, the data (1). Curators manage submissions and can modify the data, but do not have modification rights over the entire database (1). Curators provided overall quality control of the data submitted and measures of assessment for those exporting and analyzing public data (1).

We used seven housekeeping genes for *Campylobacter* MLST schemes: aspA, glnA, gltA, glyA, pgm, tkt, and uncA (2). Linking loci sequences to isolate records allowed us to assign STs to all human and non-human isolates for further analysis (4). PubMLST data were categorized into source groups, using subject matter knowledge and the published literature (2). We used dataset variables, including STs, the seven loci per isolate, and a source identifier we created for iSource to designate which group each categorized source fell into. We combined a total of 24 different source descriptions from PubMLST into five final source groups: meat, poultry, dairy, water, and environmental/animal contact, to create categories of public health significance and ensure enough data in each category. For isolates labeled as "cattle," we classified these as either a dairy or meat source based on whether they were distinguished as dairy cattle or beef cattle, respectively, in our data. Additionally, any isolates labeled as "dairy cattle" were assumed to be from a raw or unpasteurized sources (5). Other dairy sources, such as "cows milk," that were labeled in our data as "bulk tank milk," "retail raw milk," "tanker truck milk," or "raw milk" were also assumed to be isolates from raw or unpasteurized sources after reviewing related literature and the publication breakdown of datasets on PubMLST (6).

Some parameters, e.g., evolutionary parameters, were established after we ran iSource. The evolutionary parameters connected a predictive probability to each non-human source designating a migration and mutation probability, i.e., the probability that the human isolate locus sampled is attributed to an original mutant, an identical non-human source already observed, or another non-human source category (2). This parameter explained genetic migration and mutation probabilities in our animal, food, and environmental samples.

Human case isolates were obtained from ten Foodborne Diseases Active Surveillance Network sites, identified as sporadic *C. jejuni* case isolates from 1998 to 2008. We used STs and seven constitutive genetic loci to attribute each human isolate to categorized sources using the iSource software.

Models and Analysis. Our overall objective in using a simplified form of the asymmetric island model, implemented by iSource, was to attribute human isolates of *C. jejuni* to our five source category groupings, as mentioned above (7). The asymmetric island model refers to attributed probability estimates of each human *Campylobacter* illness from known non-human MLST sources modeled via genetic mutation, migration, and recombination, independent of the distance between islands, or isolates (2). Utilizing extra information from MLST, the asymmetric island model attributes human cases from an MLST pattern not observed in any sources to a likely source of human illness by comparing the genetic resemblance to other observed types in the source groups (2). Using population genetic analyses similar to STRUCTURE (8), iSource approximates a migration matrix model, a generalized form of the asymmetric island model, and runs

Markov Chain Monte Carlo (MCMC) simulations, further accounting for variances in differentiated source populations (2, 7, 9). We analyzed output from iSource using R to provide bootstrap estimates (random sampling with replacement to ensure accuracy of sample estimates) of sources for human isolates.

RESULTS

We observed a total of 686 unique genotypes (STs) among 2,829 isolates of *C. jejuni* in our dataset, containing both human and non-human isolates. There were 789 human isolates and 2,040 non-human isolates in this dataset, with a further breakdown of non-human isolates shown in Table 1. Among the 228 isolates in our dairy source, 215 (94%) had a known pasteurization status, and all of these were raw or unpasteurized milk products from cows (215/228). The most frequent genotypes were 45 and 21, comprising 12% of non-human and human STs, while 452 genotypes were observed only once. Among non-human isolates, certain genotypes commonly presented within specific sources although they were not exclusive to one category. For example, in poultry, ST 459 was the most common genotype (81/548) but was also observed in other non-human isolates.

We attributed the 789 human-case *C. jejuni* isolates to five categories: meat, poultry, dairy, environmental/animal contact, and water. Poultry had the largest average proportion of isolates attributed to human cases (70%; 95% confidence interval [CI]: 65–75%) and water had the smallest (3%; 95% CI: 0–6%) (Figure 1; Table 2). Trace plots demonstrated model convergence (Figure 2). All 789 human case isolates were probabilistically attributed to a source category (Figure 3; Figure 4).

The most likely order of the non-human sources that explained the most, second most, third most, fourth most, and fifth most number of human cases was: poultry, dairy, meat, environmental/animal contact, and water (Table 3a). Each source had a probability that it was responsible for a given proportion of human cases, whether that was the largest proportion or the smallest proportion. Poultry explained the largest number of human cases (70%) and was responsible for 100% of the human cases attributed to the poultry source (Table 3b). However, water, the source that explains the smallest number of human cases, was responsible for 67% of the human cases attributed to water sources, 32% of the human cases attributed to environmental/animal contact sources, and 1% of the human cases attributed to meat sources (Table 3b). The most probable order of sources explaining the first-most through fifth-most number of human cases had a posterior probability of 54%, whereas a less plausible order: poultry, meat, dairy, water, and environmental/animal contact, had a posterior probability of 3% (Table 4).

Given the evolutionary parameters overall, source categories were composed of identical sample parameters without correspondence to other source categories. In poultry 90% of the migration and mutation probabilities categorized within poultry while the remainder was composed of other source categories and a novel mutant category (labeled NM) (Figure 5; Figure 6).

DISCUSSION

In our study, we found that poultry was the principal source of *C. jejuni* human illness in North America. The most likely ordering of sources from most to least number of cases was: poultry, dairy, meat, environmental/animal contact, and water. Additionally,

evolutionary parameters showed that the poultry source was strongly linked to poultry migration and mutation evolution. On the other hand, water was not strongly linked to water migration and mutation evolution and had overlapping evolution with the environmental/animal contact source. This analysis, using iSource software with the asymmetric island model, provides greater insight into what human case isolates are most commonly attributed to food and non-food sources of *Campylobacter* in North America and how we may use genetic modeling tools to help attribute human illness to non-human sources in the future.

Studies in the United States, ranging from 1982 to 2017, have found that the largest proportion of outbreaks was attributed to poultry, when not accounting for dairy (10). Moreover, in another study, the highest population attributable fraction of those cases infected with *Campylobacter* are from chicken (11). As we consistently found, poultry was the largest source of the 789 U.S. human *Campylobacter* illnesses. Out of 60 *Campylobacter* outbreaks in the United States in 2013, the largest estimated proportion of illnesses was attributed to chicken (29%) after controlling for dairy products (10). The Interagency Food Safety Analytics Collaboration (IFSAC) 2017 report stated that "An attribution percentage for Dairy is not included because, among other reasons, most foodborne Campylobacter outbreaks were associated with unpasteurized milk, which is not widely consumed, and we think these over-represent Dairy as a source of Campylobacter illness," thus excluding 116 *Campylobacter* outbreaks out of 176 (10, pg. 2). Although our dairy source was entirely raw or unpasteurized products, we maintained this as a separate source in our model. We recognized that it was unlikely that raw or

unpasteurized dairy products were widely consumed and contributed to many human *Campylobacter* cases compared to the entire burden of disease in North America.

With the substantial overlap between dairy and meat sources, one study contributed both beef cattle and dairy cattle to the non-human isolates for meat and dairy, respectively, which allowed the evolutionary parameters to provide further insight into realistic links between sources (Figure 5). Two large studies in the United States contributed dairy cattle isolates, sourcing them as cattle and only labeling them as dairy cattle in the comments, thus presenting the opportunity for overlap of source categories based on realistic factors, such as fecal matter getting into dairy milk before pasteurization. Nevertheless, keeping dairy and meat source groups separate allows us to see the distribution of human cases attributed to dairy and meat sources for better understanding of food and non-food sources of *Campylobacter* illness.

Further thinking about evolutionary parameters, the environmental/animal contact source includes the farm environment and cattle feces, but there may have been overlap between dairy cattle products (raw or unpasteurized milk) and cattle feces. Feces can carry *Campylobacter* and until the milk products are pasteurized, there is no means of killing those bacteria (5). Evolutionary parameters give us insight into what sources may overlap with each other, presenting issues of how to categorize sources and distinguish the main source categories. We may never categorize source perfectly but must recognize when there is realistic overlap. For example, petting zoos have an intersection of sources, whereby a child could acquire *Campylobacter* illness from petting a goat and that source in our dataset may be labeled as goat with an unknown meat or animal contact label, and thus be placed in the meat category instead of the environmental/animal contact

categorized source. Additionally, the evolutionary parameters in Figure 4 and Figure 5 describe the migration and mutation probabilities in animal, food, and environmental source samples, using MCMC to estimate the evolutionary parameters in each source category (2). In poultry, 90% of the migration and mutation probabilities were categorized within poultry, while the remainder was composed of other source categories and a novel mutant category (labeled NA). This category shows mutant loci not associated or accounted for in other source categories (2). The environmental/animal contact source had about 62% of its migration and mutation probabilities accounted for by environmental/animal contact, 33% by water sources, and <1% accounted for by poultry, meat, dairy sources, and novel mutant sources. This signifies that many environmental/animal contact isolates were related to the already observed source categories compared to poultry.

A matched, population-based case-control study in 2004 conducted in the United States found chicken to be the dominant food-specific risk factor for *Campylobacter* infections, based on the population attributable fraction for chicken (24%) in their multivariate analysis (11). Additionally, the adjusted odds ratios (aOR) in that study showed some of the highest attributed exposures as consuming raw milk (aOR: 4.3), eating chicken prepared at a restaurant (aOR: 2.2), eating undercooked or pink chicken (aOR: 2.1), and eating turkey prepared at a restaurant (aOR: 2.5). This study coincides with our results that the largest proportion of human *Campylobacter* illness attributed to poultry. Moreover, with 95% of the 766 *Campylobacter* isolates labeled as *C. jejuni* subspecies, this 2004 matched case-control further attributes *C. jejuni* specified cases to consumption of undercooked poultry products, whether at home or at a restaurant (11).

Comparatively, our poultry source did not distinguish home cooked or restaurant cooked poultry. Nonetheless, poultry products were highly attributed to *Campylobacter* human case isolates and more studies should be conducted to help us to better understand the food and non-food sources of *Campylobacter*.

Furthermore, an older case-control study, from 1986, suggested *Campylobacter* illness was associated most with consumption of contaminated poultry products. Specifically, the consumption of contaminated chicken contributed to half of the cases (12). With chicken being the largest consumed food category, the authors further specified the poultry category to: any turkey, any chicken, rare or raw chicken, cooked chicken, and game hens. The risk ratios (RR) specific to this outbreak and the type of food subjects consumed in the week prior to the onset of symptoms showed that rare or raw chicken (RR: 7.6) was the risk of acquiring *Campylobacter* illness (12). Further, game hens (RR: 3.3), any chicken (RR: 2.4), and cooked chicken (RR: 2.3) were in the highest end of risk. However, this study looked at *C. jejuni* and *C. coli* together as *Campylobacter jejuni/coli* enteritis and did not differentiate *C. jejuni* cases or controls with *C. coli* cases or controls. Thus, this study may not be generalizable to our study, as we do not know if majority of the poultry or chicken consumption risk were from *coli* or *jejuni Campylobacter* subspecies. Yet, over 90% of *Campylobacter* cases in the United States have been reported to be for *C. jejuni* in 2001, and though this study was conducted in 1986, *Campylobacter* subspecies distribution within the last 40 years may not have changed enough to matter (1).

Our study showed that the most likely order of sources responsible for the most, second most, third most, fourth most and fifth most number of cases was: poultry, dairy,

meat, environmental/animal contact, and water. These results varied based upon source categorization in other studies. Rare or raw chicken, rare or raw fish, game hens, any chicken, processed turkey sandwich meats, and shellfish were responsible for the largest to sixth largest RR, which also lists any beef, uncured sausage, and turkey TV dinner as having protective effects from *Campylobacter jejuni/coli* in one case-control study (12). A matched case-control study had the largest population attributable fraction for those who ate chicken at a restaurant and those who ate non-poultry meat prepared at a restaurant was second largest population attributable fraction for *Campylobacter* illness, followed by contact with animal stool (10). Further generalizability of source groups between studies can differentiate the largest proportion of human case isolates attributed to non-human sources, with non-human source data being the determining factor.

The IFSAC 2017 report, a 2004 case-control study, and a 1986 case-control study indicated that their chicken sources were the largest contributor of *Campylobacter* illness in the United States (10, 11, 12). These studies supported our results that poultry, which includes chicken, chicken offal or meat, turkey, and turkey offal or meat, constitutes the greatest proportion of human cases attributable to animal, food and environmental sources (10, 11, 12).

Limitations

The temporal range of data (1980-2016) in our dataset was not weighted in the iSource software, with more recent *Campylobacter* carriers, sporadic cases, environmental isolates, or general outbreaks having a heavier weight compared to earlier data. The iSource software and the asymmetric island model do not weight data based on when

they were inputted into the dataset. Our human case isolate data from the ten Foodborne Diseases Active Surveillance Network sites ranges from 1998 to 2008. Thus, with a time range discrepancy between our non-human and human isolate data, it is difficult to understand how more recent human isolates are attributable to non-human sources as time changes and whether different sources are more highly attributable in recent years. We would like our non-human and human isolate data ranges to overlap more to understand how current human isolates are attributable to current non-human sources. We may want to use different modeling software to weight more recent human case isolates for better attribution estimates. Conversely, older case-control studies suggested consuming contaminated poultry was highly associated with *Campylobacter* enteritis and recent data maintained the same human case illness attribution for non-human *Campylobacter* sources (12).

Another issue with comparing *Campylobacter jejuni* studies is generalizability among the representation of sources. For example, various studies have found seafood to have a high relative risk [risk ratio of 4.0], but when considering the number of cases attributable to a source in the population, as we did in this study, other studies have shown that poultry accounted for the largest proportion of human *Campylobacter* cases with non-poultry meat responsible for the second largest proportion of human *Campylobacter* cases, in contrast to our results which found poultry and then dairy to be the most frequently attributed categories (11, 12). Recognizing that source categorization varies across studies, generalizability is limited when it comes to the sources responsible for the largest proportion of human *Campylobacter* illness with the available data. Further, studies may not categorize food, animal, or environmental factors in similar

fashion, which continues to create gaps in comparing studies on human *Campylobacter* attribution. For example, we categorized "dairy cattle" as a dairy source, which was originally labeled as cattle, but could be placed into a meat or animal source in other studies. The IFSAC 2017 report categorizes chicken and other meat/poultry as separate sources, thus complicating the opportunity to compare studies on *C. jejuni* for source attribution (10).

Conclusions

Our work elucidates the food and non-food sources of *Campylobacter* in North America using the genetic information of food, animal, and environmental isolates instead of looking at case-control or outbreak data. Future attribution using just genetic information could help us refine the tools to link genetic human isolates to genetic source isolates via modeling techniques. The move towards national whole genome sequencing of human cases will help us account for differences in non-human sources, refine the accuracy of unknown sources, and further unknown human source attribution to food, animal, or environment sources using genetic information.

REFERENCES

1. Dingle KE, Colles FM, Wareing DRA, et al. Multilocus Sequence Typing System for Campylobacter jejuni. *Journal of Clinical Microbiology*. 2001; 39(1):14–23.
2. Wilson DJ, Gabriel E, Leatherbarrow AJ, et al. Tracing the Source of Campylobacteriosis. PLoS Genet. 2008; 4(9): e1000203.
3. Miller P, Marshall J, French N, et al. SourceR: Classification and source attribution of infectious agents among heterogeneous populations. *PLOS Computational Biology*. 2017;13(5): 1005564.
4. Jolley KA, Maiden MCJ. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics*. 2010;11(1): 595.
5. Division of Foodborne, Waterborne, and Environmental Diseases (DFWED), National Center for Emerging and Zoonotic Infectious Diseases (NCEZID), Centers for Disease Control and Prevention. "Campylobacter (Campylobacteriosis)." https://www.cdc.gov/campylobacter/faq.html. Updated October 2, 2017. Accessed March 26, 2018.
6. UC Davis Western Institute for Food Safety & Security. "FDA's Milk Residue Study and Our Food Safety." https://www.wifss.ucdavis.edu/fdas-milk-residue-study-and-food-safety/. Published March 5, 2015. Accessed March 22, 2018.
7. Hald T, Vose D, Wegener HC, et al. A Bayesian Approach to Quantify the Contribution of Animal-Food Sources to Human Salmonellosis. *Risk Analysis*. 2004;24(1), 255-269.

8. Porras-Hurtado L, Ruiz Y, Santos C, et al. An overview of STRUCTURE: applications, parameter settings, and supporting software. *Frontiers in Genetics*. 2013; 4:98.
9. Bodmer WF, Cavalli-Sforza LL. A Migration Matrix Model for the Study of Random Genetic Drift. *Genetics*. 1967; 59:565-592.
10. Interagency Food Safety Analytics Collaboration (IFSAC). Foodborne illness source attribution estimates for 2013 for *Salmonella*, *Escherichia coli* O157, *Listeria monocytogenes*, and *Campylobacter* using multi-year outbreak surveillance data, 1998-2013. GA and D.C., United States: U.S. Department of Health and Human Services, CDC, FDA, USDA-FSIS. 2017.
11. Friedman CR, Hoekstra RM, Samuel M, et al. Risk Factors for Sporadic Campylobacter Infection in the United States: A Case-Control Study in FoodNet Sites. *Clinical Infectious Diseases*. 2004;38(3): S285-S296.
12. Harris NV, Weiss NS, Nolan CM. The role of poultry and meats in the etiology of Campylobacter jejuni/coli enteritis. *American Journal of Public Health*. 1986;76(4), 407-411.

TABLES

Table 1: Distribution of source isolates in each category, before model attribution (N = 2,829).

Source	Human Cases	Meat	Poultry	Dairy	Water	Environmental/Animal Contact
Number of isolates	789	479	548	235	445	333

Table 2: Summaries of the posterior distribution of each source, including the mean, median, standard deviation and centered 95% interval.

	Meat*	Poultry#	Dairy†	Water‡	Environmental/Animal Contact**
Mean	0.091	0.698	0.144	0.027	0.040
Median	0.090	0.698	0.144	0.025	0.039
Standard Deviation	0.024	0.026	0.024	0.015	0.020
2.5%	0.046	0.646	0.098	0.003	0.005
97.5%	0.141	0.747	0.191	0.059	0.084

Footnotes:

*Includes beef offal or meat, cattle, lamb, goat, pig and sheep

#Includes chicken offal or meat, chicken, turkey and turkey offal or meat

†Includes goat's milk, cow's milk, and dairy cattle

‡Includes environmental waters

**Includes calf feces, cattle feces, farm environment, farm slurry, dog, marmoset, duck, goose, starling, wild birds and other animals

Table 3a: The most likely order of sources responsible for the most, second most, third most, etc. number of cases.

Source:	Poultry#	Dairy†	Meat*	Environmental/Animal Contact‡	Water**

Footnotes:

*Includes beef offal or meat, cattle, lamb, goat, pig and sheep

#Includes chicken offal or meat, chicken, turkey and turkey offal or meat

†Includes goat's milk, cow's milk, and dairy cattle

‡Includes environmental waters

**Includes calf feces, cattle feces, farm environment, farm slurry, dog, marmoset, duck, goose, starling, wild birds and other animals

Table 3b: Probability that each source is responsible for the most, second most, third most, etc. number of cases.

	[, 1]	[, 2]	[, 3]	[, 4]	[, 5]
Meat	0.000	0.114	0.819	0.063	0.003
Poultry	1.000	0.000	0.000	0.000	0.000
Dairy	0.000	0.885	0.114	0.001	0.000
Water	0.000	0.000	0.013	0.318	0.669
Environmental/Animal Contact	0.000	0.000	0.055	0.617	0.328

Footnotes:

*Includes beef offal or meat, cattle, lamb, goat, pig and sheep

#Includes chicken offal or meat, chicken, turkey and turkey offal or meat

†Includes goat's milk, cow's milk, and dairy cattle

‡Includes environmental waters

**Includes calf feces, cattle feces, farm environment, farm slurry, dog, marmoset, duck, goose, starling, wild birds and other animals

Table 4: Most probable orderings and their posterior probabilities.

Most likely Source Sequence	First Source	Second Source	Third Source	Fourth Source	Fifth Source	Posterior Probability
1	Poultry	Dairy	Meat	Environmental/ Animal Contact	Water	53.71%
2	Poultry	Dairy	Meat	Water	Environmental/ Animal Contact	28.18%
3	Poultry	Meat	Dairy	Environmental/ Animal Contact	Water	0.08%
4	Poultry	Dairy	Environmental/ Animal Contact	Meat	Water	5.09%
5	Poultry	Meat	Dairy	Water	Environmental/ Animal Contact	3.42%

Footnotes:

*Includes beef offal or meat, cattle, lamb, goat, pig and sheep

#Includes chicken offal or meat, chicken, turkey and turkey offal or meat

†Includes goat's milk, cow's milk, and dairy cattle

‡Includes environmental waters

**Includes calf feces, cattle feces, farm environment, farm slurry, dog, marmoset, duck, goose, starling, wild birds and other animals

FIGURES

Figure 1: Histograms of the proportion of isolates attributable to animal, food and environmental sources.

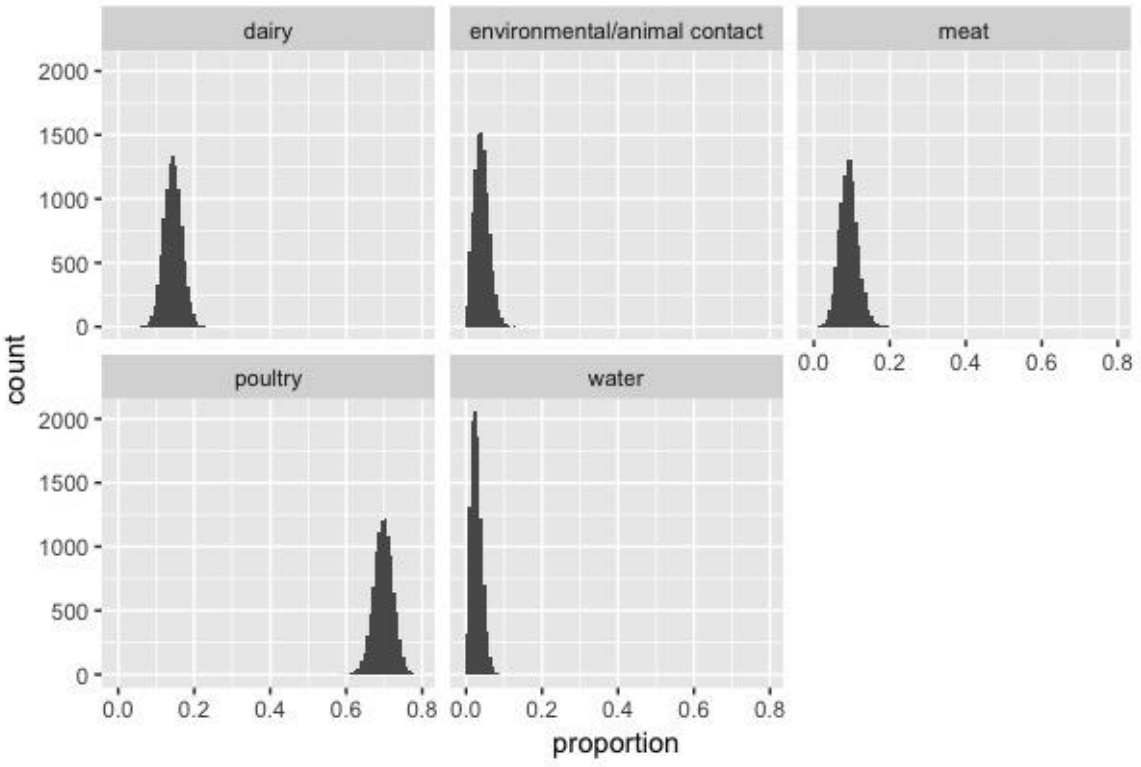

Figure 2: Trace plots of the proportion of isolates attributable to animal, food and environmental five sources demonstrating model convergence.

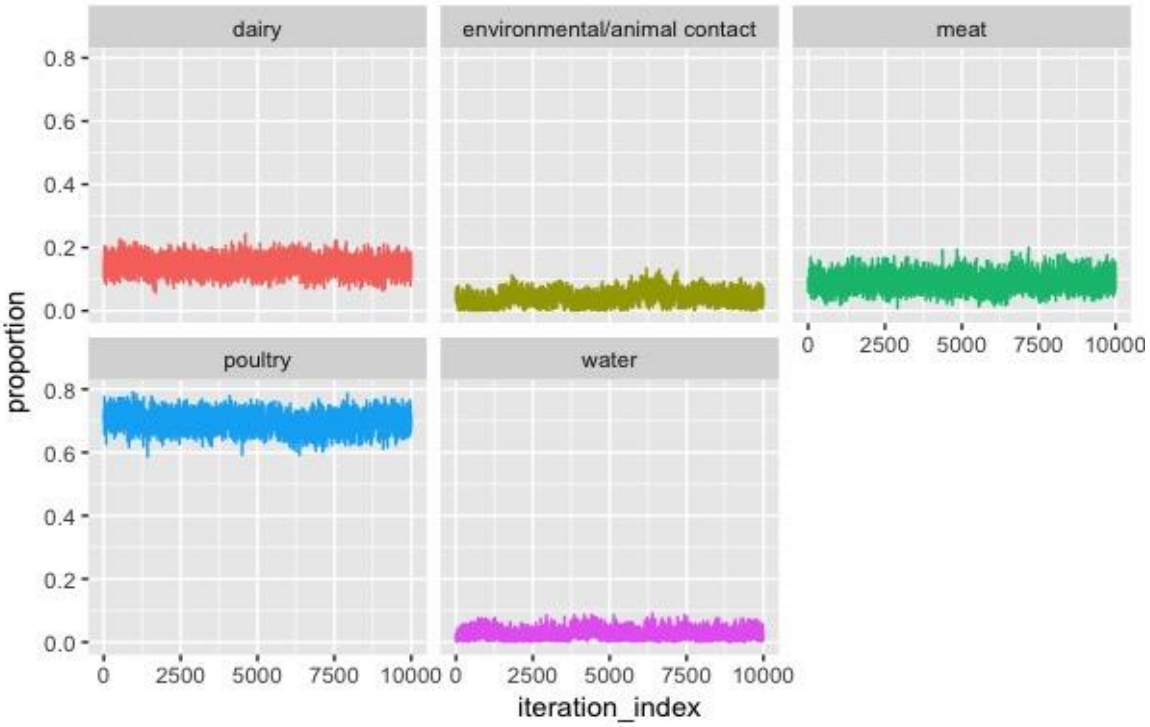

Figure 3: Estimated proportion of human cases attributable to animal, food and environmental sources. Error bars denote the 95% confidence interval for each source.

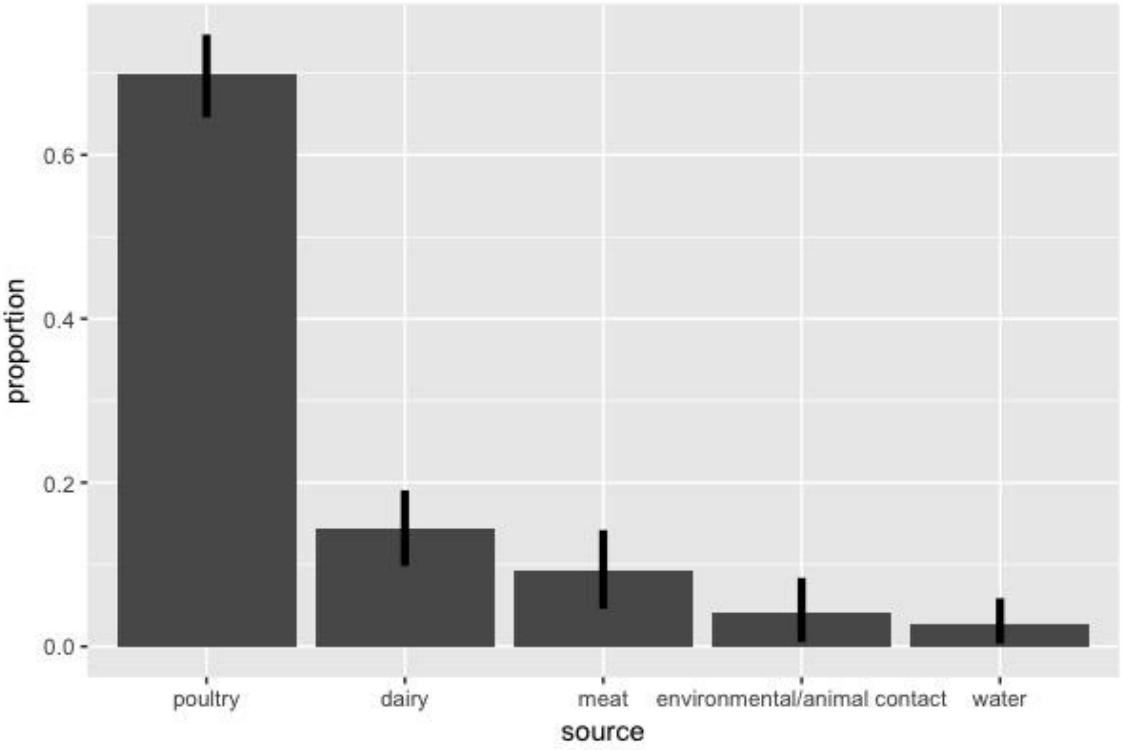

Figure 4: Probability of source for human cases (N = 789). The source probability for 789 human case isolates is denoted for Dairy (red), Environmental/Animal Contact (olive green), Meat (green), Poultry (blue), and Water (purple). These isolates have been ordered horizontally to assist visualization.

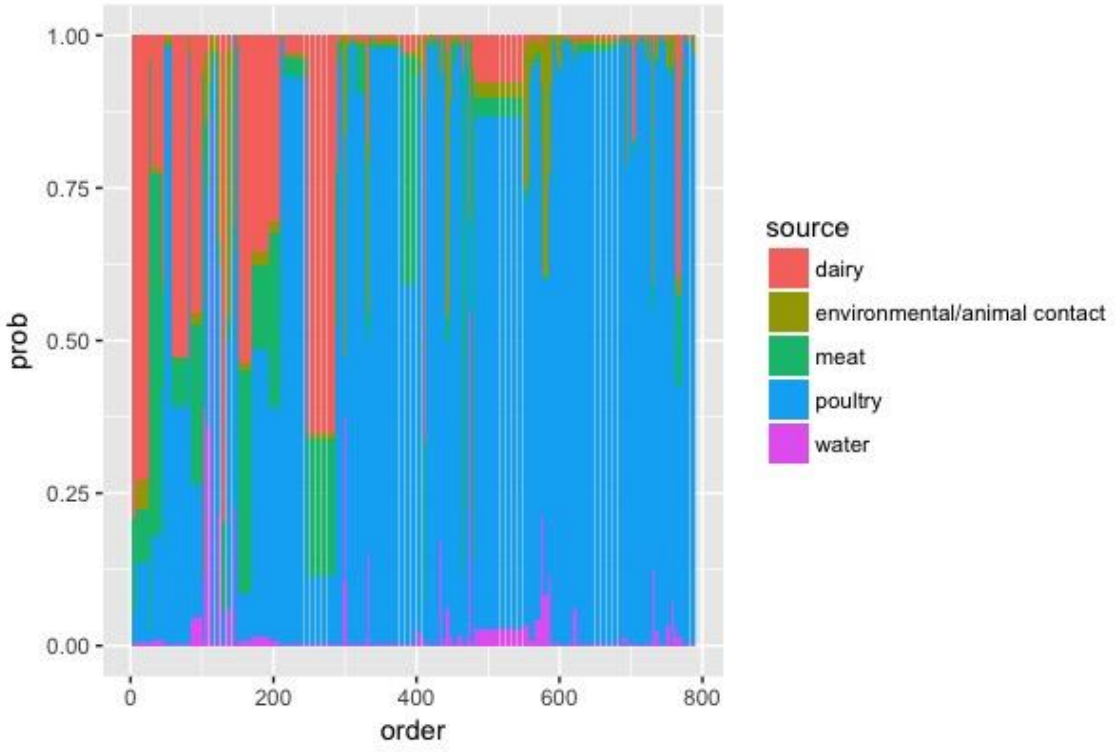

Figure 5: Migration and mutation probabilities in animal, food, and environmental samples, tracing evolutionary parameters using MCMC. Each *C. jejuni* source has a predicative probability that a newly-sampled locus is an original mutant (black line) or identical to a source already observed or another source category (colored lines: Meat-red, Poultry-lime green, Dairy-green, Water-blue, Environmental/Animal Contact-purple).

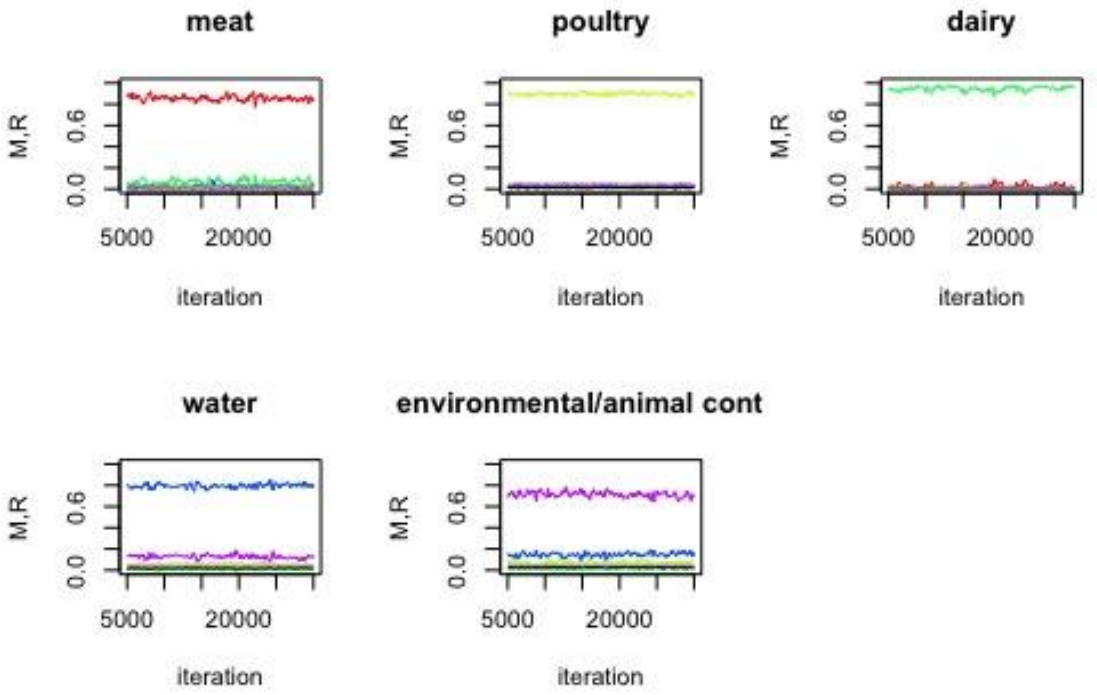

Figure 6: Migration and mutation probabilities in animal, food, and environmental samples, tracing evolutionary parameters using MCMC. Each *C. jejuni* source has a predicative probability that a newly-sampled locus is a novel mutant (NA row) or identical to a source already observed or another source category.