Source attribution of campylobacteriosis in Australia circa 2018

Angus McLure (angus.mclure@anu.edu.au)
The Australian National University
James J Smith
Queensland Health
Simon Matthew Firestone
The University of Melbourne
Martyn D Kirk
The Australian National University
Nigel French
Massey University
Emily Fearnley
Government of South Australia
Rhiannon Wallace
Agriculture and Agri-Food Canada
Mary Valcanis
The Doherty Institute for Infection and Immunity
Dieter Bulach
The Doherty Institute for Infection and Immunity
Cameron Moffatt
The Australian National University
Linda A Selvey
University of Queensland
Amy Jennison
Queensland Health
Kathryn Glass
The Australian National University

Research Article

Keywords: Campylobacter, source attribution, foodborne disease, gastroenteritis, disease reservoir, zoonosis, Bayesian analysis

Posted Date: June 14th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1743613/v1

Source attribution of campylobacteriosis in Australia circa 2018

Angus McLure*[1], James J Smith[2,3], Simon Matthew Firestone[4], Martyn D Kirk[1], Nigel French[5,6], Emily Fearnley[7], Rhiannon Wallace[8], Mary Valcanis[9,10], Dieter Bulach[9,11], Cameron Moffatt[1], Linda A Selvey[12], Amy Jennison[13], Danielle M Cribb[1], Kathryn Glass[1]

*Corresponding author: angus.mclure@anu.edu.au

[1] National Centre for Epidemiology and Population Health, The Australian National University, Canberra, Australia

[2] Food Safety Standards and Regulation, Health Protection Branch, Queensland Health, Brisbane, Australia

[3] School of Biology and Environmental Science, Faculty of Science, Queensland University of Technology, Brisbane, Australia

[4] Faculty of Veterinary and Agricultural Sciences, The University of Melbourne, Melbourne, Australia

[5] Infectious Disease Research Centre, Hopkirk Research Institute, Massey University, Palmerston North, New Zealand

[6] New Zealand Food Safety Science and Research Centre, Hopkirk Research Institute, Massey University, Palmerston North, New Zealand

[7] Department for Health and Wellbeing, Government of South Australia, Adelaide, Australia

[8] Agassiz Research and Development Centre, Agriculture and Agri-Food Canada, Agassiz, Canada

[9] The Doherty Institute for Infection and Immunity, Melbourne, Australia

[10] Microbiological Diagnostic Unit Public Health Laboratory, The University of Melbourne

[11] Melbourne Bioinformatics, The University of Melbourne, Melbourne, Australia

[12] School of Public Health, Faculty of Medicine, The University of Queensland, Brisbane, Australia

[13] Public Health Microbiology, Forensic and Scientific Services, Queensland Health, Brisbane, Australia

Abstract

Background: *Campylobacter* spp. infections are the leading cause of foodborne gastroenteritis in high-income countries, including Australia. *Campylobacter* colonises a variety of mammalian and avian hosts that are reservoirs for human campylobacteriosis. Though most Australian outbreak investigations implicate chicken meat, the proportions of sporadic cases attributable to different animal reservoirs are unknown.

Methods: *Campylobacter* isolates were obtained from notified human cases, and raw meat and offal from the major livestock in Australia: chickens, pigs, and ruminants (cattle and sheep) between 2017 and 2019. Isolates were speciated, with sequence types determined using multi-locus sequence genotyping. We used Bayesian source attribution models to estimate the proportion of human cases attributable to each livestock source by comparing the frequency of sequence types in cases and each animal source. We employed a model comparison approach with ten base models and explored adjusting these for age, gender, jurisdiction, rurality, and season. Four of the ten base models included an 'unsampled' source to estimate the proportion of cases attributable to wild, feral, or domestic animal reservoirs not sampled in our study.

Results: We included 612 food and 710 human case isolates. The best fitting models attributed >80% of *Campylobacter* cases to chickens, with a greater proportion of *Campylobacter coli* (>84%) than *Campylobacter jejuni* (>77%). The best fitting model that included an unsampled source attributed 14% (95% CrI: 0.3-32%) to the unsampled source and only 2% to ruminants (95% CrI: 0.3-12%) and 2% to pigs (95% CrI: 0.2-11%.) The best fitting model that did not include an unsampled source attributed 12% to ruminants (95%CrI: 1.3-33%) and 6% to pigs (95%CrI: 1.1-19%.) Model fit was not improved by inclusion of case covariates.

Conclusions: Chickens were the leading source of *Campylobacter* infections in our data and should remain the focus of interventions to reduce the burden in Australia.

Keywords: *Campylobacter*; source attribution; foodborne disease; gastroenteritis; disease reservoir; zoonosis; Bayesian analysis

Background

Campylobacter infections are a major cause of zoonotic foodborne gastroenteritis worldwide and the leading cause in high income countries [1], with a global burden of over 166 million cases and 3.7 million disability adjusted life years circa 2010 [2]. In Australia, foodborne *Campylobacter* causes an estimated 179,000 cases and 3,200 hospitalisations each year [3]. Although the genus *Campylobacter* comprises 41 species (as of June 2022) [4], two species, *C. coli* and *C. jejuni,* cause >90% of human campylobacteriosis cases [5]. The most widely used sub-species classification for *Campylobacter* research is seven-gene multi-locus sequence typing (MLST) [6]. *C. coli and C. jejuni* are typed in the same scheme and this '*Campylobacter*' scheme contains more than 11,800 sequence types (STs) [7]. A range of warm-blooded animals — including wild bird species and food production animals such as chickens, pigs, sheep, and cattle — can act as *Campylobacter* reservoirs. Human cases are primarily foodborne [8] and occasionally waterborne [5], with some direct zoonotic transmission [9] and rare person-to-person transmission [10]. In Australia, most *Campylobacter* infections are sporadic, i.e. not associated with recognised outbreaks [5, 11], making it difficult to identify the source of infection for most individual cases.

Bayesian source attribution models (e.g. [12-16]) have been used to estimate the proportion of human campylobacteriosis attributable to sources by comparing the relative abundance of *Campylobacter* STs observed in sources and cases. Some methods attempt to account for the relative transmissibility and virulence of different strains (e.g. [12, 14, 15]), while others attempt to model organism genetic recombination and mutation to improve estimates of relative abundance of rare types in sources (e.g. [13, 16]). In many source attribution studies, including the present study, the sources of interest are animal reservoirs (e.g. chickens, pigs) rather than individual classes of food products, risk factors for infection, transmission routes, or systemic food safety failures. Nevertheless, source attribution estimates have been used to identify the primary animal reservoirs, inform effective food safety policy and interventions [17, 18], and identify human subpopulations with differing patterns of attribution [13, 19].

The aim of this study was to estimate the proportion of *Campylobacter* infections attributable to chickens, pigs, and ruminants (cattle and sheep) in Australia circa 2018. We compared attribution estimates across a range of modelling frameworks, including models that considered a fourth 'unsampled' source to reflect potential reservoirs for which no data were available.

Methods

Data collection

Study data were collected as part of the broader CampySource project, a collaboration between Australian academic institutions, government agencies, and industry. A detailed description of study methods and data collection can be found elsewhere [20-24]. The sequence readset for each food isolate (Bioproject Accession: PRJNA591966), case isolates sampled as part of the case-control study (Bioproject Accession: PRJNA592186) and the case isolates included in the national snapshot (Bioproject Accession: PRJNA560409) are available through GenBank [25].

Detailed sampling methods for cases have been described previously [20, 23, 24]. We defined a case as a person with acute diarrhoea where *Campylobacter* spp. was cultured from stool. Cases were identified through a combination of state or regional notifiable disease surveillance systems and pathology service databases. We collected and sequenced 531 *Campylobacter* isolates from human cases between February 2018 and October 2019 in three Australian jurisdictions: The Hunter New England (HNE) public health district of New South Wales (NSW), Queensland (Qld), and the Australian Capital Territory (ACT). We collected and sequenced an additional 184 human isolates from Victoria (Vic), Western Australia (WA), Tasmania (Tas), South Australia (SA), and the Northern Territory (NT) over the period 2017 to 2019. Our case isolates included 164 collected from across all these jurisdictions over a shorter period (October 2018—February 2019) as part of national snapshot of campylobacteriosis [24]. After removing five isolates with indeterminate MLST assignments (two from Vic and three from Qld), 710 human case isolates were included in the analysis. Cases from three jurisdictions (HNE, Qld, and ACT) were interviewed regarding potential exposures for a case-control study [20]. We considered responses on age, gender, rurality of residence, notification date, and jurisdiction as potential model covariates.

Detailed sampling methods for food products have been described previously [21, 22]. Jurisdictional health departments sampled retail meat products for *Campylobacter* testing between October 2016 and March 2019 in the ACT (Canberra), NSW (Hunter Valley and Greater Sydney regions), Qld (Brisbane, Toowoomba, Rockhampton, Townsville, and Cairns hospital and health districts), and Vic (Bendigo and Melbourne; chicken only), with isolates collected after March 2017 considered for sequencing. Samples included pre-packaged (fresh or frozen) meats and delicatessen products. The prevalence of *Campylobacter* spp. contamination was expected to be low for muscle meat from pigs and ruminants [26]. Therefore, we sampled offal (kidney and liver) to maximise the total number of isolates from these sources. For chicken, a combination of muscle meat (with and without skin or bones) and offal (giblet and liver) products were sampled. We attempted to sequence all isolates from pig and ruminant samples, but due to budgetary constraints only sequenced a subsample of isolates from chicken, aiming for a total of 500 isolates from this source. In Qld, chicken *Campylobacter* spp. isolates were subsampled for genotyping to optimise coverage using a judgemental-stratified approach across the following strata: region (hospital and health district), *Campylobacter* species (*C. jejuni*, *C. coli*), abattoir/processor, and time (calendar year and quarter).

Campylobacter isolation, sequencing, and genotyping

We isolated and confirmed *C. coli* and *C. jejuni* from the meat and offal samples according to ISO 10272–1:2017 [27] and AS 5013.06.2015 [28] with minor modifications [21]. Isolation and genomic analysis of isolates from food [22] and clinical specimens [24] has been described in detail elsewhere. Briefly*, C. coli* and *C. jejuni* isolates were grown from patient faecal samples, with samples stored at 2–8°C and processed for *Campylobacter* culture within 48 h of collection. We extracted *Campylobacter* DNA from food and patient isolates using the QiaSymphony DSP DNA Mini kit (Qiagen) according to manufacturer's instructions. We used the Nextera XT DNA Library Prep kit (Illumina, San Diego, CA, USA) to prepare DNA for whole genome sequencing which was performed on the Illumina Next-Seq500 with 150 base-pair paired-end reads using the NextSeq 500 Mid Output kit (300 cycles) (Illumina). MLST was

performed on *de novo* assembled contigs, searching with a BLAST-based tool [29] against the PubMLST allele database [7].

Source attribution modelling approach

We employed a generalisation of existing Bayesian source attribution methods [12, 13] to estimate attribution proportions, include covariates for the cases, and adjust for differences between types, as applied previously [30]. We modelled the proportion of human cases in subpopulation s attributable to transmission from each source j (ξ_{js}) given the number of cases in each subpopulation s due to each pathogen type i (Y_{is}), the number of isolates of each type observed in each putative source (X_{ij}), and weights for the relative exposure of humans to these putative sources w_j. The proportion (θ_{ijs}) of cases in subpopulation s that were due to pathogen type i from source j was modelled as:

$$\theta_{ijs} \propto a_{js} w_j \, r_{ij} \, q_i$$

with constraints $\sum_{i,j} \theta_{ijs} = 1$ and $\sum_i r_{ij} = 1$, where a_{js} was the ability of source j to act as a source of infection for group s, r_{ij} was the relative abundance of type i in source j, and q_i was the relative ability of subtype i to transmit from a source and lead to a reported case (which we call *type transmission potential*). The proportion of cases in subpopulation s attributed to source *j* was:

$$\xi_{js} = \sum_i \theta_{ijs} \propto a_{js} \, w_j \sum_i r_{ij} \, q_i,$$

while the proportion of cases due to each type, μ_{is}, was:

$$\mu_{is} = \sum_j \theta_{ijs} \propto q_i \sum_j a_{js} \, w_j \, r_{ij}.$$

We considered two models for the relative abundance of sequence types in sources (r_{ij}). The first was the Dirichlet model of Liao et al. [13], which adopts independent symmetric Dirichlet priors for the relative abundance of sequence types in sources and models the numbers of isolates of each type observed in each source (X_{ij}) as independent multinomial distributions:

$$p(r_{\cdot j}) \sim Dirichlet(\alpha, \alpha, \ldots, \alpha),$$
$$p(X_{\cdot j} | r) \sim Multinomial(r_{\cdot j}).$$

The second approach was the asymmetric island model, first proposed for source attribution by Wilson et al. [16] and developed further and implemented in an R package, *islandR* by Liao et al. [13]. The asymmetric island model uses the observed number of MLST types and frequency of alleles at each locus to estimate mutation rate (new allele generation), recombination (new sequence generation types from novel allele combinations extant in the sources), and transmission between sources, to estimate the relative abundance of all types in each source (r_{ij}).

Parameters were estimated either by joint inference of all parameters (joint Dirichlet model), or in two steps (two-step Dirichlet and asymmetric island model). In two-step approaches, the relative abundance of sequence types in each source (r_{ij}) was estimated first, with all other parameters then estimated repeatedly using draws from the posterior distribution of r_{ij}. In estimating the remaining parameters, the transmission potential of each type (q_i) and the exposure weights (w_j) were assumed to be the same for each sub-population s, but the ability of each source to transmit to humans (a_{js}) was allowed to vary and modelled as:

$$a_{js} = \exp\left(\sum_n F_{sn}\beta_{nj}\right),$$

where F was a model matrix defining a linear predictor based on binary, categorical, or ordinal covariates for each subgroup s of the cases; and β was a matrix of parameters for each source j. A reference source was assigned, and the associated parameters in the matrix β were fixed to 0, while the remaining parameters were given unit normal priors.

The number of human cases in subpopulation s due to pathogen type i, were modelled as independent multinomial variables, i.e. $p(Y_{\cdot s}|\mu) \sim Multinomial(\mu_{\cdot s})$. The type transmission potential terms, q_i, were constrained with a log-normal prior:

$$p(q_i|\sigma) \sim logNormal(0, \sigma^2),$$
$$p(\sigma) \sim HalfCauchy(0, 5).$$

Since *Campylobacter* is primarily foodborne [8], the exposure weights w_j, were approximated by the relative exposure to contaminated food products derived from each source, modelled as $w_j = M_j\, k_j$, where M_j was the apparent consumption (per capita, per year) of food derived from source j [31], and k_j was the prevalence of the pathogen in muscle meat derived from source j. The prevalence of *Campylobacter* in the muscle meat of each source j, was modelled with the binomial model and a flat prior: $p(P_j|N_j, k_j) \sim Bin(N_j, k_j)$, $p(k_j) \sim Beta(1,1)$, where N_j was the number of total tests and P_j the number of positive tests. Data for prevalence in chicken muscle meat was derived from the CampySource study [21]. CampySource collected offal rather than muscle meat samples for pigs and ruminants. Since *Campylobacter* prevalence was likely to be higher on offal than muscle meat, and Australians consume relatively little offal, we estimated prevalence in these meats from a separate survey [26].

When employing the Dirichlet model for the relative abundance of sequence types in sources, an 'unsampled source' was modelled by including an additional source j^* without any observed samples: i.e. $X_{ij^*} = P_{j^*} = N_{j^*} = 0$.

The Modified Hald model proposed by Müllner et al. [12] can be seen as a special case of the modelling framework presented here, using the Dirichlet model of relative abundance of sequence types in sources and no covariates for cases. The attribution model proposed by Liao et al. [13] can also be seen as a special case of our framework, using the asymmetric island model for the prevalence of types in sources and assuming all type transmission potential terms, q_i, were equal to 1.

Source attribution models

We considered ten base models (Table 1), varying the assumptions for the type transmission potential terms and the model for the relative abundance of sequence types in sources (Dirichlet versus asymmetric island). With the asymmetric island models (M7 and M10), we assumed the mutation and recombination rates of *Campylobacter* were the same across sources. With the Dirichlet models we compared models with and without an 'unsampled source' and primarily used a flat Dirichlet (1,1,…,1) prior for the relative abundance of sequence types in sources. However, we used an even less informative Dirichlet (0.1, 0.1,…,0.1) prior as a sensitivity analysis.

We then included covariates (age, rurality, gender, jurisdiction and season) into each of the base models. We used four age categories (0-4, 5-18, 19-64, and 65 and over). Rurality was categorised as urban or rural, with those reporting residence in an inner city, urban, suburban, or town area categorised as urban, and those reporting residence in rural or remote areas categorised as rural. Season was classified by the date the case was reported as: Summer (December to February), Autumn (March to May), Winter (June to August) and Spring (September to November). Cases with a missing value for gender, age, rurality, jurisdiction, or season were excluded only from analyses involving the missing covariates.

To compare the risk associated with different animal sources we calculated a quantity we call the *relative attributable proportion* (RAP). For each source the RAP was estimated by the proportion of cases attributed to that source divided by the domestic annual consumption of meat products from that source [31], normalised against one of the sources as a reference. We chose the most commonly consumed meat product in Australia — chicken — as the reference source, which therefore had a RAP of one.

Implementation

All analyses were conducted in the R software environment [32], with data cleaning and visualisations using *tidyverse* packages [33] and *ggVennDiagram* [34]. In the two-step Dirichlet and asymmetric island models, the relative abundance of sequence types in sources was estimated using the R package, *islandR*, created by Liao et al. [13]. Inference for the asymmetric island model was done with Markov Chain Monte Carlo (MCMC) using the Metropolis-Hasting Algorithm, with 1,000 iterations of warmup and thinning post-warmup draws to one in every 100 iterations. Using 100 draws from the posterior distribution of the relative abundance of sequence types in sources, inference for remaining parameters for all models was performed using Hamiltonian MCMC using the No U-Turn Algorithm implemented with the Stan language [35] via the R package, *Rstan* [36]. The Hamiltonian MCMC step used four independent chains, 2,000 warmup iterations, and 100 post-warmup draws, for a total of 40,000 post-warmup draws per model. In fully joint models, all inference was conducted using Hamiltonian MCMC with four chains, 2,000 warmup iterations, and 2,000 post-warmup draws, for a total of 8,000 post-warmup draws per model. Convergence checking for Hamiltonian MCMC was done using the 'R hat' statistic [37] across the four chains.

Model posterior predictive fits were compared using the Watanabe-Akaike information criterion (WAIC) [38] using the R package, *loo* [39], with a difference greater than five

standard errors considered to be substantive evidence of superior model predictions. The WAIC uses log-likelihood of each datapoint averaged across the posterior to assess how well models match points in the training data. For comparing joint models to one another (including models with and without covariates for cases) each isolate from either a case or source was considered a datapoint for calculating WAIC. However, for models where inference was conducted in two steps, only the data from the second step (i.e. case isolates) could be considered and the WAIC on these data was the primary measure of model fit used to compare across all models.

Ethics approval and consent to participate

All methods were carried out in accordance with a protocol approved by the Australian National University (ANU) Human Research Ethics Committee (Protocol: 2016/426) and other partner institutions' ethics committees, and in accordance with ANU guidelines. All cases (or their parent or legal guardian in the case of children under 18) provided informed consent to participate in the study.

Results

After removing nine isolates with indeterminate ST assignments (five case isolates and four chicken isolates), the final dataset comprised genotyped isolates from 710 human cases, 480 chicken meat and offal samples, 88 ruminant (sheep and cattle) offal samples, and 44 pig offal samples. *C. coli* was more common in chickens (59%, 283/480) and pigs (64%, 28/44) and *C. jejuni* more common in human cases (82%, 585/710) and ruminants (82%, 72/88). The 1,322 isolates represented 175 different STs, 74 *C. coli* and 101 *C. jejuni*. STs shared across multiple sources, or sources and cases, were usually more common than those found only in cases or a single source. While 66 of 118 human case STs were not found in any sampled food source (18/31 *C. coli* STs and 48/87 *C. jejuni* STs), these types only accounted for 19% of all isolates from cases. Conversely the eight STs found in cases and every source (STs 21, 42, 50, 538, 827, 1181, 2083, and 7323) together accounted for 31% (221/710) of isolates from cases, 47% (41/88) of isolates from ruminants, 30% (146/480) of isolates from chickens, and 34% (15/44) of isolates from pigs. Fifteen *C. coli* STs (18 isolates) and two *C. jejuni* STs (two isolates) were found in pigs but no other sources. Three *C. jejuni* STs (three isolates) were only found in ruminants but no other sources. Similarly, 29 *C. coli* STs (101 isolates) and 31 *C. jejuni* STs (91 isolates) were only found in chicken (Figure S1 and Tables 2, S1, and S2).

In almost all models, chickens were estimated to be the most common source for both *C. jejuni* (9/10 models) and *C. coli* (all models). The proportion of cases attributed to chickens was lower for *C. jejuni* than for *C. coli* (Figure 1). Dirichlet models with an unsampled source had reduced attribution to ruminants and pigs but similar attribution to chickens in most models. Performing inference jointly compared to in two-steps resulted in only slightly higher attribution to chickens and unsampled sources and reduced attribution to ruminants and pigs.

Including covariates for cases in the joint attribution models (M1-M4) did not lead to substantive improvements in model posterior predictions as measured with WAIC. Although the point estimates for the proportions of *C. coli* and *C. jejuni* cases attributed to chickens was lower in rural than urban areas for all models except M10, the 95% credible interval for the differences included zero (no difference; Figure S3). This absence of substantive improvement in model fit was observed across all covariates (gender, age group, jurisdiction, and season)

and models. As models with covariates could only be fit to the subset of the cases enrolled via the case-control study (n = 531), models without covariates were preferred.

Our data indicate transmission potential (ability to transmit and cause disease) varied between sequence types in our study. For instance, ST48 was more common among cases (4.2%) than in any of the sources (ruminants: 1.1%, chickens: 0.8%, and pigs: 0.0%), indicating high transmission potential (Table S3). Conversely, ST827 was more common in every source (ruminants: 9.1%, chickens: 9.6%, and pigs: 4.5%) than in cases (1.8%) (Table S4), and ST832 accounted for 3.3% and 2.3% of chicken and pig isolates but none of the cases, indicating low transmission potential. Models allowing transmission potential to vary between STs (M1-M7) had a much better fit to the relative abundance of STs in cases (Figure S4) and substantially better WAIC values than models that assumed that all STs had equal transmission potential (M8-M10) (Table S5). These models with variable transmission potential (M1-M7) attributed more cases to chicken meat and the unsampled source and fewer cases to ruminants compared to those that assumed all STs had equal transmission potential (M8-M10) (Figure 1). *C. coli* STs were generally estimated to have lower transmission potential than *C. jejuni* STs, although there was substantial variation within each species (Figure 2). STs isolated from cases but none of the animal meat sources (e.g. ST2398) were estimated to occur rarely in the sampled sources but have a relatively high transmission potential. When sequence type transmission potential and relative abundance of STs in sources were estimated in two steps (M5-M7), transmission potential was more variable between STs than in models where all parameters were jointly estimated (M1-M4). When comparing models, estimates of transmission potential and abundance varied most for those types that were found in cases but not in sources (e.g. ST2398). In joint Dirichlet models, using a weaker prior or including an unsampled source increased estimates of the relative abundance of these types in sources (Table S4) and correspondingly reduced the estimates of their transmission potential (Table S6). For the Dirichlet models, transmission potential varied at most by a factor of 220 across sequence types and varied by less than a factor of 45 across the most common sequence types in cases (Figure 2 and Table S6). However, under the asymmetric island model, types like ST2398 (observed in cases but not sources) were estimated to have a trillion times greater transmission potential than other types (Figure 2 and Table S6), which is implausible.

The joint Dirichlet model with variable type transmission potential had very good convergence metrics (Rhat<1.008). When type transmission potential terms were set to be equal, the joint Dirichlet model failed to converge (Rhat>>1), further highlighting the importance of including type transmission potential terms. Two-step inference with variable or equal type transmission potential converged adequately (Rhat<1.04).

Comparison of observed and estimated relative abundance of sequence types in sources (Tables S1 and S2) suggested the default priors for these quantities in the Dirichlet model may be too strong, with better concordance under a weaker prior. Use of a weaker prior for the Dirichlet model lead to substantively poorer model fit to case data as measured with WAIC (Table S5). However, the weaker prior improved fit to source data such that the WAIC calculated over cases and sources was lower (though not substantively) for models with a weaker prior (Table S5). Using the weaker prior had little effect on the attribution proportions for *C. coli,* but substantially increased the proportion of *C. jejuni* cases attributed to ruminants and unsampled sources while reducing attribution to chickens.

The asymmetric island models produced source attribution estimates with narrower credible intervals than Dirichlet models, both at the species (Figure 1) and ST level (Figure S2). This was particularly notable for sequence types like ST2398 (which was found in cases but not in sources) for which models M7 and M10 attributed 97% (95% CrI: 83%-100%) and 65% (95%CrI: 55%-74%) of cases to chicken (Figure S2). The asymmetric island model also attributed a larger proportion of cases to chicken meat and a smaller proportion to ruminants and pigs.

Model M1 had the best predictions for the relative abundance of STs in cases as measured with WAIC (Table S5). M1 had substantively better model predictions than all other models except M2 (same as M1 but without an unsampled source) and M7 (asymmetric Island model with variable type transmission potential) (Table S5). When WAIC was calculated for predictions of sequence type abundance in cases and sources, M3 had the best predictions, although the difference between M3 and M1 was not substantive (less than 5 standard errors) (Table S5). As the only difference between M1 and M3 is that the latter has a weaker prior on relative abundance of sequence types in sources, this indicates that the strength of this prior involves a trade-off between prediction in cases compared to sources. As we were primarily interested in cases, M1 was our preferred model. With model M1, the proportions of cases attributed to each source were 80% (95% credible interval [CrI]: 61-92%) to chickens, 2% (CrI: 0.3-12%) to ruminants, 2% (CrI: 0.2-11%) to pigs, and 14% (CrI: 0.3-32%) to the 'unsampled source' (Figure 1). With model M2, the proportions of cases attributed to each source were similar to M1 for chicken (81% [CrI: 59-96%]) but with greater attribution to ruminants (12% [CrI: 1.3-33%]) and pigs (6% [CrI: 1.1-19%]) (Figure 1). Attribution proportions with M7 were similar to M1 for ruminants (2% [CrI: 0.3-9.6%]) and pigs (2% [CrI: 0.1-14%]), but with more attribution to chicken (96% [CrI: 80-99%]) (Figure 1).

When comparing the relative attributable proportion (RAP) by source — calculated by dividing attribution proportions by domestic consumption of that source and normalising against chickens — all models estimated chickens had a higher RAP than pigs (Figure S5). Half the models estimated that RAP of chicken was higher than ruminants, and the remaining models (which included the three worst fitting models, M8-M10) were inconclusive (i.e. 95% credible intervals for RAP include 1, i.e. equal RAP). Our best fitting model (M1) estimated that chickens had a RAP 22 times that of pigs (95% CrI: 3.6-240) and 27 times that of ruminants (95% CrI: 4.0-190). However, models without an 'unsampled source' attributed a greater percentage of cases to ruminants and pigs when compared to similar models with an 'unsampled source', increasing the estimates of RAP for these sources. For instance, model M2 (which had similar fit statistics to M1) estimated that chickens had a RAP 7.3 times that of pigs (95% CrI: 2.1-47) and 4.8 times that of ruminants (95% CrI: 1.3-51) (Figure S5).

Discussion

We estimate that approximately 80% of campylobacteriosis in Australia circa 2018 was attributable to transmission from chickens, with greater attribution to chickens for *C. coli* than *C. jejuni.* Our models attributed to meat sources (e.g. chicken vs. ruminant), not transmission routes (e.g. consumption of contaminated chicken meat compared to contact with chicken faeces). However, as most campylobacteriosis in Australia is believed to be foodborne [8] we can approximate the risk posed by the consumption of meat from a source with the relative

attribution proportion, i.e. by dividing the proportion of cases attributed to that source by its domestic consumption. Applying this method, we estimate consuming chicken meat posed a 22-27 times greater risk of campylobacteriosis than consuming meat from pigs and ruminants, in accord with the case-control study findings [23].

Our modelling of sporadic campylobacteriosis cases aligns with outbreak investigation findings in Australia, the majority of which have linked cases to chicken or dishes containing chicken [11]. Source attribution studies in other high income countries have also identified chicken as the leading source of *Campylobacter* infections, e.g. Switzerland [40], New Zealand [13, 17], Germany [41], Denmark [42], the Netherlands [43] and the United Kingdom [44]. In our study, only a small number of *C. coli* and *C. jejuni* campylobacteriosis cases were attributed to pigs, similar to studies conducted in Denmark [42], Switzerland [40], and the Netherlands [43] and findings for *C. jejuni* in Germany [41]. However, our attributions contrast with findings in Germany where pigs have been identified as a major source of *C. coli* campylobacteriosis. Our attribution proportion estimates for ruminants were sensitive to model assumptions (e.g., inclusion/exclusion of an 'unsampled source'), with up to one quarter of *C. coli* and one-half of *C. jejuni* infections attributable to ruminants in some models, but less than five percent of infections in other models. However, the model with best fit to data (M1) and the two models with similar fit (M2 and M7) all attributed < 12% to ruminants. We did not find inclusion of case covariates such as rurality, age, sex, or jurisdiction substantially improved model predictions. However, point estimates of attribution to chicken meat was higher in urban than in rural populations, in agreement with studies in New Zealand [13].

The number of isolates from ruminants and pigs was low despite extensive sampling, as the prevalence of *Campylobacter* was lower than anticipated. Observed ST diversity in each source indicated sampling was far from reaching saturation for pigs and ruminants. Up to one third of cases were attributed to an "unsampled source" when this was included in models. However, in the model with the best fit, only 14% of cases were attributed to the unsampled source. Models without an unsampled source attributed more cases to all three sampled sources, but primarily ruminants and pigs. Attribution to the unsampled source may genuinely indicate one or more unidentified sources (e.g., companion animals, environmental sources including from wildlife, water, overseas acquisition), but may also be driven by uncertainty introduced by low isolate numbers from ruminants and pigs and the high ST diversity.

Although all models identified chickens as the primary source of campylobacteriosis, there were substantial differences in attribution dependent on model assumptions. The most influential assumption concerned the relative transmission potential of sequence types. In our model, sequence type transmission potential captured any ST organism-intrinsic differences affecting the risk of developing a (notified) infection following exposure to a contaminated food source e.g. ST-specific differences in abundance on contaminated food, survival, or virulence. Our findings indicated transmission potential varied between STs. Models that made the strong assumption that all types had equal transmission potential had attribution estimates with narrower credible intervals, but with substantially poorer fit to the data. Unlike the Dirichlet model, the asymmetric island model incorporates locus-level information about STs relatedness by considering shared alleles. However, including differences in type transmission potential in the asymmetric island model resulted in

biologically implausible estimates of transmission potential for STs that were not detected in sources. In our models, ST48 was estimated to have high transmission potential, in agreement with a study in New Zealand [14]. Given emerging *Campylobacter* virulence factor research (e.g. [45, 46]), it may prove valuable to investigate whether these factors are associated with ST transmission potential in source attribution studies. If estimates of type transmission potential are consistent over time and across studies then transmission potential could be used to differentiate STs of substantial concern (e.g. ST48) or limited concern (e.g. ST827 and ST832).

Models that allowed for differences in transmission potential between STs suggested *C. coli* STs generally had lower transmission potentials than *C. jejuni* STs, possibly explaining why *C. coli* predominated in chickens and yet *C. jejuni* predominated in human cases. Salmonellosis source attribution models have often accounted for variable serotype transmission potential (e.g. [12, 14, 15, 47]). Our findings here are consistent with our previous study which found that assuming *Salmonella* serotypes had equal transmission potential distorted attribution estimates, particularly underestimating attribution to sources harbouring a mix of serotypes with low and high transmission potential; namely, broiler chickens [30].

Many source attribution studies rely on secondary analysis of multiple datasets, often adopting different sampling approaches for different food animals, including targeted sampling (e.g. outbreak investigations) and sampling at different stages of the food production process (e.g. farm, processor, or retail). In contrast, our study applied a common sampling approach, sampling specific animal sources retail meats and meat types over the course of two years. This included samples from the four major food animals consumed in Australia (chickens, pigs, cattle, and sheep). We included cases from every Australian state and two most populous territories (the ACT and NT). We collected food isolates from four jurisdictions including Australia's three most populous states (NSW, Qld, and Vic). We did not find any differences in the human cases between jurisdiction in this or previous studies [24] and therefore believe our estimates of source attribution are generalisable to the general Australian population. By adopting a model comparison approach, we were able to identify findings consistent across models and assumptions, including high attribution to chicken.

Our study has some limitations. Given the generally low prevalence of *Campylobacter* on muscle meat from pigs and ruminants, we chose to sample offal meats, which have previously been found to have much higher prevalence of *Campylobacter* contamination [26]. If the STs present on muscle meat and offal are substantially different, our samples may not have been representative of the STs encountered when consuming meat from ruminants and pigs. However, the STs in muscle and offal meats from chicken were similar [22] supporting the validity of our sampling approach. Despite extensive sampling efforts, the number of isolates from pigs and ruminants limited the statistical power of our study [21, 22], particularly with respect to detection sensitivity of differences in attribution by state, rurality, or other covariates. There is a general lack of theoretical work to inform questions of power and minimum sample sizes at the design phase of source attribution studies [48]. While *Campylobacter* has been isolated from a wide range of wild and domestic animal species, this study only included samples from the primary retail meat production animals in Australia. However, 62-89% of cases in Australia are believed to be foodborne [8] primarily from meat, in concordance with our best fitting models that attributed 86% to chickens, pigs, and

ruminants, and 14% to an unsampled source. Attribution to unsampled source must be interpreted with caution, as it could represent genuine unsampled reservoir(s) (e.g. wildlife and/or companion animals, water) but might also be an artefact of limited sample sizes. In particular, the large attribution proportions for the unsampled sources predicted by the model with weak priors on the relative abundance of sequence types in sources (M3) should be interpreted as an uncertain attribution rather than strong evidence of one or more major unsampled sources. Some of the attribution to an unsampled source may be due to travel associated cases, though this was unlikely to be a major factor in our study, as about 97% of Australian campylobacteriosis cases are believed to be domestically acquired [3] and 68% of the cases in our study were from a case-control study which excluded travel associated cases. The sampling periods for isolates from human cases and food were only partially overlapping (February 2018 to October 2019 vs March 2017 to March 2019). However, this was unlikely to have biased our analyses as the relative frequency of STs in sources and cases were unlikely to have changed substantially over the sampling period and the multi-year sampling periods allowed us to average over any seasonal differences. Application of the asymmetric island model to attribute relative abundance of STs in sources performed worse than anticipated, leading to a poor model fit when all STs were assumed to have equal transmission potential (the standard assumption when using the asymmetric island model), but resulting in biologically implausible estimates of transmission potential when this assumption was relaxed. This limited our ability to utilise the full richness of the genomic data (e.g. core genome MLST or single nucleotide polymorphisms) as the Dirichlet model will perform poorly when most or all isolates are assigned a unique type. Further work is also required to better account for differences in type transmission potential when using the asymmetric island model.

Our best fitting model estimated that chickens account for about 80% of campylobacteriosis and that even after adjusting for the relatively high rates of chicken meat consumption chicken poses a risk of campylobacteriosis approximately five to thirty times higher than ruminants and seven to twenty-two times higher than pigs. Chickens should therefore remain the priority target for reducing burden of *Campylobacteriosis* in Australia, including increased promotion of safe food handling practices for all raw meats and reduction of meat contamination during production and processing.

Abbreviations

MLST: multi-locus sequence typing;
ST: sequence type;
NSW: New South Wales (an Australian state);
HNE: Hunter New England (a public health district of NSW);
Qld: Queensland (an Australian state);
ACT: Australian Capital Territory (an Australian territory);
Vic: Victoria (an Australian state);
WA: Western Australia (an Australian state);
Tas: Tasmania (an Australian state);
SA: South Australia (an Australian state);
NT: Northern Territory (an Australian territory);
ISO: International Organisation for Standardisation;

AS: Standards Australia;
DNA: deoxyribonucleic acid;
USA: United States of America;
CA: California (a state of the USA);
BLAST: Basic Local Alignment Search Tool;
RAP: relative attributable proportion;
MCMC: Markov chain Monte Carlo;
WAIC: Watanabe-Akaike information criterion;
ANU: Australian National University.

Declarations

Ethics approval and consent to participate

All methods were carried out in accordance with relevant guidelines and regulations. Informed consent was provided by all participants (or from a parent or legal guardian for cases under 18 years old) prior to participation in the study. The study protocol was approved by the Australian National University (ANU) Human Research Ethics Committee (Protocol: 2016/426), ACT Health Human Research Ethics Committee (Protocol: ETH.8.17.168), Hunter New England Human Research Ethics Committee (Protocol: 17/08/16/4.03), Qld Health Human Research Ethics Committee (Protocol: RD007108), and the University of Melbourne Office of Research Ethics and Integrity (Protocol: 1750366.1).

Consent for publication

Not applicable

Availability of data and material

The sequence readset for each food isolate (Bioproject Accession: PRJNA591966), case isolates sampled as part of the case control study (Bioproject Accession: PRJNA592186) and the case isolates included in the national snapshot (Bioproject Accession: PRJNA560409) are available through GenBank.

Competing interests

The authors declare that they have no competing interests.

Funding

The CampySource Project was funded by a National Health and Medical Research Council grant (NHMRC GNT1116294), AgriFutures, Australian Government Department of Health, Food Standards Australia New Zealand, New South Wales Food Authority, Queensland Health, and ACT Health. MDK was supported by a National Health and Medical Research Council fellowship (GNT1145997). AM was supported through an Australian Research Council Discovery Project Grant (DP180100246). DC was supported by an Australian Government Research Training Program (AGRTP) scholarship.

Authors' contributions

Conceptualisation: MDK, KG, NF; Methodology: KG, AM; Software: AM; Formal analysis: AM, DB; Investigation: CM, MV, AJ, JS; Resources: AJ, MV, DB; Data curation: RW, DB, DC; Visualisation: AM; Supervision: KG; Project administration: DC; Funding acquisition: KG, MDK, NF, MV, DB, EF. AM prepared the original draft and all authors reviewed and edited the manuscript. The final version of the manuscript was approved by all authors.

Acknowledgements

The authors would like to thank Jonathan Marshall for valuable conversations about the asymmetric island model and for making changes to the islandR package to accommodate equal mutation and recombination rates across sources.

The authors would like to thank the extended CampySource Project team, reference panel, and additional contributors to the study. The CampySource Project Team comprised three working groups and a reference panel. The working groups focused on food and animal sampling, epidemiology and modelling, and genomics. The reference panel included expert representatives from government and industry.

The study includes the following partner organisations: the Australian National University, Massey University, University of Melbourne, Queensland Health, Queensland Health Forensic and Scientific Services, New South Wales Food Authority, New South Wales Health, Hunter New England Health, Victorian Department of Health and Human Services, Food Standards Australia New Zealand, Commonwealth Department of Health, and AgriFutures Australia–Chicken Meat Program. CampySource also collaborated with the following organisations: ACT Health, Sullivan Nicolaides Pathology, University of Queensland, Primary Industries and Regions South Australia, Department of Health and Human Services Tasmania, Meat and Livestock Australia, and New Zealand Ministry for Primary Industries. We would particularly like to acknowledge the public health and environmental health officers from our partner and collaborating organisations who carried out the project sampling.

The CampySource Project Team included: Nigel P French, Massey University, New Zealand; Mary Valcanis, The University of Melbourne; Dieter Bulach, The University of Melbourne; Emily Fearnley, South Australian Department for Health and Wellbeing; Russell Stafford, Queensland Health; Amy Jennison, Queensland Health; Trudy Graham, Queensland Health; Keira Glasgow, Health Protection NSW; Kirsty Hope, Health Protection NSW; Themy Saputra, NSW Food Authority; Craig Shadbolt, NSW Food Authority; Arie Havelaar, The University of Florida, USA; Joy Gregory, Department of Health and Human Services, Victoria; James Flint, Hunter New England Health; Simon Firestone, The University of Melbourne; James Conlan, Food Standards Australia New Zealand; Ben Daughtry, Food Standards Australia New Zealand; James J Smith, Queensland Health; Heather Haines, Department of Health and Human Services, Victoria; Sally Symes, Department of Health and Human Services, Victoria; Barbara Butow, Food Standards Australia New Zealand; Liana Varrone, The University of Queensland; Linda Selvey, The University of Queensland; Tim Sloan-Gardner, ACT Health; Deborah Denehy, ACT Health; Radomir Krsteski, ACT Health; Natasha Waters, ACT Health; Kim Lilly, Hunter New England Health; Julie Collins, Hunter New England Health; Tony Merritt, Hunter New England Health; Rod Givney, Hunter New England Health; Joanne Barfield, Hunter New England Health; Ben Howden, The University of Melbourne; Kylie Hewson, AgriFutures Australia–Chicken Meat Program; Dani Cribb, The Australian National University; Rhiannon Wallace, The Australian National University; Angus McLure, The Australian National University; Ben Polkinghorne, The Australian National University; Cameron Moffatt, The Australian National University and Queensland Health; Martyn Kirk, The Australian National University; and Kathryn Glass, The Australian National University.

References

1. Li M, Havelaar AH, Hoffmann S, Hald T, Kirk MD, Torgerson PR, et al. Global disease burden of pathogens in animal source foods, 2010. PLOS ONE. 2019;14(6):e0216545. doi: 10.1371/journal.pone.0216545.
2. Kirk MD, Pires SM, Black RE, Caipo M, Crump JA, Devleesschauwer B, et al. World Health Organization Estimates of the Global and Regional Disease Burden of 22 Foodborne Bacterial, Protozoal, and Viral Diseases, 2010: A Data Synthesis. PLOS Medicine. 2015;12(12):e1001921. doi: 10.1371/journal.pmed.1001921.
3. Kirk M, Ford L, Glass K, Hall G. Foodborne Illness, Australia, Circa 2000 and Circa 2010. Emerging Infectious Diseases. 2014;20(11):1857-64. doi: 10.3201/eid2011.131315.
4. List of Prokaryotic names with Standing in Nomenclature: Genus Campylobacter. [cited 2021/09/01]. Available from: https://lpsn.dsmz.de/genus/campylobacter.
5. Kaakoush NO, Castaño-Rodríguez N, Mitchell HM, Man SM. Global Epidemiology of Campylobacter Infection. Clinical Microbiology Reviews. 2015;28(3):687-720. doi: 10.1128/cmr.00006-15.
6. Dingle KE, Colles FM, Wareing DRA, Ure R, Fox AJ, Bolton FE, et al. Multilocus Sequence Typing System for *Campylobacter jejuni*. Journal of Clinical Microbiology. 2001;39(1):14-23. doi: 10.1128/jcm.39.1.14-23.2001.
7. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. Wellcome Open Research. 2018;3:124. doi: 10.12688/wellcomeopenres.14826.1.
8. Vally H, Glass K, Ford L, Hall G, Kirk MD, Shadbolt C, et al. Proportion of illness acquired by foodborne transmission for nine enteric pathogens in Australia: an expert elicitation. Foodborne pathogens and disease. 2014;11(9):727-33.
9. Varrone L, Glass K, Stafford RJ, Kirk MD, Selvey L. A meta-analysis of case-control studies examining sporadic campylobacteriosis in Australia and New Zealand from 1990 to 2016. Australian and New Zealand Journal of Public Health. 2020;44(4):313-9. doi: 10.1111/1753-6405.12998.
10. Kuhn K, Hvass A, Christiansen A, Ethelberg S, Cowan S. Sexual Contact as Risk Factor for Campylobacter Infection, Denmark. Emerging Infectious Diseases. 2021;27(4):1133-40.
11. Moffatt CRM, Fearnley E, Bell R, Wright R, Gregory J, Sloan-Gardner T, et al. Characteristics of Campylobacter Gastroenteritis Outbreaks in Australia, 2001 to 2016. Foodborne Pathogens and Disease. 2020;17(5):308-15. doi: 10.1089/fpd.2019.2731.
12. Mullner P, Jones G, Noble A, Spencer SEF, Hathaway S, French NP. Source Attribution of Food-Borne Zoonoses in New Zealand: A Modified Hald Model. Risk Analysis. 2009;29(7):970-84. doi: 10.1111/j.1539-6924.2009.01224.x.
13. Liao S-J, Marshall J, Hazelton ML, French NP. Extending statistical models for source attribution of zoonotic diseases: a study of campylobacteriosis. Journal of The Royal Society Interface. 2019;16(150):20180534. doi: 10.1098/rsif.2018.0534.
14. Miller P, Marshall J, French N, Jewell C. sourceR: Classification and source attribution of infectious agents among heterogeneous populations. PLOS Computational Biology. 2017;13(5):e1005564. doi: 10.1371/journal.pcbi.1005564.
15. Hald T, Vose D, Wegener HC, Koupeev T. A Bayesian Approach to Quantify the Contribution of Animal-Food Sources to Human Salmonellosis. Risk Analysis. 2004;24(1):255-69. doi: 10.1111/j.0272-4332.2004.00427.x.
16. Wilson DJ, Gabriel E, Leatherbarrow AJH, Cheesbrough J, Gee S, Bolton E, et al. Tracing the Source of Campylobacteriosis. PLoS Genetics. 2008;4(9):e1000203. doi: 10.1371/journal.pgen.1000203.

17. Mullner P, Spencer SEF, Wilson DJ, Jones G, Noble AD, Midwinter AC, et al. Assigning the source of human campylobacteriosis in New Zealand: A comparative genetic and epidemiological approach. Infection, Genetics and Evolution. 2009;9(6):1311-9. doi: 10.1016/j.meegid.2009.09.003.
18. Sears A, Baker MG, Wilson N, Marshall J, Muellner P, Campbell DM, et al. Marked Campylobacteriosis Decline after Interventions Aimed at Poultry, New Zealand. Emerging Infectious Diseases. 2011;17(6):1007-15. doi: 10.3201/eid1706.101272.
19. Lake RJ, Campbell DM, Hathaway SC, Ashmore E, Cressey PJ, Horn BJ, et al. Source attributed case-control study of campylobacteriosis in New Zealand. International Journal of Infectious Diseases. 2021;103:268-77. doi: 10.1016/j.ijid.2020.11.167.
20. Varrone L, Stafford RJ, Lilly K, Selvey L, Glass K, Ford L, et al. Investigating locally relevant risk factors for Campylobacter infection in Australia: protocol for a case–control study and genomic analysis. BMJ Open. 2018;8(12):e026630. doi: 10.1136/bmjopen-2018-026630.
21. Walker LJ, Wallace RL, Smith JJ, Graham T, Saputra T, Symes S, et al. Prevalence of Campylobacter coli and Campylobacter jejuni in Retail Chicken, Beef, Lamb, and Pork Products in Three Australian States. Journal of Food Protection. 2019;82(12):2126-34. doi: 10.4315/0362-028x.jfp-19-146.
22. Wallace RL, Bulach DM, Jennison AV, Valcanis M, Mclure A, Smith JJ, et al. Molecular characterization of Campylobacter spp. recovered from beef, chicken, lamb and pork products at retail in Australia. PLOS ONE. 2020;15(7):e0236889. doi: 10.1371/journal.pone.0236889.
23. Cribb DM, Varrone L, Wallace RL, Mclure AT, Smith JJ, Stafford RJ, et al., Risk Factors for Campylobacteriosis in Australia: Outcomes of a 2018–2019 Case-Control Study, 2022-03-10, PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-1387051/v1].
24. Wallace RL, Bulach D, McLure A, Varrone L, Jennison AV, Valcanis M, et al. Antimicrobial Resistance of Campylobacter spp. Causing Human Infection in Australia: An International Comparison. Microbial Drug Resistance. 2021. doi: 10.1089/mdr.2020.0082.
25. Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Research. 2016;44(D1):D67-D72. doi: 10.1093/nar/gkv1276.
26. New South Wales Food Authority. Campylobacter in meat and offal: Microbiological quality of beef, lamb and pork meat cuts and offal. Sydney, Australia: New South Wales Food Authority, 2018.
27. International Organization for Standardization. ISO 10272–1:2017 Microbiology of the food chain—Horizontal method for detection and enumeration of Campylobacter spp. Part 1: Detection method. Geneva: Inter- national Organization for Standardization; 2017.
28. Standards Australia. Food microbiology, Method 6: Examination for specific organisms - Campylobacter. Sydney: Standards Australia; 2015.
29. Seemann T, mlst. GitHub https://github.com/tseemann/mlst
30. Mclure A, Shadbolt C, Desmarchelier PM, Kirk MD, Glass K. Source Attribution of Salmonellosis by Time and Geography in New South Wales, Australia. 2021.
31. Agricultural commodity statistics [Internet]. Australian Bureau of Agricultural and Resource Economics and Sciences. [cited 2020/03/07]. Available from: https://www.agriculture.gov.au/abares/research-topics/agricultural-outlook/data.
32. R Core Team, R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing https://www.R-project.org/
33. Wickham H, Averick M, Bryan J, Chang W, McGowan LDA, François R, et al. Welcome to the Tidyverse. Journal of open source software. 2019;4(43):1686.

34. Gao C-H, ggVennDiagram: A 'ggplot2' Implement of Venn Diagram. https://github.com/gaospecial/ggVennDiagram
35. Stan Development Team. Stan Modeling Language Users Guide and Reference Manual, 2.25. 2020.
36. Stan Development Team. RStan: the R interface to Stan. 2020.
37. Vehtari A, Gelman A, Simpson D, Carpenter B, Bürkner P-C. Rank-Normalization, Folding, and Localization: An Improved 'R hat' for Assessing Convergence of MCMC. Bayesian Analysis. 2021;16(2):667-718. doi: 10.1214/20-BA1221.
38. Watanabe S, Opper M. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. Journal of machine learning research. 2010;11(12).
39. Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and Computing. 2017;27:1413-32.
40. Kittl S, Heckel G, Korczak BM, Kuhnert P. Source Attribution of Human Campylobacter Isolates by MLST and Fla-Typing and Association of Genotypes with Quinolone Resistance. PLoS ONE. 2013;8(11):e81796. doi: 10.1371/journal.pone.0081796.
41. Rosner BM, Schielke A, Didelot X, Kops F, Breidenbach J, Willrich N, et al. A combined case-control and molecular source attribution study of human Campylobacter infections in Germany, 2011–2014. Scientific Reports. 2017;7(1). doi: 10.1038/s41598-017-05227-x.
42. Boysen L, Rosenquist H, Larsson JT, Nielsen EM, Sørensen G, Nordentoft S, et al. Source attribution of human campylobacteriosis in Denmark. Epidemiology and Infection. 2014;142(8):1599-608. doi: 10.1017/s0950268813002719.
43. Mughini-Gras L, Kooh P, Augustin J-C, David J, Fravalo P, Guillier L, et al. Source Attribution of Foodborne Diseases: Potentialities, Hurdles, and Future Expectations. Frontiers in Microbiology. 2018;9. doi: 10.3389/fmicb.2018.01983.
44. Thépault A, Méric G, Rivoal K, Pascoe B, Mageiros L, Touzain F, et al. Genome-Wide Identification of Host-Segregating Epidemiological Markers for Source Attribution in Campylobacter jejuni. Applied and Environmental Microbiology. 2017;83(7):AEM.03085-16. doi: 10.1128/aem.03085-16.
45. Zhang T, Luo Q, Chen Y, Li T, Wen G, Zhang R, et al. Molecular epidemiology, virulence determinants and antimicrobial resistance of Campylobacter spreading in retail chicken meat in Central China. Gut Pathogens. 2016;8(1). doi: 10.1186/s13099-016-0132-2.
46. Bolton DJ. Campylobacter virulence and survival factors. Food microbiology. 2015;48:99-108.
47. Mughini-Gras L, Barrucci F, Smid JH, Graziani C, Luzzi I, Ricci A, et al. Attribution of human *Salmonella* infections to animal and food sources in Italy (2002–2010): adaptations of the Dutch and modified Hald source attribution models. Epidemiology and Infection. 2014;142(5):1070-82. doi: 10.1017/s0950268813001829.
48. Smid JH, Mughini Gras L, De Boer AG, French NP, Havelaar AH, Wagenaar JA, et al. Practicalities of Using Non-Local or Non-Recent Multilocus Sequence Typing Data for Source Attribution in Space and Time of Human Campylobacteriosis. PLoS ONE. 2013;8(2):e55029. doi: 10.1371/journal.pone.0055029.

Table 1 Source attribution models considered, with model numbers for those models that converged.

Model	Type transmission potential terms (q)[1]:	Model of relative abundance of STs in each source | prior		Joint vs two-step inference?[2]	With unsampled source?
M1	Varied	Dirichlet	Dirichlet(1,…,1)	Joint	Yes
M2					No
M3			Dirichlet(0.1,…,0.1)		Yes
M4					No
M5			Dirichlet(1,…,1)	Two-step	Yes
M6					No
M7		Asymmetric island			
*	Equal	Dirichlet	Dirichlet(1,…,1)	Joint	Yes
*					No
*			Dirichlet(0.1,…,0.1)		Yes
*					No
M8			Dirichlet(1,…,1)	Two-step	Yes
M9					No
M10		Asymmetric island			

* The fitting procedure for these models failed to converge, so results are omitted.

[1] Type transmission potential terms were either all set to 1 or estimated using a log-normal hyperprior with unknown variance.

[2] In joint models, all parameters were estimated simultaneously in fully joint Bayesian inference; otherwise, the posterior distribution of the relative abundance of sequence types (STs) in each source was estimated separately, and 100 draws from this posterior were used to estimate the remaining parameters in a second inference step.

Table 2 Number of distinct sequence types (STs) and isolates from humans and food sources. Percentages denote of the fraction of all STs/isolates from that origin. See Table S1 for *C. coli* numbers and Table S2 for *C. jejuni* numbers.

Origin	STs	Isolates	Unique* STs (%)	Isolates from Unique* STs (%)	STs without cases (%)	Isolates from STs without cases (%)
Human	118	710	66 (55.9%)	134 (18.9%)	N/A	N/A
Chicken	88	480	60 (68.2%)	192 (40%)	38 (43.2%)	90 (18.8%)
Pig	33	44	17 (51.5%)	20 (45.5%)	21 (63.6%)	25 (56.8%)
Ruminant	25	88	3 (12%)	3 (3.4%)	2 (8%)	2 (2.3%)

*Unique types for sources are types that were found in that source and no other source (except potentially in humans), while unique types for humans are those found in cases but not in any of the three sources.

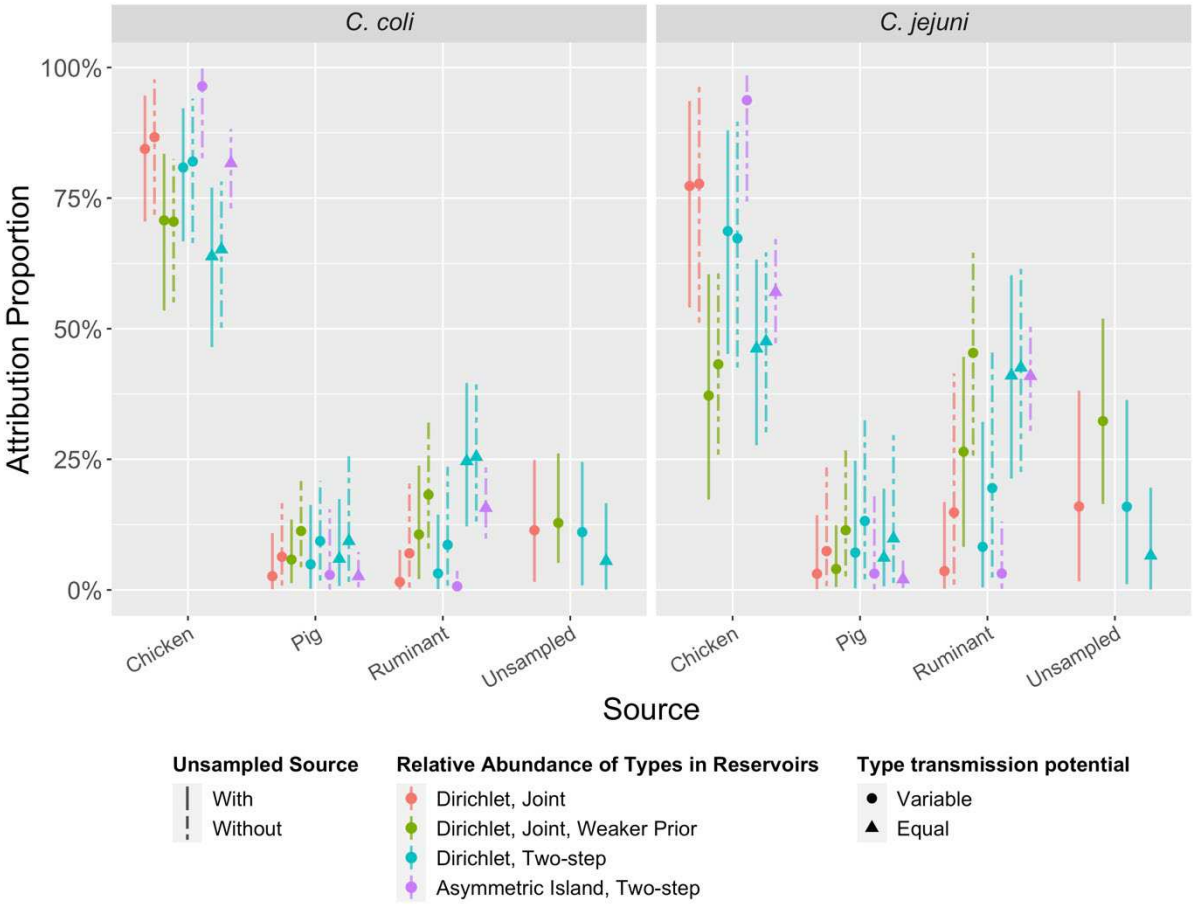

Figure 1 Source attribution proportions of *C. jejuni* and *C. coli* cases to three sampled sources (chicken, pig, ruminant) in ten models (M1-M10, left to right). Four models (M1, M3, M5, M8) also include a fourth, 'unsampled source'. The asymmetric island model is intrinsically unable to accommodate an unsampled source. See Table 1 for the list of assumptions for each model.

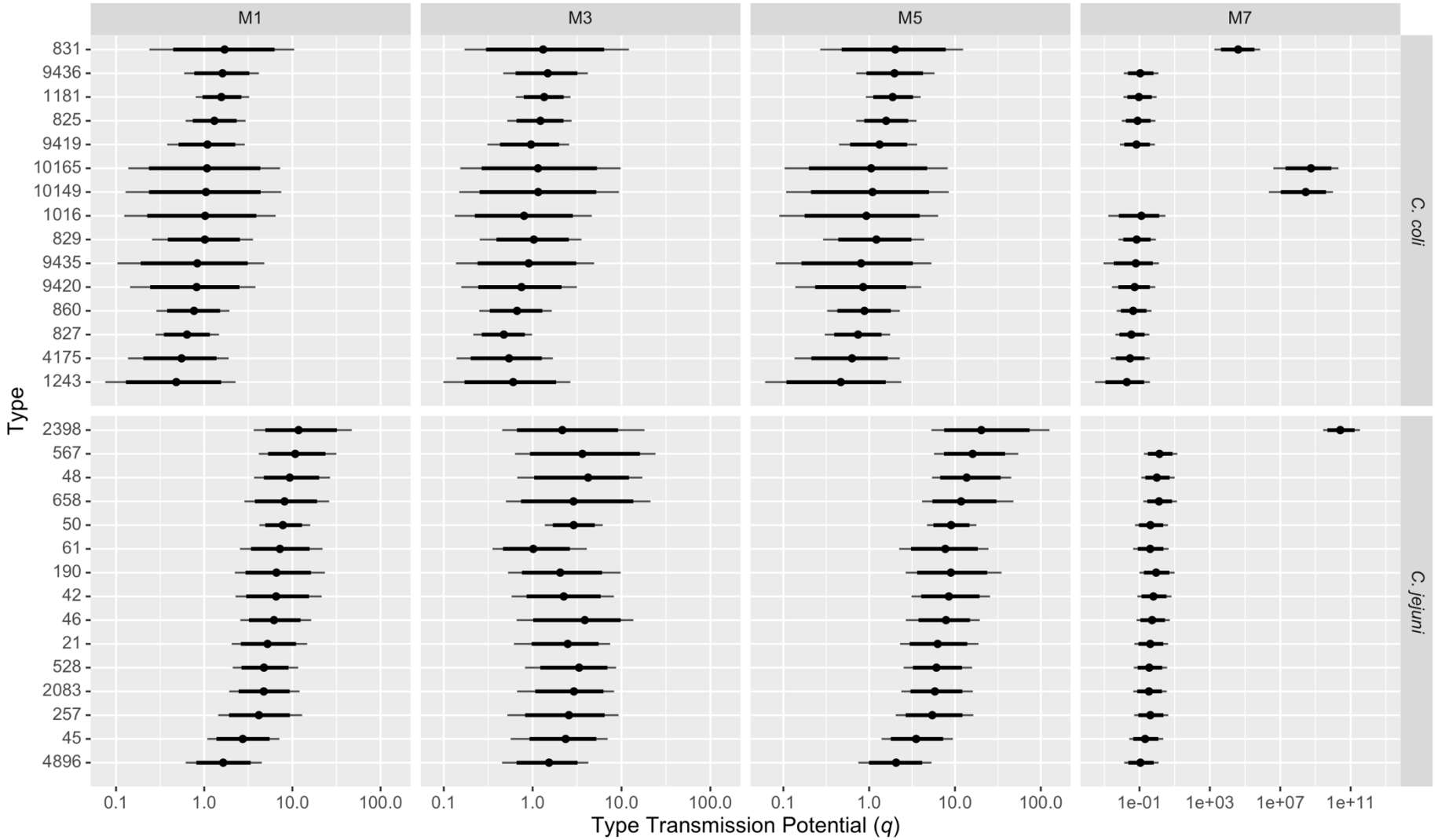

Figure 2 Posterior median and 95% credible intervals for type transmission potential (relative ability of a sequence type to transmit from a source and lead to a reported campylobacteriosis case) of the fifteen most common *C. coli* and *C. jejuni* multi-locus sequence types (STs) in human cases for four models. STs have been ordered by posterior median type transmission potential in model M2. Note the x-axis is on a log scale, with wider limits for M7 (asymmetric island model). See Table 1 for details of the four models.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- McLureSupplementaryFileCampySourceAttributionAustralia.docx