SCIENTIFIC REPORTS

OPEN

Source attribution of Campylobacter jejuni shows variable importance of chicken and ruminants reservoirs in non-invasive and invasive French clinical isolates

Received: 13 December 2018
Accepted: 17 May 2019
Published online: 30 May 2019

Elvire Berthenet[1], Amandine Thépault[2], Marianne Chemaly[2], Katell Rivoal[2], Astrid Ducournau[1], Alice Buissonnière[1], Lucie Bénéjat[1], Emilie Bessède[1,3], Francis Mégraud[1,3], Samuel K. Sheppard[4] & Philippe Lehours[1,3]

Campylobacter jejuni **is the most common cause of bacterial gastroenteritis worldwide. Mainly isolated from stool samples, *C. jejuni* can also become invasive. *C. jejuni* belongs to the commensal microbiota of a number of hosts, and infection by this bacterium can sometimes be traced back to exposure to a specific source. Here we genome sequenced 200 clinical isolates (2010–2016) and analyzed them with 701 isolate genomes from human infection, chicken, ruminants and the environment to examine the relative contribution of different reservoirs to non-invasive and invasive infection in France. Host-segregating genetic markers that can discriminate *C. jejuni* source were used with STRUCTURE software to probabilistically attribute the source of clinical strains. A self-attribution correction step, based upon the accuracy of source apportionment within each potential reservoir, improved attribution accuracy of clinical strains and suggested an important role for ruminant reservoirs in non-invasive infection and a potentially increased contribution of chicken as a source of invasive isolates. Structured sampling of *Campylobacter* in the clinic and from potential reservoirs provided evidence for variation in the contribution of different infection sources over time and an important role for non-poultry reservoirs in France. This provides a basis for ongoing genomic epidemiology surveillance and targeted interventions.**

Campylobacter jejuni is one of the most common bacterial enteropathogens in both high and low income countries[1,2]. In clinical microbiology laboratories, *C. jejuni* is mainly isolated from stools, but 1–2% of cultured strains are isolated from blood[3,4]. Infection symptoms vary from watery diarrhea to bloody stools, accompanied by fever, abdominal pain, vomiting and dehydration[5,6]. Post-infectious complications can occur, including Guillain-Barré syndrome[7]. Because of it's clinical importance, determining the source of *C. jejuni* infection is a high priority. However, this is challenging as *C. jejuni* is part of the commensal microbiota of many mammal and bird species and is commonly isolated from poultry[8,9], ruminants[9,10], pigs[11,12], wild birds[13,14] and companion animals (dogs and cats)[15], as well as the environment[16,17].

In the last two decades, characterization of strain variation within populations, using DNA sequence based methods such as Multi-Locus Sequencing Type (MLST)[18] and whole genome sequencing (WGS)[19], has improved

[1]French National Reference Center for Campylobacters & Helicobacters, Bordeaux, France. [2]Unit of Hygiene and Quality of Poultry & Pork Products, Laboratory of Ploufragan-Plouzané-Niort, French Agency for Food Environmental and Occupational Health & Safety (ANSES), Ploufragan, France. [3]Univ. Bordeaux, INSERM, UMR1053 Bordeaux Research in Translational Oncology, BaRITOn, 33076, Bordeaux, France. [4]The Milner Centre for Evolution, Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath, United Kingdom. Correspondence and requests for materials should be addressed to P.L. (email: philippe.lehours@u-bordeaux.fr)

understanding of *Campylobacter* ecology, epidemiology and evolution. In particular, the degree to which lineages are associated with different hosts, reflecting segregating genetic variation that has resulted from the physical isolation of populations in discrete niches as well as adaptations that promote survival in a given host[20]. For example, sequence types (STs) belonging to the ST-257 and ST-353 clonal complexes are most commonly isolated from chickens while ST-61 or ST-42 complex isolates are associated with ruminants[14,21]. As such these lineages can be described as host specialists[21,22]. An applied advantage of understanding the genomics of lineage-host association is that the origin of isolates from human infection can potentially be determined by comparison to genome sequenced isolates from putative reservoir sources, and quantitative probabilistic models have been developed for source attribution of clinical strains[23].

Attribution studies using MLST data have successfully identified the relative contribution of different host reservoirs to human infection and contamination from chicken reservoirs was implicated in several countries[24,25]. However, a limitation to these approaches is that some of the most common strains infecting humans are found in multiple hosts. These ecological generalist strains cannot be easily assigned to one source as recent host transitions erode the signal of host association[26,27]. While this remains a challenging, decreasing costs and increasing availability of large WGS datasets[28] is improving understanding of the genes and genetic elements that promote *C. jejuni* host adaptation[29,30] and survival[31,32] in particular niches. These elements represent candidate markers for source attribution studies and recent work analyzing the pan-genome of 4 *C. jejuni* reference strains in 884 genomes identified 15 host-segregating markers that were used for source attribution of specialist and generalist genotypes[33]. This last method allowed a better accuracy of attributions compared with MLST loci and a higher host segregation of isolates even in host generalist clonal complexes[34].

Implementing these approaches has highlighted the importance of chickens and ruminants as sources of *C. jejuni* strains that infect people in France[34]. The French National Reference Centre receives strains sent by a network of around 200 clinical laboratories spread all over the French territory. Among these strains, most are isolated from stools, but the number of invasive *C. jejuni* strains, isolated from blood, has been steadily increasing since 2014, and in 2017 bacteraemia cases caused by *C. jejuni* exceeded those caused by *Campylobacter fetus* for the first time. The reason for this increase is not known but some clonal complexes have been associated with invasive disease. For example, ST-677 clonal complex isolates that are a common cause of diarrhoeal disease in Finland[35] were also a common cause of invasive infection[4]. However, this clonal complex is less common in some other surveyed countries, including France[34], and it is unclear if particular lineages are over represented in invasive disease.

Source attribution studies are helping to describe the previously cryptic transmission networks of sporadic disease caused by *Campylobacter*. However, effective implementation for epidemiological monitoring is hampered by limitations in attribution study design. First, these studies typically represent epidemiological snapshots representing a discrete period of time, meaning that long term variation and fine scale trends in source-sink dynamics are overlooked. Second, incomplete segregation of genomic markers by source (host) can lead to a weak signal of self-attribution and bias in the overall attribution results. Third, attribution studies typically treat all *Campylobacter* strains equally and, therefore, do not identify the potential sources of strains causing severe infection (i.e. invasive disease). In this study we aimed to improve understanding of the source dynamics of *C. jejuni* infection in France over the last 10 years. Sampling and sequencing (WGS) contemporary isolates (non-invasive and invasive) and analyzing them with available clinical and potential source isolates, we use host-segregating markers[33] and an enhanced source attribution model incorporating a self-attribution correction step. Our analysis describes fluctuations in contribution of host reservoirs over time. Chicken remained a major contributor to non-invasive and invasive disease but there is evidence of ruminant reservoirs as a source of strains associated with invasive disease.

Results

Genetic structure and organization of the dataset. Consistent with previous studies[36], a core-genome tree, based upon 1422 genes shared by at least 90% of sources strains and clinical strains (Fig. 1) showed no evidence of segregation of the French isolates from potential sources compared to those from other countries. Furthermore, clinical strains clustered with strains from potential sources. These two findings confirmed the validity of using a world-wide dataset of source strains to study French clinical isolates. MLST profiles of the complete dataset were concordant with previously published data[22,37–39], with both host generalist and host specialist clonal complexes. Among the host generalist clonal complexes, CC-21 (176 strains) and CC-45 (92 strains) were the most abundant. Known host specialist clonal complexes were also identified including the chicken-associated CC-353 (35 strains) and CC-354 (31 strains) and ruminant-associated CC-42 (20 strains).

Self-attribution of isolates from chicken and ruminants reservoirs. The accuracy of the attribution based upon probabilistic assignment of 15 host segregating markers with STRUCTURE was tested using isolates of known origin (self-attribution). A subset of 20 strains from each of the two reservoirs was randomly selected and 10 replicates of attribution tests were performed by comparison to the remaining isolates from the same host. The average probability of each provenance for the subsets of strains was then analyzed (Fig. 2A). The probabilities of correct self-attribution were estimated at 93% for the chicken reservoir, 69% for the ruminant reservoir and 54% for the environment. These results are acceptable for the chicken reservoir, but highlight a bias for the ruminant and environment in favor of the chicken reservoir, with a risk of under-estimation of test isolates attributed to the latter two sources and over-estimation of the proportion of chicken-attributed isolates. Correction of the bias, using the new method based on a system of 3 equations, gave probabilities of correct self-attribution estimated at 97% for the chicken reservoir, 90% for the ruminant reservoir and 91% for the environment reservoir (Fig. 2B). With correct self-attribution of more than 90% for each of the reservoirs, the correction method provides a useful method for attribution of French clinical isolates.

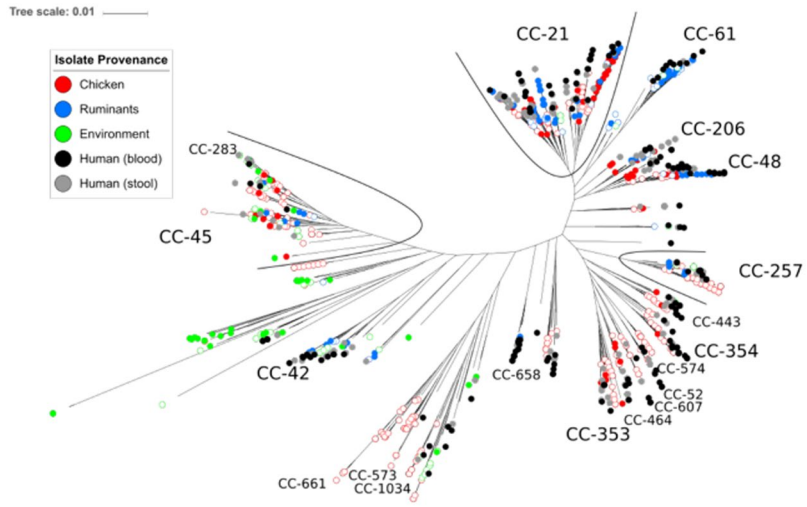

Figure 1. Maximum-likelihood tree based on 1422 concatenated core genes from the 899 strains of *C. jejuni*. Filled circles represent French strains, white circles represent strains from the rest of the world. Clonal complexes obtained from MLST that contain more than 10 strains are labelled in small font. Clonal complexes containing more than 20 strains are labelled in larger font.

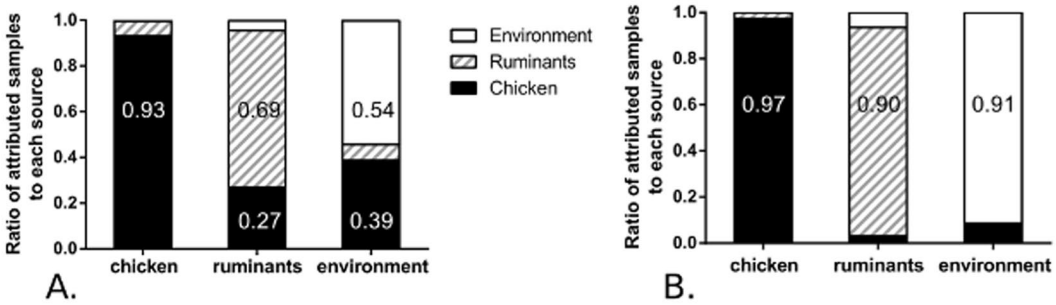

Figure 2. Self-attribution of isolates from chicken, cattle and environment sources. Attribution tests were performed using the STRUCTURE software with 10 replicates. For each sub-dataset of 20 isolates coming from a known source, the proportion of isolates attributed to each reservoir is represented. (**A**) Original self-attribution, uncorrected STRUCTURE results. (**B**) Corrected self-attribution, after correction step based on a system of equations that balances the bias observed in the self-attribution internal results of each attribution test.

Attribution of french clinical isolates. Attribution of the most recent French clinical isolates (from 2014 to 2016) was performed using the 15 host-segregating markers and STRUCTURE software. Corrected attribution results (Fig. 3) show that the proportion of non-invasive isolates attributed to the chicken reservoir (39%) was lower than in previous studies[34] were 63% of clinical isolates from 2015 were attributed to chicken using the same 15 host-segregating markers. Isolates attributed to ruminants (58%) were predominant in the non-invasive isolates from 2014 to 2016. The proportion of chicken attributed isolates in invasive strains was much higher (60%) compared to non-invasive isolates. The remaining invasive isolates were attributed equally to the ruminants and the environment (21% and 19%, respectively).

Attribution of non-invasive strains was performed independently for strains isolated in 2009, 2015 and 2016 (39, 78 and 26 strains, respectively), since a variation in source attribution according to the year of isolation was observed[34]. Corrected attribution results (Fig. 4A) show that the proportion of environment attributed strains remained low (below 10%) over time. The majority of isolates were attributed to ruminants in 2009 and 2015 (54.6% and 64.5%, respectively), but switched to a majority of chicken attributed isolates in 2016 (59%). Attribution of invasive strains was performed independently year-by-year for strains isolated in 2011, 2012, 2013, 2014, 2015 and 2016 (17, 18, 33, 35, 33 and 37 strains respectively). Corrected attribution results (Fig. 4B) show that the proportion of environment attributed strains, increased after 2014 from below 10% (8.8%, 0% and 6.5% for 2011, 2012 and 2013 respectively) to over 10% (20.1%, 13.5% and 21.4% for 2014, 2015 and 2016 respectively). The majority of isolates (57.6% to 86.5%) were attributed to chicken every year except 2014 were only 37.7% of isolates were attributed to chicken.

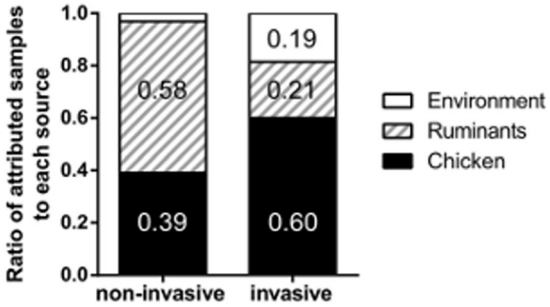

Figure 3. Attribution of French clinical non-invasive (n = 104) and invasive (n = 105) isolates, collected between 2014 and 2016, to isolates from chicken, ruminants and environment sources (352, 136 and 95 isolates respectively). Attribution tests were performed using the STRUCTURE software with 10 replicates. The corrected proportion of isolates attributed to each reservoir is represented after the model correction step based on the observed self-attribution results for each attribution test.

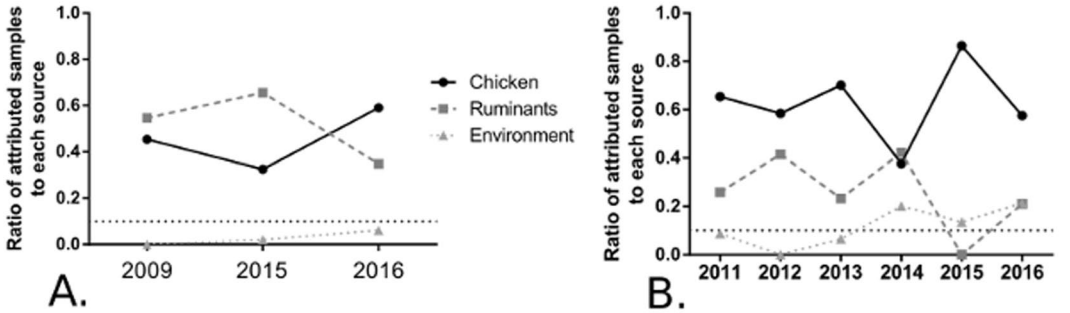

Figure 4. Attribution of French clinical isolates collected between 2009 and 2016 over time based on a collection. Attribution tests were performed using the STRUCTURE software with 10 replicates. A corrected proportion of isolates attributed to each reservoir (after model correction) is represented according to the year of isolation. (**A**) Attribution of French non-invasive clinical isolates collected in 2009 (n = 39), 2015 (n = 78) and 2016 (n = 26). (**B**) Attribution of French invasive clinical isolates collected in 2011 (n = 17), 2012 (n = 18), 2013 (n = 33), 2014 (n = 35), 2015 (n = 33) and 2016 (n = 37).

Discussion

The continued importance of campylobacteriosis as a major preventable cause of gastroenteritis means that effective monitoring is a high priority in many countries. *C. jejuni* infection is typically associated with contaminated food or drink but as peoples dietary habits vary by country, exposure and source of infection may also vary. This means that source attribution studies in, for example, the UK may not be an accurate reflection of the source of infection elsewhere. In this study we aimed to provide an improved understanding of *C. jejuni* source dynamics in France that incorporated the strengths of: (i) national scale reference laboratory surveillance; (ii) selection of genome-wide host segregating markers; (iii) probabilistic model (STRUCTURE) correction based upon self-attribution. This genomic epidemiology surveillance program was implemented to investigate the French clinical isolates from both invasive and non-invasive infection over the last 10 years.

As in previous studies in France[34] and other countries[8,23–25,40,41], chicken is a major source of infection in this study. However, among non-invasive strains isolated between 2014 and 2016, the overall proportion attributed to chicken was lower than previously described, only 39%. Attribution of non-invasive strains to chicken in 2016 increased and even if this trend does not continue, there is ample justification for the continuation of prevention efforts currently in place to control the rates of infection at the slaughterhouse level of the production chain[42]. Efforts to reduce the contamination of chicken at every step of the production chain would also help reducing the risks of campylobacteriosis[42].

Contaminated chicken is not the only source of campylobacteriosis and previous studies have highlighted ruminants as an important source[8,34,35]. Consistent with this, our study revealed ruminants to be a potentially important source. This could be related to French dietary habits in terms of meat consumption, with a more diverse diet compared to countries such as the United Kingdom[43]. Evidence of the importance of the ruminant reservoir could also be driven by the correction of model bias that was included in the analysis. Indeed, results for attribution in our study showed increased ruminant proportion after the correction step compared with the ruminant proportion before correction (Supp. Fig. 1). Recent chicken-ruminant host transitions leading to incorrect attribution of ruminant strains to chicken could reduce the signal of ruminant attribution in studies that are not correcting the attribution results.

The environment is not thought to be a major reservoir for infection, not least because *C. jejuni* is not thought to thrive outside the host gut. This is a composite niche reflecting contamination from multiple hosts. For this

Year	Invasive *C. jejuni*	Non-invasive *C. jejuni*
2009	—	39[a]
2011	17[b]	—
2012	18[b]	—
2013	33[b]	—
2014	35[b]	—
2015	1[a] + 32[b]	78[a]
2016	37[b]	26[b]
Total	**173**	**143**

Table 1. Clinical *Campylobacter jejuni* isolates from French patients. [a]Previously published isolates. [b]Newly sequenced isolates.

reason, attribution to this source is potentially more complex. However, after model correction self-attribution of environmental strains provided sufficient discrimination for attribution of clinical strains. Before 2014, the proportion of clinical strains attributed to the environment was below 10%, and increased to approximately 20% in 2014 and beyond. This recent increase may suggest that environmental strains reflect another infection reservoir. For example, *C. jejuni* has been isolated in seafood such as mussels in small proportions[17].

There is evidence for increased incidence of invasive disease caused by *C. jejuni*. This may be related to host factors or mobile genetic elements that confer virulence on particular strains. Neither of these were addressed here, however, it is also possible that certain reservoir sources have increased relative importance as a source of strains associated with invasive disease. In the clinical invasive strains isolated between 2014 and 2016, chicken were the major contributor (60% of cases), with 21% of strains attributed to ruminants and 19% to environment. This varied over time, with an increase in environmental attributed strains after 2014 from about 10% to about 20%, and a drop in the ruminant proportion particularly important in 2015.

Probabilistic attribution models, such the STRUCTURE based method used here, have considerable potential for improving understanding the epidemiology and spread of *Campylobacter* and can form an important part of reference surveillance and targeted interventions. However, there are limitations. First, source reservoirs are identified *a priori* excluding the possibility of attribution to unknown reservoirs. Second, models typically assume that the isolates from a given source are representative. Third, transition of strains between hosts can make definitive attribution difficult. Despite the strength of our method, highlighted by the high rates of self-attribution, we did not reach 100% self-attribution. This means that there is still a risk of erroneous attribution. Finally, source populations of different sizes may effect the probability of attribution of clinical isolates. In this study we used all the available data to maximize the reservoir strains available. As more isolates are sampled and genome sequenced from multiple sources the impact of these limitations will be reduced. Differences observed in attribution results between our study and previous ones could also be a consequence of the evolution of training datasets available to perform attribution studies at different times. Moreover, despite the homogeneous distribution of our source and clinical datasets verified in Fig. 1, there is no concordance of time and space between the source and clinical datasets which could introduce a bias.

In conclusion, this study confirmed not only the importance of the chicken reservoir, but also the importance of ruminants and the environment reservoirs for human campylobacteriosis in France. Furthermore, potential differences in the source of invasive and non-invasive clinical strains suggest that chickens may be a source of more serious infections. This study provides a basis for ongoing genomic epidemiology surveillance of *Campylobacter* in France, and reveal a need for investigation on genomics traits associated with invasive strains that will be carried out using GWAS methods[31,44].

Material and Methods

***C. jejuni* isolates and genome sequencing.** The collection of *C. jejuni* genomes from chicken, ruminants and the environment were obtained from previously published studies[33,34]. This collection was comprised of 352, 136 and 95 isolates from chicken, ruminants and the environment, respectively, including isolates from France and from the rest of the world (Supp. Table 1). Consistent with previous studies[23,31,40], isolates from caecal content, carcass, chicken farms or organs were grouped into a single "chicken" category in order to increase the numbers of strains included as training dataset to increase the efficiency of analyses. Strains from the environment linked to chicken or ruminant farms were included in the chicken or ruminant reservoirs, respectively. Previously published strains of *C. jejuni* isolated from patients in France in 2009 (40 strains)[33,34] and in 2015 (79 strains)[33] were included in our dataset. The provenance of the isolates (blood or stools) was traced back and confirmed as stools for 39 of the 40 strains from 2009 and for 78 of the 79 strains from 2015. One isolate from 2015 was isolated from blood. The isolate of unknown provenance from 2009 was not used in our study (Supp. Table 2).

A representative collection of clinical strains comprised 198 clinical strains isolated in France between 2011 and 2016 (Table 1). These isolates were received as single colonies by the CNRCH from French laboratories and hospitals participating in its surveillance network. In this dataset, 63 were sent by private laboratories and 135 were sent by public hospitals spread among 54 of the 102 French departments (Supp. Fig. 2). All invasive strains available to us for the studied years of isolation were selected, and a random selection of non-invasive strains isolated in 2016 was used to complete the collection of non-invasive strains publicly available. Upon reception, stocks of single colonies were maintained at −80 °C in brucella broth with 25% glycerol. Bacterial pellets were digested using MagNA Pure 96 DNA Bacteria Lysis Buffer and proteinase K. DNA extraction was performed on

a MagNA Pure 96 System (Roche Applied Science, Manheim, Germany) using the MagNA Pure 96 DNA and Viral NA SV Kit (Roche Applied Science). Quantification and purity checks (260/280 and 260/230 ratios) were determined by spectrophotometry (NanoDrop Technologies, Wilmington, DE, USA) before sequencing (performed by Helixio, Clermont-Ferrand, France). Qubit quantification was carried out prior to sequencing. Library preparation was made using the Nextera XT DNA Library Preparation Kit (Illumina Inc, San Diego, CA, USA) from 1 ng of DNA, and validation of the libraries was performed on the bioanalyzer with the High Sensitivity DNA Assay kit (Agilent, Santa Clara, CA, USA) in order to obtain sizes ranging from 250 to 1,500 bp. Paired-end sequencing was then performed on a NextSeq. 500 (Illumina Inc). Quality was controlled using FastQC v0.11.3[45]. De novo assemblies were produced using SPAdes (v3.10.1)[46]. An average of 20.6 contigs were obtained for the 198 sequenced strains, with a median value of 19 contigs. The average total size was 1,676,574 bp. (Supp. Table 3).

All of the genomic sequences, and associated information, were stored on a web-based Bacterial Isolates Genomic Sequences database (BIGSdb, http://zoo-dalmore.zoo.ox.ac.uk/)[47].

Genetic composition of the dataset. The BLAST algorithm implemented in BIGSdb was used to perform gene-by-gene alignment on the 899 *C. jejuni* genomes of our dataset using the 1,572 coding sequences from the reference strain NCTC 11168 (acc. Number: NC_002163.1). The concatenated alignment obtained for the core genes (present in at least 90% of the strains) was used to produce a phylogenetic tree using FastTree2 software annotated by iTOL v3[48]. MLST typing was performed automatically on all of the 899 strains using a MLST scheme implemented in BIGSdb.

Preparation of sequences for source attribution. Sequences for the 15 host-segregating markers were downloaded from BIGSdb and used for attribution (Supp. Table 4). The list of 15 host-segregating markers was used to perform a nucleotide BLAST on the 899 strains of our dataset using the genome comparator tool implemented in BIGSdb. The genome comparator tool attributed a unique identification number to each allele of the 15 host-segregating markers. The resulting matrix, identifying the allele present in each strain for each of the 15 host-segregating markers, was then re-formatted and used as input in STRUCTURE[49].

Self-attribution of isolates from the 3 putative sources and attribution of french clinical isolates using 15 host-segregating markers. Self-attribution tests were performed using only the 583 training dataset strains from the 3 putative sources. For these self-attribution tests, 20 isolates from each source or reservoir (chicken, ruminants and the environment) were randomly selected to constitute 3 test datasets. Strains belonging to the test datasets were flagged with a 0 using POPFLAG, and all remaining strains, constituting the training dataset were flagged with a 1 using POPFLAG. The origin of each strain (chicken, ruminant or environment) was indicated using POPDATA.

Attribution tests were performed using all 583 animal or environmental isolates as well as the strains of interest for each attribution test (French clinical strains from different years of isolation and different origin). Strains belonging to the test dataset (clinical strains) were flagged with a 0, and all source strains, constituting the training dataset were flagged with a 1 (POPFLAG parameter). The origin of each strain (chicken, ruminants, environment or clinical) was indicated (POPDATA parameter).

Analyses were performed with 100,000 burn-in cycles followed by 100,000 MCMC repetitions with the parameters using source population information (USEPOPINFO parameter) with no admixture model assumed and allele frequency independent model. All analyses were repeated 10 times to insure the reproducibility of the attribution test. Average scores for each attribution were considered.

Correction of attribution scores. The principle of correction made to the attribution is simple: it relies on the hypothesis that the errors FineStructure makes while self-attributing strains to a reservoir are also made when it attributes strains for which we don't know the provenance. That means that if when we give the software 100 strains of chicken, 100 of ruminants, 100 of environment (these 3 populations being the training dataset), and 100 of human (being the ones we want to test), if it correctly self-attributes the 300 stains from the training dataset, there is no problem as far as we know with the learning, and the correction will have absolutely no effect on the attribution of the human cases. But if the software is wrongly attributing 20 of ruminant isolates, by attributing them to the chicken group, that means that part of the human isolates attributed to chicken will actually more likely be ruminants. The correction step does that and the calculations were built accordingly. The results from STRUCTURE can be viewed in the form of a matrix presenting the proportion of membership of each pre-defined population in each of the source clusters. The first three rows correspond to the 3 groups from the training dataset (chicken, ruminant and the environment); the following rows correspond to the test dataset (Supp. Fig. 3). In out-of-the-bag results, if the proportion of correctly self-attributed strains from chicken, ruminant and environment populations (respectively C_C, R_R and E_E) are below 0.9, there is a bias introduced in the proportion of clinical strains attributed to chicken, ruminant or environment (respectively T_C, T_R and T_E) due to the presence of samples in the training dataset that are similar to samples from another source. As this was the case here, a system of equations (Supp. Fig. 4) was implemented in order to correct this bias. Specifically, the proportion of strains wrongly attributed from the training dataset was used to estimate the proportion of strains wrongly attributed in the tested population.

This system was solved using an online equation solver (https://matrixcalc.org). Unbiased numbers of isolates (T_C*, T_R*and T_E*) were then turned into proportions based on the number of isolates from the test dataset N.

Data Availability

All 198 newly sequenced genomes were deposited in the Genbank and SRA public databases under the BioProject PRJNA497209.

References

1. Kaakoush, N. O., Castaño-Rodríguez, N., Mitchell, H. M. & Man, S. M. Global Epidemiology of *Campylobacter* Infection. *Clin. Microbiol. Rev.* **28**, 687–720 (2015).
2. Platts-Mills, J. A. & Kosek, M. Update on the burden of *Campylobacter* in developing countries. *Curr. Opin. Infect. Dis.* **27**, 444–450 (2014).
3. Louwen, R. *et al*. *Campylobacter* bacteremia: A rare and under-reported event? *Eur. J. Microbiol. Immunol.* **2**, 76–87 (2012).
4. Feodoroff, B. *et al*. Clonal distribution and virulence of *Campylobacter jejuni* isolates in blood. *Emerg. Infect. Dis.* **19**, 1653–5 (2013).
5. Bessède, E., Lehours, P., Labadi, L., Bakiri, S. & Mégraud, F. Comparison of characteristics of patients infected by Campylobacter jejuni, Campylobacter coli, and Campylobacter fetus. *J. Clin. Microbiol.* **52**, 328–30 (2014).
6. Chlebicz, A. & Śliżewska, K. Campylobacteriosis, Salmonellosis, Yersiniosis, and Listeriosis as Zoonotic Foodborne Diseases: A Review. *Int. J. Environ. Res. Public Health* **15** (2018).
7. Nyati, K. K. & Nyati, R. Role of *Campylobacter jejuni* Infection in the Pathogenesis of Guillain-Barré Syndrome: An Update. *Biomed Res. Int.* **2013**, 1–13 (2013).
8. Skarp, C. P. A., Hänninen, M.-L. & Rautelin, H. I. K. Campylobacteriosis: the role of poultry meat. *Clin. Microbiol. Infect.* **22**, 103–109 (2016).
9. Epps, S. *et al*. Foodborne *Campylobacter*: Infections. *Metabolism, Pathogenesis and Reservoirs. Int. J. Environ. Res. Public Health* **10**, 6292–6304 (2013).
10. Thépault, A. *et al*. Prevalence of Thermophilic *Campylobacter* in Cattle Production at Slaughterhouse Level in France and Link Between *C. jejuni* Bovine Strains and Campylobacteriosis. *Front. Microbiol.* **9**, 471 (2018).
11. Jensen, A. N., Dalsgaard, A., Baggesen, D. L. & Nielsen, E. M. The occurrence and characterization of *Campylobacter jejuni* and *C. coli* in organic pigs and their outdoor environment. *Vet. Microbiol.* **116**, 96–105 (2006).
12. Wieczorek, K. & Osek, J. Antimicrobial Resistance and Genotypes of *Campylobacter jejuni* from Pig and Cattle Carcasses Isolated in Poland During 2009–2016. *Microb. Drug Resist.* **24**, 680–684 (2018).
13. Hepworth, P. J. *et al*. Genomic variations define divergence of water/wildlife-associated *Campylobacter jejuni* niche specialists from common clonal complexes. *Environ. Microbiol.* **13**, 1549–1560 (2011).
14. Sheppard, S. K. *et al*. Niche segregation and genetic structure of *Campylobacter jejuni* populations from wild and agricultural host species. *Mol. Ecol.* **20**, 3484–3490 (2011).
15. Acke, E. Campylobacteriosis in dogs and cats: a review. *N. Z. Vet. J.* **66**, 221–228 (2018).
16. Nilsson, A. *et al*. Genomic and phenotypic characteristics of Swedish *C. jejuni* water isolates. *PLoS One* **12**, e0189222 (2017).
17. Rincé, A. *et al*. Occurrence of Bacterial Pathogens and Human Noroviruses in Shellfish-Harvesting Areas and Their Catchments in France. *Front. Microbiol.* **9**, 2443 (2018).
18. Dingle, K. E. *et al*. Multilocus sequence typing system for *Campylobacter jejuni*. *J. Clin. Microbiol.* **39**, 14–23 (2001).
19. Méric, G. *et al*. A reference pan-genome approach to comparative bacterial genomics: identification of novel epidemiological markers in pathogenic *Campylobacter*. *PLoS One* **9**, e92798 (2014).
20. Sheppard, S. K., Guttman, D. S. & Fitzgerald, J. R. Population genomics of bacterial host adaptation. *Nat. Rev. Genet.* **19**, 549–565 (2018).
21. Sheppard, S. K. *et al*. Cryptic ecology among host generalist *Campylobacter jejuni* in domestic animals. *Mol. Ecol.* **23**, 2442–2451 (2014).
22. Gripp, E. *et al*. Closely related *Campylobacter jejuni* strains from different sources reveal a generalist rather than a specialist lifestyle. *BMC Genomics* **12**, 584 (2011).
23. Sheppard, S. K. *et al*. *Campylobacter* Genotyping to Determine the Source of Human Infection. *Clin. Infect. Dis.* **48**, 1072–1078 (2009).
24. Muellner, P. *et al*. Molecular-based surveillance of campylobacteriosis in New Zealand–from source attribution to genomic epidemiology. *Euro Surveill.* **18** (2013).
25. Rosner, B. M. *et al*. A combined case-control and molecular source attribution study of human *Campylobacter* infections in Germany, 2011–2014. *Sci. Rep.* **7**, 5139 (2017).
26. Dearlove, B. L. *et al*. Rapid host switching in generalist *Campylobacter* strains erodes the signal for tracing human infections. *ISME J.* **10**, 721–9 (2016).
27. Woodcock, D. J. *et al*. Genomic plasticity and rapid host switching can promote the evolution of generalism: a case study in the zoonotic pathogen *Campylobacter*. *Sci. Rep.* **7**, 9650 (2017).
28. Motro, Y. & Moran-Gilad, J. Next-generation sequencing applications in clinical bacteriology. *Biomol. Detect. Quantif.* **14**, 1–6 (2017).
29. Méric, G. *et al*. Convergent Amino Acid Signatures in Polyphyletic *Campylobacter jejuni* Subpopulations Suggest Human Niche Tropism. *Genome Biol. Evol.* **10**, 763–774 (2018).
30. Sheppard, S. K. *et al*. Genome-wide association study identifies vitamin B5 biosynthesis as a host specificity factor in. *Campylobacter. Proc. Natl. Acad. Sci. USA* **110**, 11923–7 (2013).
31. Yahara, K. *et al*. Genome-wide association of functional traits linked with *Campylobacter jejuni* survival from farm to fork. *Environ. Microbiol.* **19**, 361–380 (2017).
32. Pascoe, B. *et al*. Enhanced biofilm formation and multi-host transmission evolve from divergent genetic backgrounds in *Campylobacter jejuni*. *Environ. Microbiol.* **17**, 4779–89 (2015).
33. Thépault, A. *et al*. Genome-Wide Identification of Host-Segregating Epidemiological Markers for Source Attribution in *Campylobacter jejuni*. *Appl. Environ. Microbiol.* **83** (2017).
34. Thépault, A. *et al*. Ruminant and chicken: important sources of campylobacteriosis in France despite a variation of source attribution in 2009 and 2015. *Sci. Rep.* **8**, 9305 (2018).
35. de Haan, C. P. A., Kivisto, R., Hakkinen, M., Rautelin, H. & Hanninen, M. L. Decreasing Trend of Overlapping Multilocus Sequence Types between Human and Chicken *Campylobacter jejuni* Isolates over a Decade in Finland. *Appl. Environ. Microbiol.* **76**, 5228–5236 (2010).
36. Sheppard, S. K. *et al*. Host Association of *Campylobacter* Genotypes Transcends Geographic Variation. *Appl. Environ. Microbiol.* **76**, 5269–5277 (2010).
37. Vidal, A. B. *et al*. Genetic Diversity of *Campylobacter jejuni* and *Campylobacter coli* Isolates from Conventional Broiler Flocks and the Impacts of Sampling Strategy and Laboratory Method. *Appl. Environ. Microbiol.* **82**, 2347–2355 (2016).
38. Oh, J.-Y. *et al*. Epidemiological relationships of *Campylobacter jejuni* strains isolated from humans and chickens in South Korea. *J. Microbiol.* **55**, 13–20 (2017).
39. Ramonaite, S., Tamuleviciene, E., Alter, T., Kasnauskyte, N. & Malakauskas, M. MLST genotypes of *Campylobacter jejuni* isolated from broiler products, dairy cattle and human campylobacteriosis cases in Lithuania. *BMC Infect. Dis.* **17**, 430 (2017).
40. Sheppard, S. K. *et al*. *Campylobacter* genotypes from food animals, environmental sources and clinical disease in Scotland 2005/6. *Int. J. Food Microbiol.* **134**, 96–103 (2009).
41. Ravel, A. *et al*. Source attribution of human campylobacteriosis at the point of exposure by combining comparative exposure assessment and subtype comparison based on comparative genomic fingerprinting. *PLoS One* **12**, e0183790 (2017).
42. Meunier, M., Guyard-Nicodème, M., Dory, D. & Chemaly, M. Control strategies against *Campylobacter* at the poultry production level: biosecurity measures, feed additives and vaccination. *J. Appl. Microbiol.* **120**, 1139–1173 (2016).

43. Gallay, A. *et al.* Risk Factors for Acquiring Sporadic *Campylobacter* Infection in France: Results from a National Case-Control Study. *J. Infect. Dis.* **197**, 1477–1484 (2008).
44. Berthenet, E. *et al.* A GWAS on *Helicobacter pylori* strains points to genetic variants associated with gastric cancer risk. *BMC Biol.* **16**, 84 (2018).
45. Wingett, S. W. & Andrews, S. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Research* **7**, 1338 (2018).
46. Bankevich, A. *et al.* SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
47. Jolley, K. A. & Maiden, M. C. J. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* **11**, 595 (2010).
48. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* **44**, W242–W245 (2016).
49. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–59 (2000).

Acknowledgements

We thank Lindsay Mégraud for her help with correcting the manuscript. We also thank Michèle Gourmelon for her work isolating the environmental strains, as well as everyone who participate in some way to increase the size and quality of the isolates publically available. Their effort makes studies like ours exist. This work was supported by internal funding of the CNRCH. SS is funded by UK Medical Research Council (MRC) grants MR/L015080/1, MR/M501608/1 and G0801929, Biotechnology and Biological Sciences Research Council (BBSRC) grant BB/I02464X/1, and the Wellcome Trust.

Author Contributions

E.B. performed the analyses and wrote the manuscript. A.T., M.C. and K.R. advised the first author on the attribution experiments. A.D., A.B., L.B., E.B. and F.M. helped gathering the clinical samples used in this study. S.K.S. advised the first author on the analyses of results. P.L. originated and supervised this study. All the authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-019-44454-2.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.