Evidence of land-sea transfer of the zoonotic pathogen Campylobacter to a wildlife marine sentinel species

Received Date : 06-Aug-2014

Revised Date : 11-Nov-2014

Accepted Date : 13-Nov-2014

Article type : Original Article

Evidence of land-sea transfer of the zoonotic pathogen *Campylobacter* to a wildlife marine sentinel species

Johanna L. Baily[1,2]#, Guillaume Méric[3], Sion Bayliss[3]§, Geoffrey Foster[4], Simon E. Moss[2], Eleanor Watson[1], Ben Pascoe[3,5], Jane Mikhail[3], Romain Pizzi[6,7], Robert J. Goldstone[8], David G.E. Smith[1,8], Kim Willoughby[1], Ailsa J. Hall[2], Samuel K. Sheppard[3,5,9]* and Mark P. Dagleish[1]*

[1]Moredun Research Institute, Pentlands Science Park, Bush Loan, Penicuik;

[2]SeaMammal Research Unit, University of St Andrews;

[3]College of Medicine, Institute of Life Science, Swansea University, Swansea;

[4]Scottish Rural College, Inverness;

[5]MRC CLIMB Consortium, Institute of Life Science, Swansea University, Swansea;

[6]Royal Zoological Society of Scotland, Edinburgh;

[7]Scottish SPCA National Wildlife Rescue Centre, Fishcross, Alloa;

[8]Institute of Infection, Immunity & Inflammation, University of Glasgow, Glasgow;

[9]Department of Zoology, University of Oxford, Oxford, United Kingdom;

Present address: College of Medicine and Veterinary Medicine, Royal (Dick) School of Veterinary Studies, University of Edinburgh, United Kingdom.

§ Present address: School of Biology and Biochemistry, Bath University, Bath, United Kingdom.

*Corresponding authors contributed equally to this manuscript: Samuel K. Sheppard: s.k.sheppard@swansea.ac.uk; Mark P. Dagleish: mark.dagleish@moredun.ac.uk

Keywords: environmental health, *Campylobacter*, grey seals, source attribution, genomics

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/mec.13001

Abstract

Environmental pollution often accompanies the expansion and urbanisation of human populations where sewage and wastewaters commonly have an impact on the marine environments. Here we explored the potential for faecal bacterial pathogens, of anthropic origin, to spread to marine wildlife in coastal areas. The common zoonotic bacterium *Campylobacter* was isolated from grey seals (*Halichoerus grypus*), an important sentinel species for environmental pollution, and compared to isolates from wild birds, agricultural sources and clinical samples to characterize possible transmission routes. *Campylobacter jejuni* was present in half of all grey seal pups sampled (24/50 dead and 46/90 live pups) in the breeding colony on the Isle of May (Scotland), where it was frequently associated with histological evidence of disease. Returning yearling animals (19/19) were negative for *C. jejuni* suggesting clearance of infection whilst away from the localised colony infection source. The genomes of 90 isolates from seals were sequenced and characterized using a whole-genome multilocus sequence typing (MLST) approach, and compared to 192 published genomes from multiple sources using population genetic approaches and a probabilistic genetic attribution model to infer the source of infection from MLST data. The strong genotype-host association has enabled the application of source attribution models in epidemiological studies of human campylobacteriosis, and here assignment analyses consistently grouped seal isolates with those from human clinical samples. These findings are consistent with either a common infection source or direct transmission of human *campylobacter* to grey seals, raising concerns about the spread of human pathogens to wildlife marine sentinel species in coastal areas.

Introduction

Marine mammals can act as sentinel species with their health reflecting that of the wider coastal marine ecosystem (Bossart 2011; Moore 2008; Reddy *et al.* 2001). Contamination of environmental waters is commonly associated with sewage either from municipal waste water treatment works or runoff from agricultural land. Several studies have shown that marine mammals can be infected with pathogens of known terrestrial origin and this may be the result of faecal to oral transmission via sewage. For example, multidrug resistant *Escherichia coli* have been isolated from free-ranging bottlenose dolphins (*Tursiops truncatus*) in South Carolina and Florida (Greig *et al.* 2007) and *Toxoplasma gondii* transmitted from cat faeces is currently a major cause of mortality in the threatened southern sea otters (*Enhydra lutris nereis*) (Conrad *et al.* 2005; Shapiro *et al.* 2012). Two zoonotic

gastrointestinal pathogens, *Campylobacter* and *Salmonella*, have also been found in juvenile northern elephant seals, with a higher prevalence in seals that became stranded along the coast of central California than in seals on their natal beaches that had never entered the water (Stoddard *et al.* 2005).

Campylobacter is of particular importance as it is currently a major public health concern in the UK, being the most commonly reported cause of bacterial gastroenteritis, with over 500,000 community cases, 20,000 hospitalisations and 110 deaths estimated to occur annually (Strachan & Forbes 2010; Tam *et al.* 2012). Despite this, the epidemiology of this organism is poorly understood as most human *Campylobacter* infections are sporadic and the source of outbreaks can be difficult to detect. A large number of mammalian and avian hosts (wild and domestic) have been found to carry *Campylobacter* and transmission from reservoir hosts to humans has been well studied (Sheppard *et al.* 2009a; Sheppard *et al.* 2009b; Strachan *et al.* 2009). However, little is known about the transmission of *Campylobacter* to wild animals.

Campylobacter populations are highly structured into clusters of related lineages that reflect the clonal ancestry under the influence of mutation, horizontal gene transfer and natural selection. Multilocus sequence typing (MLST) of *Campylobacter jejuni* and *C. coli* has proved effective for examining this, and assigning genotypes that share 4 or more alleles at 7 MLST loci to the same 'clonal complex' (Dingle *et al.* 2005). By characterizing isolates from multiple populations in this way it became clear that there is substantial genetic differentiation between *C. jejuni* populations from different host species (Miller *et al.* 2006; Miller *et al.* 2005; Sheppard *et al.* 2011). Some clonal complexes are strongly host associated while others display a generalist distribution (Gripp *et al.* 2011; Sheppard *et al.* 2011). More recently, whole genome analysis of *C. jejuni* populations has built on this research to show accessory genome differences between lineages suggesting that host adaptation may be reflected in the genome (Sheppard *et al.* 2013b) and that different lineages may occupy different niches within a single host (Sheppard *et al.* 2014).

Quantitative information about the degree of niche segregation among isolates from different host species is useful for understanding *Campylobacter* ecology and evolution but also has another practical use. Specifically, the segregation of gene pools allows the attribution of isolates to original source populations. Population genetic attribution models have been used

to attribute human *Campylobacter* infection to host species of origin, based on reference data sets from a range of animal species and the environment (Boysen *et al.* 2013; de Haan *et al.* 2010; Kittl *et al.* 2013; Mullner *et al.* 2009; Sheppard *et al.* 2009b; Wilson *et al.* 2008). These models, which generally implicate poultry as a major source of human infection, are robust to temporal and spatial variation in the data because of the strength of the host association signal (Sheppard *et al.* 2010), and the increasingly large MLST data archives for isolates from multiple sources (Jolley & Maiden 2010) provide a basis for investigation of the transmission of isolates to other host species.

Grey seals (*Halichoerus grypus*) are top predators in the British marine environment and are potential sentinels for the health of marine food webs. Until now, little work has been done to investigate *Campylobacter* in this species. In this study we investigate the prevalence of *Campylobacter* infection in neonatal and juvenile Scottish grey seals from a breeding colony on the Isle of May (Scotland, UK) and relate this information to evidence of pathogenicity. Using MLST and comparative genomics, the seal-associated *Campylobacter* population is analysed and a quantitative source attribution model is used to determine the likely origin of *C. jejuni* infections and investigate evidence of a possible land-sea transfer of this important zoonotic pathogen.

Materials and Methods

Bacterial sampling from grey seals

During autumn 2011, rectal swabs were taken from 90 wild-caught live grey seal pups and 19 live wild-caught yearling grey seals on their natal colony of the Isle of May (56° 11′ 9″ N, 2° 33′ 27″ W), as well as 32 live grey seal pups that had become stranded along the Scottish coastline. Live free-ranging grey seal pups were sampled on their natal colony from 3 sites with different ground substrates (tidal boulder beach, muddy grassy slope and rocky stagnant pools), each at 3 different time-points (early, mid and late pupping season). Stranded seals were sampled within 24 hours of admission for rehabilitation to the Scottish Society for the Prevention of Cruelty to Animals (SPCA) National Wildlife Centre, in Fife, Scotland. In addition, rectal swabs were taken from 50 dead free-ranging grey seal pups found on the Isle of May and 9 stranded grey seal pups that died or were euthanized on humane grounds at the Scottish SPCA National Wildlife Centre. All dead pups (colony and stranded) were sampled within 48h of death. Amongst the stranded seals, 9 animals died and the delay in processing

results provided unreliable results, so these were not further analysed. Three sediment samples were also taken from each of two pupping locations within the colony (muddy grassy slope and rocky stagnant pools); samples from the boulder beach were not taken due to the nature of the substrate. All sampling of live free-ranging animals was carried out under UK Home Office Project (No. 60/4009) and Personal Licences as issued to the Sea Mammal Research Unit under the Animals (Scientific Procedures) Act, 1986. Stranded grey seal pups were sampled within 24h of arrival at the rehabilitation centre as part of the routine health assessment procedure. Date of sampling and location of sampling/stranding was recorded for each animal. Rectal swabs were placed in Amies transport medium with charcoal (Medical Wire and Equipment (MWE), Corsham, UK) and processed within 8h in the field laboratory on the Isle of May for the free-ranging live pups, dead pups and live yearlings. Swabs from stranded live or dead grey seal pups were sent by first class post to SAC Consulting Veterinary Services, Inverness, where they were processed, incurring a delay of up to 4 days between sampling and processing in some cases.

Biochemical identification of *Campylobacter* sp.

Rectal swabs were plated onto *Campylobacter* selective blood-free agar (CCDA, Oxoid, Basingstoke, UK), incubated at 37°C under a microaerophilic environment (Campygen sachets, Oxoid) in airtight containers and assessed at 48 hours, 4 days and 6 days post inoculation for suspect *Campylobacter* colonies. From each plate, a maximum of 3 distinct colonies were selected for further identification. *Campylobacter* identification was performed using Gram stain and a wide range of phenotypical tests (On 1996; On & Holmes 1991, 1992). Briefly, biochemical assays tested the activity of catalase (Biomerieux, Basingstoke, UK), oxidase (1% w/v N-N-N'-N'- tetramethyl-p-phenylenediamine dihydrochloride, Thermo Fisher Scientific, Loughborough, UK); hippuricase (hippurate hydrolysis) (Rosco Diagnostica, Taastrup, Denmark); alkaline phosphatase (Rosco); urease (Rosco); gamma glutamyl aminopeptidase (Bioconnections, Leeds, UK); acetate esterase (Rosco) and the ability to reduce nitrate (Rosco). Sensitivity to nalidixic acid (Oxoid), cefoperazone (Oxoid), cephalexin (Oxoid) and bile (Rosco) along with H_2S production on Triple Sugar Iron (TSI) slopes (Oxoid) was assessed under a microaerophilic environment. Growth at 22°C, 25°C and 42°C on Columbia agar with sheep blood (Oxoid) was assessed. In addition, and growth on 4 different agar media (MacConkey agar (Oxoid), nutrient agar (Oxoid) supplemented with 1% glycine (Scientific Laboratory Supplies, Nottingham, UK), nutrient agar supplemented with 1.5% NaCl (Scientific Laboratory Supplies)) and nutrient agar supplemented with 3.5% NaCl

was assessed under a microaerophilic environment. Growth was also assessed under anaerobic (Oxoid), CO_2 enriched (Oxoid) and aerobic atmospheres. For each individual animal from which *Campylobacter* was isolated, one to three strains were selected for whole genome sequencing on the basis of their phenotypic characteristics: arbitrarily hippurate hydrolysis; reduction of nitrate; growth on agar supplemented with 1% glycine and resistance to nalidixic acid. In total, 90 strains of *Campylobacter* were selected from 74 seals and 2 sediment samples. The prevalence of *Campylobacter* in different conditions was analysed using a generalized linear model (GLM) with binomial errors to compare sampling site for live pups on the colony, time point for live pups on the colony, weight, sex, stranded vs. colony, group (dead colony, dead rehab, live colony, live rehab, yearling colony) and age for pup and yearling on the colony only. The significance of individual comparisons of prevalence were tested using Fisher's exact tests, for example between *Campylobacter* in pups with and without colitis.

Histopathology and immunohistochemistry of dead grey seal pups

A full post-mortem examination including extensive histopathology was performed on each of the 50 dead animals from the colony, from which bacteria were sampled. Samples of large intestine from all these 50 dead pups were fixed in 10% neutral buffered formalin and processed routinely to paraffin wax. Sections (5µm) were stained with haematoxylin and eosin (HE) and graded as to the degree of colitis in each section (normal to minimal, mild, moderate to severe). Normal to minimal inflammation of colonic mucosa was characterised by small numbers (≤5) of lymphocytes and plasma cells in the lamina propria between crypts with well differentiated crypt epithelium. Mild inflammation was characterised by an increased number of inflammatory cells, which filled the inter-cryptal region and mildly increased separation of crypts. Mild crypt epithelial dysplasia was present with occasional colonic crypts dilated by cellular debris. Moderate to severe inflammation was characterised by inflammatory cells filling the inter-cryptal region and moderately to markedly increasing separation of crypts with moderate to marked crypt epithelial dysplasia and occasional crypt abscessation.

Fisher's exact tests were performed in R to assess the association of colitis with *Campylobacter* presence. Immunohistochemistry was performed on selected sections of large intestine to detect *Campylobacter* (100 µl, dilution 1:1000, clone: Ab54125, mouse monoclonal, anti-*Campylobacter* primary antibody, BGN/2E10, 1 mg ml^{-1}, Abcam Plc,

Cambridge, UK) as previously described (Haddock *et al.* 2010). Immunohistochemistry was analysed with reference to the effect of *Campylobacter jejuni* on the caecum and colon of experimentally infected pigs with positive and negative control tissues. Negative-control sections were incubated with isotype matched, IgM, mouse Ig(dilution 1:500, Sigma M5909, 0.2 mg ml^{-1}) in place of the primary antibodies. The immunoreactions were visualised with Nova red (Vector Laboratories, Peterborough, UK; 10 mins at RT) and sections were counterstained with Mayer's haematoxylin.

Whole-genome sequencing

Two sets of *C. jejuni* and *C. coli* genomes were used in this comparative population study: (a) 74 *C. jejuni* and 14 *C. coli* isolate genomes from seals and (b) 131 *C. jejuni* and 61 *C. coli* previously published genomes (Sheppard *et al.* 2014; Sheppard *et al.* 2013a; Sheppard *et al.* 2013b) for a total of 280 whole genome sequences (**Table S1**). *Campylobacter* isolates from seals were subcultured and grown on selective CCDA agar, and their total DNA was extracted using the Masterpure DNA purification kit (Epicentre, Madison, WI, USA) according to the manufacturer's instructions with minor modifications (pellets were incubated at 65°C with proteinase K and tissue lysis buffer for 30 minutes; ribonuclease (RNase) incubation was extended to 2 hours and elution was carried out with 100µl elution buffer). Sequencing was performed using an Illumina MiSeq sequencer at Glasgow Polyomics, Wolfson Wohl Cancer Research Centre, Glasgow, UK. A multiplex sequencing approach was used, involving 12 separately tagged libraries sequenced simultaneously in two lanes of an eight channel GAII flow cell. The standard Illumina Indexing protocol involved fragmentation of 2 µg genomic DNA by acoustic shearing to enrich for 200 bp fragments, A-tailing, adapter ligation and an overlap extension PCR using the Illumina 3 primer set to introduce specific tag sequences between the sequencing and flow cell binding sites of the Illumina adapter. DNA cleanup was carried out after each step to remove DNA <150 bp using a 1:1 ratio of AMPure® paramagnetic beads (Beckman Coulter, Inc., USA), and a qPCR was used for final DNA quantification. Contiguous sequences of 10-200 kb were assembled *de novo* for each isolate individually using the Velvet software (Zerbino & Birney, 2008), using default parameters to generate the consensus sequences. The average number of contiguous sequences (contigs) for all assemblies was 138.6 (average N_{50}=84,457 bp; average N_{90}=32,101 bp). All genome sequences were archived and analysed using the gene-by-gene approach implemented in the BIGSdb platform (Jolley & Maiden 2010). Assembled genome data and individual genome coverage information are available in the Dryad repository

(doi:10.5061/dryad.8p984). Short reads were uploaded to the Short Read Archive (SRA) on the NCBI repository.

Phylogenetic analyses

A reference pan genome approach was used to characterise the isolate genome contiguous sequence files (Meric *et al.* 2014). Briefly, gene-by-gene alignments were constructed, using the MUSCLE software (Edgar 2004), for 2,335 *Campylobacter* genes that constituted a reference pan-genome list. This list was composed of the genes from 7 *Campylobacter* genomes: *C. jejuni subsp. jejuni* strains NCTC11168, 81–176, 81116 and M1; *C. jejuni subsp. doylei* strain 269.97; *C. coli* strains 76339 and CVM N2970. Duplicate genes were excluded, and these were defined as having >70% sequence identity over ≥ 10% of the gene sequence (Meric *et al.* 2014). Allele numbers were assigned for each locus, including the determination of 7-locus MLST allelic profiles. Phylogenetic trees were constructed using concatenated sequences of 595 core genes shared by all *C. jejuni* and *C. coli* isolates in this study (Sheppard *et al.* 2014). Sequences were aligned gene-by-gene and concatenated into a single 362,598 bp sequence alignment. An approximation of the maximum likelihood algorithm was used to reconstruct the trees, using FastTree2 (Price *et al.* 2010). Trees were visualised from the Newick output of FastTree2 using MEGA6 (Tamura *et al.* 2013).

Representative isolate collection for source attribution

Seven-locus sequence types (STs) derived from genome sequenced seal isolates (n=74) were compared to STs of *C. jejuni* isolates from 5 source populations from previously published datasets (Sheppard *et al.* 2010; Sheppard *et al.* 2011). These included: clinical faecal samples (n=1298); chicken meat and faeces (n=1298); cattle faeces (n=597); sheep faeces (n=250); wild bird faeces (n=247). The genetic heterogeneity within and between groups was analysed using φ-statistics with analyses of molecular variance (Excoffier & Lischer 2010). The seven loci were concatenated and the number of polymorphic sites was determined. Permutation tests were used to assess significance, using 999 permutations. Genetic differentiation between groups, rather than within groups, and significance was performed using pairwise nested analysis of molecular variance (AMOVA). Low F-statistic values indicated low genetic differentiation.

Source attribution modelling

Source attribution and associated analyses were performed using iSource, also termed the asymmetric island model (Sheppard *et al.* 2009b; Wilson *et al.* 2008). Within the iSource evolutionary model, *C. jejuni* STs from different hosts are considered to represent discrete islands where there is homogenous mixing within the islands and some migration between them. The model is a generalisation of Wright's island model, called the migration matrix model (Kimura & Weiss 1964), and incorporates an evolutionary model where loci are considered to be in linkage disequilibrium using a recombination model suitable for bacteria. Here, seal isolates were assumed to reflect their source populations and non-seal populations were used to estimate migration, mutation and recombination. These estimates were used to generate posterior probability (F) and assign a source for test isolates with 95% confidence intervals (C.I.), estimated from the mean of the posterior distribution. The model was run with 100,000 iterations and the state of the Markov Chain Monte Carlo (MCMC) was sampled every 50 iterations and each run utilised a symmetrical Dirichlet prior with α of 1. For subsequent analysis a 1000 iteration burn-in was applied.

The robustness of the attribution model for application to the *C. jejuni* isolates used in the present study was tested using empirical cross-validation, also termed self-attribution. This model testing allowed for a number of performance indicators to be assessed including: model sensitivity to sample size variation; the suitability of the resulting approximations to be used for inference; the robustness of the approach to genetic heterogeneity. This empirical cross-validation was performed on 100 semi-randomised datasets. In each cross-validation iteration, approximately half of the STs from one of the source populations was removed, designated as test isolates, and assigned to the remaining source populations. The accuracy of the model for correct self-assignment was quantified, for example the number of times that test isolates from cattle were correctly assigned to cattle. A total of 100 test isolate datasets underwent source attribution using standard attribution settings. From the resulting outputs three performance indicators were calculated: predicted proportion of isolates correctly assigned, bias and root mean square error (RMSE). These were calculated as previously detailed (Wilson *et al.* 2008).

Results

Prevalence of *Campylobacter* in grey seals

Historically, the cause of death and prevalence of umbilical infection has been reported to vary for seals colonising different areas of the Isle of May breeding colony (Baker & Baker 1988), such as rocky stagnant pools, muddy grassy slopes or tidal boulder beaches (**Figure 1**). Here, we investigated the prevalence of *Campylobacter* in seals from these three areas of the Isle of May breeding colony. *Campylobacter* was recorded in 48.0% (24/50) dead and 51.1% (46/90) live grey seal pups on the colony (**Figure 1; Table S2**). *Campylobacter* was isolated from 11.1% (1/9) stranded dead and 12.5% (4/31) stranded live seals (**Table S2**). *Campylobacter* was not isolated from any of the returning yearling animals (0/19) but from 2/6 sediment samples, both taken from the stagnant rocky pools. *Campylobacter jejuni*, *C. coli* and *C. lari* were identified amongst the isolates. Among free-ranging live grey seal pups, those sampled at the tidal boulder beach site were 2 to 3 times less likely to harbour *Campylobacter* when compared with seals sampled at the muddy, grassy slope site (OR=1.97; p=0.197) or rocky stagnant pool site (OR=2.98; p =0.04) (**Table S2**). Sampling time also influenced the likelihood of isolating *Campylobacter* spp. from live free-ranging grey seal pups with a statistically significantly lower prevalence in the mid-season when compared to both early (p=0.021) and late season (p=0.001) (**Table S2**). There was no correlation with sex or pup developmental stage.

Campylobacter was infrequently isolated from grey seals admitted for rehabilitation to the Scottish SPCA National Wildlife Centre, in Fife, Scotland (Table S2). However, we did not compare the sampling from these animals with the sampling performed on the Isle of May, as although we sampled the animals within 24h of their admission, the delay between sampling and processing was long and variable for logistic reasons, and could have affected the counts. Also, these animals had typically been found by members of the public and transported back to the rehabilitation centre. Consequently, they could have been exposed to many possible other sources of contamination and are far from a closed population, as on the Isle of May.

Intestinal histopathological and immunohistochemical examination of dead grey seals

Histological examination of sections of large intestine showed a statistically significant correlation between the isolation of *Campylobacter* spp. from rectal swabs and the presence of moderate to severe colitis in dead grey seal pups on the Isle of May (Fisher's exact test, p=0.02). Using specific immunohistochemistry, we observed that *Campylobacter* bacteria

were located within sections of the large intestine (**Figure 2A**). Well-defined curved bacterial organisms, consistent with *Campylobacter*, labelled specifically with IHC were located within the lumen and intestinal crypts (**Figure 2A**), principally on the apical surface of the enterocytes. The lack of invasion into enterocytes suggests that *Campylobacter* retains a relatively superficial location which nonetheless, given the general level of infection, could lead to significant faecal fluid loss and dehydration by acceleration of intestinal transit and malabsorbtion (Gelberg 2007). This may have contributed to the death of the grey seal pups although no causal link was made.

Population structure of *C. jejuni* and *C. coli* from seals

Phylogenetic trees were constructed using an approximation of the maximum-likelihood algorithm using 362,598 bp-long alignments from genes shared by *C. jejuni* and *C. coli* (Sheppard *et al.* 2014) which were extracted from 74 seal and 131 agricultural and non-agricultural *C. jejuni* genomes, and 14 seal and 61 agricultural and non-agricultural *C. coli* genomes (**Table S1, Figure 3**). *Campylobacter* isolates from seals did not form an isolated cluster, separate from isolates from other sources. *C. jejuni* isolates from seals belonged to the ST-45 clonal complex (34 isolates), the ST-21 complex (22 isolates) and the ST-22 complex (14 isolates). Additionally, two isolates belonged to two distinct clonal complexes (ST-1034 and ST-1332 complex) (**Figure 3A**). All *C. coli* isolates from seals belonged to the introgressed clade 1 ST-828 clonal complex (Sheppard *et al.* 2013a; Sheppard *et al.* 2008), with 12 ST-827 isolates (**Figure 3B**). For both species, isolates from seals were always clustered with agricultural *Campylobacter*, indicating very recent divergences between them, and consistent with the recent spread of isolates. There was some differentiation among the genome types isolated from different parts of the colony with higher proportions of ST-583 and ST-827 isolates being isolated from dead pups compared to live pups but more sampling would be necessary to make robust inference of spatial variation.

Core and accessory genome analysis of seal and agricultural *C. jejuni* isolates

A total of 2,177 genes from a larger *Campylobacter* reference pan-genome (Meric *et al.* 2014) were identified using BLAST in 205 *C. jejuni* isolates and 75 *C. coli* from source populations (clinical, chicken, cattle, wild birds) and seals. Due to the low number of isolates of *C. coli* from seals and their low genetic diversity (12/14 isolates were ST-827), these isolates were not included in the comparative analysis. There was a median of 1,662 detected genes per seal isolate genome, slightly fewer than the median of genes detected on the

genomes present in the NCBI repository as of March 2014 (1,760 genes; n=89), but comparable to the number of genes of the reference strain NCTC11168 (1,670 genes). No single gene was present or absent in all seal isolates compared to the background agricultural *C. jejuni* population. However, some genes showed from around 40% to 50% difference in prevalence between seal and non-seal isolates (**Table 1**). Interestingly, these include 4 genes of a cluster that are involved in ferric enterobactin acquisition in *Campylobacter* (Zeng *et al.* 2013), which were detected in 40% less isolates from seals than from other sources. Additionally, a CRISPR-associated protein initially found in the *C. coli* strain 76339 annotation (BN865_15240c) was detected in almost half as many seal isolates than isolates from other sources (55.6% vs. 81.8%, respectively; **Table 1**). Some genes were also found to be more prevalent in seal isolates than in other sources, including a gene putatively encoding for an adhesin (Cj_81-176_3910; **Table 1**). Although some genes were found to differ in prevalence, and could possibly indicate slight variations in host association, no obvious genotype was linked to seal isolates. This is consistent with the recent spread of isolates common in agricultural/clinical sources.

The number of unique alleles per isolate was determined for different host groups for 704 loci shared by all *C. jejuni* isolates (**Figure S1**). Allelic diversity at core loci varied in isolates from different host groups. Isolates from wild birds had the highest core genome allelic diversity, with an average of 0.70 unique alleles per locus. Clinical isolates were also diverse, with around 0.46 unique alleles per locus. Chicken and cattle isolates showed a reduced allelic diversity with 0.31 and 0.30 unique alleles per locus respectively. The allelic diversity in seal isolates (0.47) was comparable with that in clinical isolates.

Molecular variation in source populations

Source attribution of seal isolates was estimated based on the frequency of MLST alleles in different source populations. Source populations comprised published MLST datasets from chicken meat and faeces (n=1298); cattle faeces (n=597); sheep faeces (n=250); wild bird faeces (n=247) for which isolates were compiled representing multiple sample locations (Sheppard *et al.* 2010; Sheppard *et al.* 2011). These data can be used for attribution of seal isolates from a discrete geographic location as host association is robust to variations in geographic structuring (Sheppard *et al.* 2010) and allele frequencies are considered not to differ sufficiently between studies or locations to erase this host signal. In order to test the differentiation between and within groups, analysis of molecular variance (AMOVA) was

used (Excoffier & Lischer 2010). Gene frequencies between and within groups were assessed using F-statistics (**Table 2**) with F_{ST} representing the genetic differentiation within each group and F_{GT}, the genetic differentiation between the groups.

Significant heterogeneity was observed between sub-groupings of the same population. These ranged from an F_{ST} of 1.47 % in chickens to 10.6 % in wild birds. The increased heterogeneity within the wild bird samples is likely to reflect the diversity of wild bird hosts. In order to perform source attribution effectively the differences must be larger between groups than within them in order to assign an isolate to a population. This was estimated using a nested AMOVA between paired populations (**Table 2**). All paired populations showed significant differentiation from one another to a greater or lesser degree. The F_{GT} values ranged from 1.01 % between cattle and sheep to 18.44 % between sheep and wild birds. There are two paired groupings that showed low, but significant, differentiation. Cattle and sheep exhibited the lowest F_{GT} value, indicating very low genetic differentiation. Additionally, humans and chickens showed low differentiation.

Attribution model validation

Empirical cross validation of the attribution model was performed with the source population datasets. A total of 100 simulated datasets were generated where the source information of half of the isolates were randomly assigned as test isolates (**Table 3**). From 100 simulations 74% of test isolates were correctly assigned to their original source populations. The model was considered to be well calibrated as it only slightly overestimated the true proportion (76% predicted to be correctly attributed). For chicken, sheep and wild birds the model slightly underestimated the number of test cases that were attributable to these source populations. A slight overestimation was exhibited for cattle. In terms of model coverage, the number of simulations out of 100 in which the 95% credible interval incorporated the true value, both sheep and wild birds exhibited 95 or above. Chicken and cattle fell below this value, at 89 and 84 respectively, indicating that perhaps the slight bias led to an incorrect estimation of the proportion within these populations.

Source attribution of seal isolates

Statistical models, implemented in iSource (Wilson *et al.* 2008), were used to attribute the source of seal *C. jejuni* and *C. coli* isolates using two source datasets, first with chicken, cattle, sheep and wild birds isolates, and then with the addition human clinical isolates to this

dataset (**Figure 4**, **Table 4**). Human clinical isolates are themselves a population from mixed sources, as persistent carriage of *C. jejuni* is not thought to occur in humans. However, in the case of environmental contamination, human isolates can represent a source population with more delimited agricultural and non-agricultural sources.

The majority of seal isolates were attributed to a human source 78.8% (C.I. 53.5%-93.7%) (**Table 4**). Interestingly, when human clinical isolates were removed from the sample dataset, the majority of seals isolates were attributed to the chicken (63.3%; C.I. 36.3%-87.6%) (**Table 4**). As both of these important reservoirs are linked (Sheppard *et al.* 2009a; Sheppard *et al.* 2009b; Wilson *et al.* 2008), this result suggests that the majority of seal isolates are most similar to *C. jejuni* isolates associated with human and poultry sources. It is interesting to note that the attribution of seal isolates to the human source was more robust than to the chicken source (**Figure 4, Table 4**).

When human isolates were excluded from the source dataset, all seal isolates from the ST-45 (n=34), ST-21 (n=22) and ST-22 (n=14) clonal complexes were attributed to the chicken source population (**Table S3**). Three seal isolates, from ST-696 (ST-1332 clonal complex), ST-1256 and ST-1457, were attributed to the wild bird source population. When human isolates were included in the source dataset, almost all clonal complexes were attributed to humans, except ST-1256. Within the ST-45 clonal complex, the ST-1003 did not have a clear attribution pattern, displaying a posterior probability of 0.41 to humans and 0.425 to wild birds (**Table S4**).

Discussion

Urbanisation and human expansion can be accompanied by chemical and biological contamination of the surrounding natural environment. This is especially true in coastal areas with high human population density, where sewage and wastewaters are discharged. Bacteria and protozoa associated with human activity have been isolated from several marine mammal species that live off the coast of urban areas (Conrad *et al.* 2005; Greig *et al.* 2007; Shapiro *et al.* 2012; Stoddard *et al.* 2005). Long lived marine mammals are important sentinel species, potentially providing evidence of the increasing anthropic biological contamination of the seas but few studies have considered the possibility that human pathogens could cause disease in these species. In this study, we examined the potential for the human pathogen *Campylobacter* to spread to marine wildlife in coastal areas. This organism has been isolated

from other seal species, such as the Northern Elephant seal (*Mirounga angustirostris*) (Stoddard *et al.* 2005), the Antarctic fur seal (*Arctocephalus gazella*) and the Weddell seal (*Leptonychotes weddellii*) (Garcia-Pena *et al.* 2010) and the present study represents the first isolation of *C. jejuni*, *C. coli* or *C. lari* in the grey seal. The lack of *C. insulaenigrae*, a putative marine mammal specific species of *Campylobacter* (Foster et al. 2004; Stoddard et al, 2007; Garcia-Pena et al. 2010; Gonzalez et al. 2011), in this study, is worthy of note. However, the high prevalence of *Campylobacter* in the grey seals (*Halichoerus grypus*) inhabiting the breeding colony of the Isle of May in the Firth of Forth (Scotland, UK) may reflect the intense human activity in this area and possible infection from human or livestock sources.

Asymptomatic carriage is usually assumed in the absence of clear evidence of pathology among animals infected with *Campylobacter*. However, experimental infections have demonstrated that *Campylobacter* can be pathogenic in terrestrial mammals including dogs, ferrets, immunodeficient mice and gnotobiotic piglets (Boosinger & Powe 1988; Hodgson *et al.* 1998; Macartney *et al.* 1988; Nemelka *et al.* 2009). While clinical signs were not identified in live healthy grey seals infected with *Campylobacter,* the correlation of *Campylobacter* in rectal swabs of dead seal pups on the colony with moderate to severe colitis and immunohistochemical demonstration of *Campylobacter* within intestinal crypts, is strongly suggestive of pathogenicity.

Initial clues about the spread of *Campylobacter* from terrestrial sources comes from the distribution of infected seals. Specifically, prolonged residence on the Isle of May was associated with a higher prevalence of *Campylobacter* in seals, whereas returning yearling animals, that left the island and travelled around the North Sea, had lower infection rates, potentially having cleared the infection. Infection levels were also lower in seal pups living on the tidal boulder beach of the island and stranded seal pups on the coasts of the Scottish mainland. This distribution of infection is consistent with a scenario of contamination by exposure, in which ecological factors associated with the Isle of May influenced *Campylobacter* infection of seals.

Genomic characterization of *C. jejuni* and *C. coli* isolates from seals revealed that they did not represent divergent seal-associated genetic clusters as has been observed for some lineages from wild birds (Sheppard *et al.* 2011; Griekspoor *et al.* 2013). In fact, isolates from

seals clustered phylogenetically with MLST clonal complexes commonly associated with agricultural and clinical sources. This included isolates belonging to the ST-45 and ST-21 clonal complexes which are among the most common lineages in seals and human disease. The allelic diversity of seal and human/agricultural isolates was also comparable. In the absence of obvious divergence between strains from seals, human clinical isolates and agricultural animals, contamination from external sources and multiple recent *Campylobacter* infections of grey seals is the most parsimonious explanation.

To obtain a quantitative estimate of the relative contribution of different *Campylobacter* reservoirs to seal infection, it is necessary to carry out source tracking methods that provide a probability that each isolate originated in one of the source populations. These methods have been used to attribute human *Campylobacter* infection to host sources, often implicating the consumption of contaminated poultry as a major cause of human infection (Sheppard *et al.* 2009a; Sheppard *et al.* 2009b). Because the host association in *C. jejuni* is robust to temporal and spatial variation in isolate sampling, and there are large numbers of available isolates that have been sequenced at 7-MLST loci, it was possible to adapt these attribution methods to identify the putative infection source of grey seals.

The Isle of May is an important habitat for several wild bird species, which are present at high density, and would be expected to be the major source of infection for seals. However, this does not appear to be the case as attribution modelling implicates chicken and cattle reservoirs as the principal sources of *C. jejuni* in grey seals. The exact route of transmission to seals is unknown but in the absence of livestock on the island, it is possible that land run-off or other farming activity around the estuary could contribute as a source of *Campylobacter.* Added to this is the possibility that human sewage could represent a source. Persistent carriage is not thought to occur in humans and therefore *Campylobacter* in clinical samples are representative of the source population, including cattle and poultry. Consistent with this, isolates infecting grey seals that are of cattle and poultry origin could have come via human sewage. The pathways of the land-sea transfer remain unclear, but the location of the Isle of May within a major shipping lane and the densely populated coastline of the Firth of Forth could be important factors. The role of wild birds as vectors, however, would warrant further investigation in case they are acting as vectors for transmission of poultry and cattle *Campylobacter* strains by moving between refuse sites on the mainland and the Isle of May. Similarly, investigation into the prevalence of *Campylobacter* in grey seal colonies

located in more remote areas may help elucidate risk factors for the presence of this pathogen.

With increasing urbanization, the pressure on national nature reserves intensifies. This study demonstrates the spread of a human pathogen to a sentinel marine mammal species inhabiting a national nature reserve, probably through faecal contamination from agricultural land or human sewage. This has strong implications for understanding how the degradation of water quality of coastal habitats can influence the spread of important human pathogens.

Acknowledgements

This work and JLB's PhD studentship were funded by the Moredun Research Institute and the Royal Zoological Society of Scotland. We wish to thank the Sea Mammal Research Unit at the University of St Andrews for assistance with field sampling (Matt Bivins, Kelly Robinson, Paula Redman, Chris McKnight and Amanda Stansbury), mapping and statistical advice (Bernie McConnell and Mike Lonergan); the Scottish SPCA National Wildlife Rescue Centre (Colin Seddon, Claire Stainfield and staff) for sample collection and coordination; Clare Underwood, Jeanie Finlayson and Val Forbes of Moredun Research Institute Pathology department for excellent histopathological and immunohistochemical preparations; Glasgow Polyomics for performing the whole genome sequencing and Scotland's Rural College, Inverness for assistance with microbiological methods. We also thank Paul Everest, currently University of Glasgow, for providing positive control tissue for *Campylobacter* IHC. All sampling of live free-ranging animals was carried out under UK Home Office Project (No. 60/4009) and Personal Licences as issued to the Sea Mammal Research Unit under the Animals (Scientific Procedures) Act, 1986. Stranded grey seal pups were sampled as part of the routine health assessment procedure. SKS is funded by the Biotechnology and Biological Sciences Research Council (BBSRC), the Medical Research Council (MR/L015080/1) and the Wellcome Trust.

References

Baker JR, Baker R (1988) Effects of Environment on Grey Seal (Halichoerus-Grypus) Pup Mortality - Studies on the Isle of May. *Journal of Zoology* **216**, 529-537.

Boosinger TR, Powe TA (1988) Campylobacter jejuni infections in gnotobiotic pigs. *American Journal of Veterinary Research* **49**, 456-458.

Bossart GD (2011) Marine mammals as sentinel species for oceans and human health. *Veterinary Pathology* **48**, 676-690.

Boysen L, Nauta M, Duarte AS, Rosenquist H (2013) Human risk from thermotolerant Campylobacter on broiler meat in Denmark. *Int J Food Microbiol* **162**, 129-134.

Conrad PA, Miller MA, Kreuder C*, et al.* (2005) Transmission of Toxoplasma: Clues from the study of sea otters as sentinels of Toxoplasma gondii flow into the marine environment. *Int J Parasitol* **35**, 1155-1168.

de Haan CP, Kivisto RI, Hakkinen M, Corander J, Hanninen ML (2010) Multilocus sequence types of Finnish bovine Campylobacter jejuni isolates and their attribution to human infections. *BMC Microbiol* **10**, 200.

Dingle KE, Colles FM, Falush D, Maiden MC (2005) Sequence typing and comparison of population biology of Campylobacter coli and Campylobacter jejuni. *J Clin Microbiol* **43**, 340-347.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797.

Excoffier L, Lischer HE (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* **10**, 564-567.

Garcia-Pena FJ, Perez-Boto D, Jimenez C*, et al.* (2010) Isolation and characterization of Campylobacter spp. from Antarctic fur seals (Arctocephalus gazella) at Deception Island, Antarctica. *Applied and Environmental Microbiology* **76**, 6013-6016.

Greig TW, Bemiss JA, Lyon BR, Bossart GD, Fair PA (2007) Prevalence and Diversity of Antibiotic Resistant Escherichia coli in Bottlenose Dolphins (Tursiops truncatus) from the Indian River Lagoon, Florida, and Charleston Harbor Area, South Carolina. *Aquatic Mammals* **33**, 185-194.

Gripp E, Hlahla D, Didelot X*, et al.* (2011) Closely related Campylobacter jejuni strains from different sources reveal a generalist rather than a specialist lifestyle. *Bmc Genomics* **12**, 584.

Haddock G, Mullin M, MacCallum A*, et al.* (2010) Campylobacter jejuni 81-176 forms distinct microcolonies on in vitro-infected human small intestinal tissue prior to biofilm formation. *Microbiology* **156**, 3079-3084.

Hodgson AE, McBride BW, Hudson MJ, Hall G, Leach SA (1998) Experimental campylobacter infection and diarrhoea in immunodeficient mice. *Journal of Medical Microbiology* **47**, 799-809.

Jolley KA, Maiden MC (2010) BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics* **11**, 595.

Kimura M, Weiss GH (1964) The Stepping Stone Model of Population Structure and the Decrease of Genetic Correlation with Distance. *Genetics* **49**, 561-576.

Kittl S, Heckel G, Korczak BM, Kuhnert P (2013) Source attribution of human Campylobacter isolates by MLST and fla-typing and association of genotypes with quinolone resistance. *PLoS One* **8**, e81796.

Macartney L, Al-Mashat RR, Taylor DJ, McCandlish IA (1988) Experimental infection of dogs with Campylobacter jejuni. *Veterinary Record* **122**, 245-249.

Meric G, Yahara K, Mageiros L*, et al.* (2014) A reference pan-genome approach to comparative bacterial genomics: identification of novel epidemiological markers in pathogenic Campylobacter. *PLoS One* **9**, e92798.

Miller WG, Englen MD, Kathariou S*, et al.* (2006) Identification of host-associated alleles by multilocus sequence typing of Campylobacter coli strains from food animals. *Microbiology* **152**, 245-255.

Miller WG, On SL, Wang G*, et al.* (2005) Extended multilocus sequence typing system for Campylobacter coli, C. lari, C. upsaliensis, and C. helveticus. *J Clin Microbiol* **43**, 2315-2329.

Moore SE (2008) Marine mammals as ecosystem sentinels. *Journal of Mammalogy* **89**, 534-540.

Mullner P, Spencer SE, Wilson DJ*, et al.* (2009) Assigning the source of human campylobacteriosis in New Zealand: a comparative genetic and epidemiological approach. *Infection Genetics and Evolution* **9**, 1311-1319.

Nemelka KW, Brown AW, Wallace SM*, et al.* (2009) Immune response to and histopathology of Campylobacter jejuni infection in ferrets (Mustela putorius furo). *Comparative Medicine* **59**, 363-371.

On SL (1996) Identification methods for campylobacters, helicobacters, and related organisms. *Clin Microbiol Rev* **9**, 405-422.

On SL, Holmes B (1991) Reproducibility of tolerance tests that are useful in the identification of campylobacteria. *J Clin Microbiol* **29**, 1785-1788.

On SL, Holmes B (1992) Assessment of enzyme detection tests useful in identification of campylobacteria. *J Clin Microbiol* **30**, 746-749.

Price MN, Dehal PS, Arkin AP (2010) FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* **5**, e9490.

Reddy ML, Dierauf LA, Gulland FMD (2001) Marine Mammals as Sentinels of Ocean Health. In: *Marine Mammal Medicine* (eds. Dierauf LA, Gulland FMD), pp. 3-14. CRC Press LLC, Florida.

Shapiro K, Miller M, Mazet J (2012) Temporal association between land-based runoff events and california sea otter (enhydra lutris nereis) protozoal mortalities. *J.Wildl.Dis.* **48**, 394-404.

Sheppard SK, Cheng L, Meric G*, et al.* (2014) Cryptic ecology among host generalist Campylobacter jejuni in domestic animals. *Mol Ecol* **23**, 2442-2451.

Sheppard SK, Colles F, Richardson J*, et al.* (2010) Host association of Campylobacter genotypes transcends geographic variation. *Appl Environ Microbiol* **76**, 5269-5277.

Sheppard SK, Colles FM, McCarthy ND*, et al.* (2011) Niche segregation and genetic structure of Campylobacter jejuni populations from wild and agricultural host species. *Mol Ecol* **20**, 3484-3490.

Sheppard SK, Dallas JF, MacRae M*, et al.* (2009a) Campylobacter genotypes from food animals, environmental sources and clinical disease in Scotland 2005/6. *Int J Food Microbiol* **134**, 96-103.

Sheppard SK, Dallas JF, Strachan NJ*, et al.* (2009b) Campylobacter genotyping to determine the source of human infection. *Clinical Infectious Diseases* **48**, 1072-1078.

Sheppard SK, Didelot X, Jolley KA*, et al.* (2013a) Progressive genome-wide introgression in agricultural Campylobacter coli. *Mol Ecol* **22**, 1051-1064.

Sheppard SK, Didelot X, Meric G*, et al.* (2013b) Genome-wide association study identifies vitamin B5 biosynthesis as a host specificity factor in Campylobacter. *Proc Natl Acad Sci U S A* **110**, 11923-11927.

Sheppard SK, McCarthy ND, Falush D, Maiden MC (2008) Convergence of Campylobacter species: implications for bacterial evolution. *Science* **320**, 237-239.

Stoddard RA, Gulland MD, Atwill ER*, et al.* (2005) Salmonella and Campylobacter spp. in northern elephant seals, California. *Emerging Infectious Diseases* **11**, 1967-1969.

Strachan NJ, Forbes KJ (2010) The growing UK epidemic of human campylobacteriosis. *Lancet* **376**, 665-667.

Strachan NJ, Gormley FJ, Rotariu O*, et al.* (2009) Attribution of Campylobacter infections in northeast Scotland to specific sources by use of multilocus sequence typing. *Journal of Infectious Diseases* **199**, 1205-1208.

Tam CC, Rodrigues LC, Viviani L, *et al.* (2012) Longitudinal study of infectious intestinal disease in the UK (IID2 study): incidence in the community and presenting to general practice. *Gut* **61**, 69-77.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* **30**, 2725-2729.

Wilson DJ, Gabriel E, Leatherbarrow AJ, *et al.* (2008) Tracing the source of campylobacteriosis. *Plos Genetics* **4**, e1000203.

Zeng X, Xu F, Lin J (2013) Specific TonB-ExbB-ExbD energy transduction systems required for ferric enterobactin acquisition in Campylobacter. *FEMS Microbiol Lett* **347**, 83-91.

Data Accessibility

The 88 assembled genome sequences generated in this study, core genome alignment, additional histopathology and immunohistochemistry pictures, and seal data information were deposited on the Dryad repository (doi:10.5061/dryad.8p984). The short reads have been uploaded to the NCBI Short Read Archive (SRA) and are available on: http://www.ncbi.nlm.nih.gov/bioproject/PRJNA264388.

Figures and tables legends

Figure 1. Prevalence of *Campylobacter* in seals sampled from the Isle of May, Scotland. Map of the Isle of May grey seal breeding colony with sampling sites of *Campylobacter* positive (red) and negative (blue) dead seals indicated. The three different sampling sites of live pups on the island, and their description, are shown with prevalence of positive (red) and negative (blue) pups represented

Figure 2. **Intestinal histopathological and immunohistochemical examination of samples from dead grey seals and association of disease severity with *Campylobacter.*** (A) Representative colon histology (H&E; haematoxylin and eosin stain) and immunohistochemical (IHC; counterstained with Mayer's haematoxylin) visualisations of large intestinal crypts of dead free-ranging grey seal pups. *Campylobacter* organisms are located on the apical surface of enterocytes and within the intestinal crypt lumen (arrows). (B) Association of isolate culture and detection from rectal swabs from wild dead grey seals with the severity of inflammation as observed by histopathological observations.

Figure 3. Population structure of seal-associated *C. jejuni* and *C. coli*. Phylogenetic tree of 74 *C. jejuni* (A) and 14 C. *coli* (B) isolates of grey seal origin along with 131 *C. jejuni* and 61 *C. coli* previously published genome sequences, from agricultural and non-agricultural sources. The tree is based on concatenated sequences of 595 core genes established using an approximation of the maximum likelihood algorithm using FastTree2 and visualised using MEGA6. Pink circles represent isolates from grey seals; yellow circles represent isolates from chicken; orange circles represent isolates from human clinical cases; blue circles represent isolates from cattle; grey represent isolates from sheep and white circles represent isolates from wild bird and riparian sources. Numeric labels correspond to sequence types of seal isolates, numeric labels in bold correspond to clonal complexes of seal isolates.

Figure 4. **Source attribution of *C. jejuni* from grey seals.** A model of source attribution implemented in iSource (Wilson *et al.*, 2008) was used to infer probability of sources from a representative isolate collection of MLST profiles from various possible host sources. (A, B) The origin of seal isolates was inferred from pools of chicken, cattle, sheep and wild birds isolates. (C, D) The origin of seal isolates was inferred using the same pools, with the addition of human clinical isolates, used as a possible source.

Figure S1. Core genome allelic diversity in *C. jejuni* isolates from grey seals and other hosts. Core genome allelic diversity was calculated as the number of unique alleles found at each of the 703 core loci in all isolates from a particular host group divided by the number of isolates from the corresponding host group. The starting position of the 703 core loci was mapped onto the *C. jejuni* strain NCTC11168 annotation. The plots represent moving averages of allelic diversity values with a period of 20bp for each host group. The colour of the lines represent the different host groups, with grey: wild birds, yellow: humans, red: chicken, blue: cattle and pink: grey seals. The asterisk indicates a genomic region of lower diversity in each host group, corresponding to housekeeping ribosomal genes showing a lower allelic diversity across the whole species.

Table 1. Highest differences in gene prevalence between isolates from seals and from agricultural, human clinical isolates and wild birds hosts.

Table 2. Genetic variation of *C. jejuni* within groups (F_{ST}) and between groups (F_{GT}). Significant F-statistics are printed in bold. A Bonferroni correction was applied to control the family-wise error rate.

Table 3. Performance of the source attribution model during empirical cross-validation. Isolates were assigned to a source population *a posteriori* based upon their most likely assignment probability (proportion of isolates correctly assigned). Coverage was the proportion of simulations where the true proportion fell within the 95% credible interval. Bias and RMSE were calculated for the sub-populations that were assigned pseudo-human during empirical cross-validation.

Table 4. Summary of the posterior distribution (F) after source attribution for seal source populations against source datasets with and without human clinical isolates.

Table S1. List of *Campylobacter* genomes used in this study.

Table S2. Categorical risk factors of *Campylobacter* carriage in grey seals. These were calculated using univariate analysis, for all grey seals that are harbouring *Campylobacter* spp. (n=: group size; OR: odds ratio; 95% CI: 95% confidence interval; Sign: Statistical significance of results; NS: non-significant; *: $p<0.05$; **: $p<0.01$; ***: $p<0.001$)

Table S3. Summary of the sequence types of the 74 *C. jejuni* seal isolates used in the current study and their posterior probability after source attribution using chicken, cattle, sheep and wild bird isolates as sources.

Table S4. Summary of the sequence types of the 74 *C. jejuni* seal isolates used in the current study and their posterior probability after source attribution using human clinical, chicken, cattle, sheep and wild bird isolates as sources.

Table 1. **Highest differences in gene prevalence between isolates from seals and from agricultural, clinical and wild birds hosts**.

Gene name[a]	Description of predicted function	Gene prevalence (%)					Difference of prevalence between seals and non-seals isolates
		seals (n=17)[b]	cattle (n=33)	chicken (n=54)	human (n=27)	wild birds (n=11)	
11168_Cj1677	lipoprotein	1 (5.6%)	16 (48.5%)	29 (53.7%)	18 (66.7%)	5 (45.5%)	-47.1%
11168_Cj0628	lipoprotein	2 (11.1%)	16 (48.5%)	30 (55.6%)	18 (66.7%)	5 (45.5%)	-42.3%
11168_Cj1365c	secreted serine protease	5 (27.8%)	26 (78.8%)	34 (63.0%)	22 (81.5%)	2 (18.2%)	-39.4%
Cc76339__15240c	CRISPR-associated protein, Csn1 family	9 (55.6%)	31 (93.9%)	52 (96.3%)	27 (100%)	9 (81.8%)	-39.1%
11168_Cj0177	iron transport protein	4 (22.2%)	18 (54.5%)	34 (63.0%)	21 (77.8%)	2 (18.2%)	-38.1%
11168_exbB1	biopolymer transport protein	4 (22.2%)	18 (54.5%)	34 (63.0%)	21 (77.8%)	2 (18.2%)	-38.1%
11168_exbD1	biopolymer transport protein	4 (22.2%)	18 (54.5%)	34 (63.0%)	21 (77.8%)	2 (18.2%)	-38.1%
11168_tonB1	TonB transport protein	4 (22.2%)	18 (54.5%)	34 (63.0%)	21 (77.8%)	2 (18.2%)	-38.1%
11168_Cj1305c	hypothetical protein	17 (100%)	22 (66.7%)	37 (68.5%)	8 (29.6%)	9 (81.8%)	38.9%
Cj_81-176_8305	dimethyl sulfoxide reductase subunit A	12 (72.2%)	13 (39.4%)	16 (29.6%)	5 (18.5%)	8 (72.7%)	37.9%
Cj_81-176_3910	adhesin	10 (61.1%)	13 (39.4%)	10 (18.5%)	5 (18.5%)	1 (9.1%)	37.4%
Cj_81-176_4945	hypothetical protein	12 (72.2%)	15 (45.5%)	14 (25.9%)	7 (25.9%)	8 (72.7%)	37.1%

a. The gene name correspond to the genes from the reference pan-genome list and specifies the annotation of origin in the first part of the name.

b. The number of unique clones based on the population structure observed on a phylogenetic tree was chosen to calculate prevalence in isolates from seals.

Table 2. Genetic variation of *C. jejuni* within groups (F_{ST}) and between groups (F_{GT}). Significant F-statistics are printed in bold. A Bonferroni correction was applied to control the family-wise error rate.

Genetic differentiation within groups (F_{ST})						
	Chicken	**Cattle**	**Sheep**	**Wild Bird**	**Human**	**Seal**
F_{ST}	**0.015**	**0.024**	**0.036**	**0.106**	-	-
p	0.00	0.00	0.00	0.00		
Genetic variation between groups (F_{GT}/*p*)						
Chicken	-	0.000	0.000	0.000	0.000	0.000
Cattle	**0.078**	-	0.002	0.000	0.000	0.000
Sheep	**0.080**	**0.010**	-	0.000	0.000	0.000
Wild Bird	**0.124**	**0.168**	**0.184**	-	0.000	0.000
Human	**0.015**	**0.056**	**0.048**	**0.161**	-	0.000
Seal	**0.091**	**0.153**	**0.159**	**0.133**	**0.142**	-

Table 4. Summary of the posterior distribution (F) after source attribution for seal source populations against source datasets with and without human clinical isolates.

	Proportion of isolates attributed to source population (F)				
	Human	**Chicken**	**Cattle**	**Sheep**	**Wild bird**
Excluding human clinical isolates					
Mean	-	0.633	0.050	0.183	0.135
Median	-	0.635	0.035	0.169	0.122
Standard deviation	-	0.133	0.050	0.118	0.083
2.50% quantile	-	0.363	0.001	0.010	0.016
97.50% quantile	-	0.876	0.183	0.444	0.331
Including human clinical isolates					
Mean	0.788	0.089	0.026	0.036	0.060
Median	0.806	0.063	0.0190	0.025	0.051
Standard deviation	0.104	0.086	0.026	0.037	0.041
2.50% quantile	0.535	0.002	0.001	0.001	0.007
97.50% quantile	0.937	0.322	0.095	0.135	0.164

Table 4. Summary of the posterior distribution (F) after source attribution for seal source populations against source datasets with and without human clinical isolates.

	Proportion of isolates attributed to source population (F)				
	Human	**Chicken**	**Cattle**	**Sheep**	**Wild bird**
Excluding human clinical isolates					
Mean	-	0.633	0.050	0.183	0.135
Median	-	0.635	0.035	0.169	0.122
Standard deviation	-	0.133	0.050	0.118	0.083
2.50% quantile	-	0.363	0.001	0.010	0.016
97.50% quantile	-	0.876	0.183	0.444	0.331
Including human clinical isolates					
Mean	0.788	0.089	0.026	0.036	0.060
Median	0.806	0.063	0.0190	0.025	0.051
Standard deviation	0.104	0.086	0.026	0.037	0.041
2.50% quantile	0.535	0.002	0.001	0.001	0.007
97.50% quantile	0.937	0.322	0.095	0.135	0.164