Chapter from the book *Current Topics in Phylogenetics and Phylogeography of Terrestrial and Aquatic Systems*


Interested in publishing with IntechOpen?
Contact us at book.department@intechopen.com
Ecological Factors that Influence Genetic Structure in Campylobacter coli and Campylobacter jejuni

Helen M. L. Wimalarathna¹ and Samuel K. Sheppard¹,²

¹The University of Oxford, Department of Zoology,
²The University of Swansea, College of Medicine
United Kingdom

1. Introduction

Campylobacter is the leading cause of human bacterial gastroenteritis worldwide (Friedman et al. 2000). Campylobacteriosis, caused principally by the organisms C. jejuni and C. coli, is characterized by severe diarrhoea, usually accompanied by fever, abdominal pain, nausea and malaise (Allos 2001). Campylobacter infection accounts for an estimated 2.5 million cases of gastro-intestinal disease in the United States and 1.3 million cases in the United Kingdom each year (Kessel et al. 2001), and the estimated economic burden of Campylobacter infection is $8 billion in the US and £500 million in the UK. Though pathogenic in humans, these Campylobacter species are wide-spread commensals in the digestive tracts of many wild and domesticated animals. Because of its public health significance, considerable effort has gone into understanding how this common organism is transmitted from these reservoir hosts to humans through contaminated meat, poultry, water, milk and contact with animals (Niemann et al. 2003).

1.1 Molecular typing

Molecular typing of pathogenic bacteria has enhanced many epidemiological studies, including the identification of food-borne outbreaks of infection due to E. coli O157: H7 (Bender et al. 1997), Salmonella enteritica (Bender et al. 2001) and Listeria monocytogenes (Olsen et al. 2005) and early identification of an outbreak source can enable effective disease containment (Olsen et al. 2005) (Rangel et al. 2005). However, in many species of bacteria it is impossible to predict the lineage from which a pathogenic phenotype will arise. For example, in Bacillus cereus, pathogenicity is associated with mobile elements which mediate spore-formation and toxin production (Raymond et al. 2010). These elements can be acquired by distantly related lineages and it is, therefore, difficult to predict the likelihood that a strain will be pathogenic from genotypes derived from non-plasmid DNA alone.

In species such as Campylobacter, particular genetically related groups often display similar disease associated phenotypes. The consistency of Campylobacter genotypes within sub-populations, and the variation between sub-populations can be exploited in order to
determine the source of human infection by comparing clinical isolate genotypes data with large reference sets isolated from known host-species.

Methods such as PFGE and serotyping have shown that far from being monomorphic pathogenic clones like *Yersinia pestis* or *Mycobacterium lepris* (Achtman 2008), *Campylobacter* populations are highly structured with complex associations among lineages at different levels of relatedness. The DNA sequence-based typing method of Multi Locus Sequence Typing (MLST) has provided considerable insight into population structure in recombining organisms such as *Campylobacter*. MLST is an unambiguous high-resolution genotyping method, exploiting genetic variation in fragments of seven separate housekeeping genes. Each locus is approximately 500bp in length, with a defined start and end point. Each unique sequence at a given locus is assigned an allele number, and a Sequence Type (ST) is identified by a unique series of seven numbers, referring to the specific alleles present at each locus. Related STs can be grouped into Clonal complexes, in which sequences are identical at four or more loci (Maiden et al. 1998).

This high degree of genetic structuring is the result of a complex interplay of mutation, which leads to the gradual divergence of clonally related lineages, and horizontal gene transfer (HGT), that can lead to the replacement of homologous DNA with sequence from another lineage or in extreme cases the introduction of new genes. While *Campylobacter* can be highly recombinogenic (Wilson et al. 2009), mutation and HGT have not been sufficient to erase the clonal signal of descent from the genomes and in the following sections we will investigate some of the ways in which this high degree of genetic structuring can tell us about the biology of this organism and how this relates to disease.

2. Disease-associated lineages

Analysis of the genotypes of *Campylobacter* isolated from human disease cases has shown that the vast majority of campylobacteriosis cases are caused by *C. jejuni* and *C. coli* lineages also found in other potential disease reservoirs, particularly chickens and cattle (Figure 1). Most human disease strains also occur as commensal organisms in domesticated animals, and clinical isolates are a non-random subset of these lineages. This is particularly marked in the case of *C. coli*, in which the average diversity per locus is 13 alleles in disease cases compared with 55 alleles in the general *C. coli* population.

The apparent absence of asymptomatic carriage of *C. jejuni* and *C. coli* among individuals in the UK suggests that humans may not be a natural host for these organisms in high income countries, and have undergone a relatively short history of co-evolution. For this reason, an appreciation of phylogenetic relationships, together with an examination of the ecology of *Campylobacter* can enhance understanding of the origin and causes of human campylobacteriosis and this forms the basis for much of the recent work to explain the epidemiology of these organisms (de Haan et al. 2010; Hastings et al. 2011; Jorgensen et al. 2011). (Kittl et al. 2011; Lang et al. 2010; Magnusson et al. 2011; Mullner et al. 2010; Sheppard et al. 2011a; Sproston et al. 2011; Sproston et al. 2010; Thakur & Gebreyes 2010).

3. Contrasting population structure of *Campylobacter jejuni* and *Campylobacter coli*

A neighbour joining tree (Figure 2) shows that *C. jejuni* and *C. coli* display markedly different population structures. *C. jejuni* populations are highly structured into clonal
complexes (Figure 3), clusters of related lineages that share alleles at four or more MLST loci. When each locus is considered separately, there is evidence of considerable recombination within *C. jejuni*, with alleles from disparate locations around the tree appearing within the same STs. In contrast, *C. coli* displays far greater genetic diversity, with three deep-branching clades, of which clade 1 contains the vast majority of lineages described to date. The ST-828 clonal complex, part of clade 1, accounted for around 70% of the 2289 *C. coli* isolates submitted to the PubMLST database (http://pubmlst.org/campylobacter) before September 5th 2011, with most of the remainder sharing alleles with these, and therefore also being related. The second most common clonal complex (ST-1150 complex), also from clade 1, accounts for only 2% of isolates.

Fig. 1. Human clinical isolates and those from chicken and cattle (faeces and meat). The relative abundance of clonal complexes (responsible for >1% of total UK disease) of isolates from human *C. jejuni* and *C. coli* infections and those from published chicken and cattle isolate collections (Sheppard et al. 2010a; Sheppard et al. 2009b; Sheppard et al. 2010b). All of the 21 most common disease causing clonal complexes are also found in cattle, chickens or both.
Fig. 2. The genetic relatedness of 1341 C. jejuni and C. coli genotypes based on concatenated MLST alleles (3309bp) from published studies (Sheppard et al. 2010a; Sheppard et al. 2010b). Contrasting tree topologies are visible on the neighbour-joining trees with three deep branching clades present among C. coli genotypes.
Ecological Factors that Influence Genetic Structure in Campylobacter coli and Campylobacter jejuni

3.1 Trefoil structure in Campylobacter coli

The emergence and maintenance of the 3-clade structure in C. coli implies that three distinct bacterial gene pools exist, and that although recombination is evident within C. coli clades, there are or have been barriers to recombination between clades. Recombinational barriers can be considered in three broad categories: (i) adaptive, implying selection against hybrids; (ii) mechanistic, imposed by homology dependence of recombination or other factors promoting DNA specificity; (iii) ecological, a consequence of physical separation in different ecological niches (Sheppard et al. 2010b). In order to consider the relative importance of each type of barrier in the evolution of a clade structure in C. coli it may be useful to look at both C. coli and C. jejuni in context.

Campylobacter jejuni and C. coli are approximately 12% divergent at the nucleotide level and are considered distinct microbial species, however there is strong evidence for a degree of hybridisation between the species through a process of horizontal gene transfer (HGT) (Sheppard et al. 2008) (Sheppard et al. 2011b). Statistical model-based approaches have been used to investigate the sharing of both whole alleles and recombined elements or ‘mosaic alleles’ between C. jejuni and C. coli. While C. coli clade 1 remains distinct from clades 2 & 3, there is evidence of gene flow between C. jejuni and C. coli clade 1. Analysis of 1738 alleles from a total of 2953 Sequence Types identified 31 mosaic alleles, of which 25 had demonstrably been acquired by C. coli from C. jejuni, and the remaining 6 had originated in C. coli and been acquired by C. jejuni. With the exception of a single mosaic allele, having originated in C. coli clade 3 and being acquired by C. jejuni, all genetic exchange events identified involved C. coli clade 1 as either donor or recipient.
The existence of hybrids and the maintenance of alleles of *C. jejuni* origin within the *C. coli* gene pool demonstrates that mechanistic barriers are not preventing interspecies gene flow. Furthermore, the resultant hybrid lineages are not sufficiently maladapted to prevent their proliferation (adaptive barrier). Ecological barriers to recombination are therefore likely to have been important in generating and maintaining the observed population structure in *C. coli* and *C. jejuni* species, clades and clonal complexes.

4. Ecology and host association

There is evidence of association between clusters of related genotypes and the source or host from which the bacteria were isolated. At the species level, *C. jejuni* and *C. coli* have subtly different host ranges. Both species are found in a wide range of wild and farm animals but *C. jejuni* dominate numerically in most sampled wild bird species (Colles *et al.* 2008a; Sheppard *et al.* 2011a) as well as chickens and cattle (Sheppard *et al.* 2009a). *C. coli* (clade 1) are also common in chicken and cattle, usually constituting around 10% of the *Campylobacter* population in these host animals (90% *C. jejuni*); but are more abundant than *C. jejuni* in pigs (Miller *et al.* 2006). Within *C. coli*, isolates belonging to clades 2 & 3 are far less common and are usually isolated from environmental sources where they may be associated with waterfowl.

The host-genotype relationship goes further than this. In *C. coli* and *C. jejuni* there is a strong association between specific clonal complexes (mainly *C. jejuni*), STs, and alleles and host species (McCarthy *et al.* 2007; Miller *et al.* 2006; Sheppard *et al.* 2010a). This association is stronger than spatial or temporal signals and statistical assignment analyses consistently correctly grouped isolates from a range of host animal sources regardless of geographical source (Sheppard *et al.* 2010a). For example, a population of *C. jejuni* isolates from UK chickens is strikingly similar to a population of *C. jejuni* isolates from chickens in the US, mainland Europe or Senegal. The equivalent is true of cattle, pigs and turkeys (Sheppard *et al.* 2010a). This host allelic signature between diverse lineages inhabiting the same ecological niche creates a pool of alleles common to a given source (McCarthy *et al.* 2007) and this signal of host association has been widely used to assign the origin reservoir of clinical isolates (Mullner *et al.* 2009; Sheppard *et al.* 2009b; Strachan *et al.* 2009; Wilson *et al.* 2008). All of these studies identify farm (especially chicken) associated isolates as the main source of human infection.

5. Why do farm associated isolates cause disease?

There are two possible explanations for the strong correlation between genotypes that cause human disease, and those that are associated with farm animals, especially chickens and ruminants. First, this could be the result of differential exposure. By definition, humans are more frequently exposed to domesticated food animals than to wild reservoirs of infection. The main risk factors for human campylobacteriosis include handling and consumption of raw or under-cooked poultry (Kapperud *et al.* 1992) (Friedman *et al.* 2004); handling and consumption of barbequed meat (Studahl & Andersson 2000); contact with farm animals (Friedman *et al.* 2004) and consumption of unpasteurised milk (Niemann *et al.* 2003). These risk factors are all common behaviours which present opportunities for exposure to
domesticated animals and animal products, whilst exposure to wild and environmental sources of *Campylobacter* may be less common. It is therefore possible that all *Campylobacter* strains are equally infective and the dominance of farm associated genotypes in human disease is simply reflective of greater exposure to these strains.

Alternatively, it is possible that certain strains are more likely to cause acute infection than others. If it were the case that agricultural strains were more pathogenic to humans then they would be over represented in surveys of reported clinical cases. While this may be a less likely explanation than simple differential exposure, some genotypes do appear particularly well adapted to very specific ecological niches and in an evolutionary trade-off their ability to colonise diverse hosts may have been lost. There are numerous examples of host restricted STs among strains found only in specific wild bird species (Waldenstrom et al. 2007) (Colles et al. 2008b). Genetic isolation could explain this but different colonization capacity could also be important. For example, *C. jejuni* strains (ST 3704) that are routinely found in the gut of bank voles are unable to colonise the chicken gut in laboratory experiments (Williams et al. 2010). In a similar experiment, using a European Robin (*Erithacus rubecula*) infection model, *C. jejuni* from song thrushes (*Turdus philomelos*) successfully colonized but *C. jejuni* from human disease did not (Waldenstrom et al. 2010).

As already mentioned, *C. jejuni* and *C. coli* have different host ranges and there is evidence that they exhibit colonisation and virulence factors differentially in response to different growth conditions, which may relate to host preferences (Leach et al. 1997). For example, there are a wide range of carbon sources that *C. jejuni* utilize more effectively at 42°C rather than the lower temperature of 37°C (Line et al. 2010). The average core temperature of a chicken is 42°C, while a pig is 39°C so this could influence the ability of *C. jejuni* to colonize different hosts. Serine dehydratase, encoded by the *sdaA* gene has been demonstrated to be an essential colonisation factor in *C. jejuni*. This gene is also expressed in *C. coli*, but the functionality of the enzyme is highly dependent on temperature. In *C. coli* there is little or no serine dehydratase activity at 42°C, but at the lower temperature of 37°C activity is significantly increased, this could provide a partial explanation for the porcine host association with *C. coli*.

Colonisation and virulence factors in *Campylobacter* are not well understood, but evidence of differential abilities to invade the cells of different hosts points to a possible explanation for the relationship between specific STs and human disease. Explanations based on differential exposure and colonization capacity are not mutually exclusive. It is plausible that those lineages that are found in a niche to which humans are routinely exposed have acquired the necessary colonisation factors to persist in this environment, and opportunistically to infect humans.

### 6. Dating lineage divergence

Genotyping isolates from various sources can offer insight into the causes of the genetic structuring in *Campylobacter* populations. However, a more comprehensive understanding of the evolution of the genus can be obtained if the time scale for the divergence of lineages can be overlaid upon the tree of genetic relatedness. By cross-referencing estimated dates of divergence within the genus *Campylobacter* with ecological data it is possible to make
inferences about the conditions which created the specific barriers which led to speciation, the formation of the lineage structure, and the gene-flow between certain clades and clonal complexes.

The traditional method for dating bacterial evolution is based on the rate of sequence divergence between *Escherichia coli* and *Salmonella typhimurium*, which is assumed to be 1% 16S rRNA divergence per 50 million years (Ochman & Wilson 1987). Applying this method to the *Campylobacter* genus estimates the *C. coli* – *C. jejuni* split to have occurred approximately 10 million years ago, and the divergence of 3 *C. coli* clades about 2.5 million years ago. An alternative dating method, using a molecular clock based on intra-specific diversity in *C. jejuni*, places these splitting events much more recently (Wilson et al. 2009). The speciation of *C. coli* and *C. jejuni* has been estimated to have occurred around 6,500 years ago, with *C. coli* clade divergence occurring 1,000-1,700 years ago (Sheppard et al. 2010b). While this large disparity between estimates is difficult to explain, there are reasons for favouring the more recent estimates for *Campylobacter* divergence. Methods that provide recent estimates are based on knowledge of genetic variation within the genus *Campylobacter* and not on the divergence of genera (*E. coli* and *S. typhimurium*) only distantly related to *Campylobacter*; additionally there is an increasing number of studies that use similar approaches and infer a more rapid rate of molecular evolution than in traditional models of bacterial evolution (Falush et al. 2001; Feng et al. 2008; Perez-Losada et al. 2007; Wilson et al. 2009).

If the diversification leading to the population structure in extant *Campylobacter* populations is placed within the last 6,500 years then it correlates with important changes in human behaviour. For example, the development of agriculture, which began in the middle east about 10,000 years ago and became common in Europe about 5,000-3,00 BC (Ammerman & Cavalli-Sforza 1984; McCorriston & Hole 1991; Zvelebil & Dolukhanov 1991) or the establishment of the first cities and the rise of urbanization. Clearly this could have provided novel opportunities for *Campylobacter* to expand into new host species and infect humans in a way that is, to some extent, mirrored in modern society and may have begun to shape the population structure that we observe today.

7. Conclusion

It is evident that the genetic structure that has been described in *C. coli* and *C. jejuni* populations is related to phenotypic factors, such as the animal host from which the isolate was sampled. Furthermore, experimental infections show that genotype is a strong predictor of the host-specific behaviour of a given isolate. Practical applications have effectively exploited this ecology-driven genetic differentiation to attribute the source of human infection but many questions remain about the nature of the forces that result in the highly diverse *Campylobacter* populations. For example, the host association of a particular MLST allele may be influenced by numerous factors including selection for isolates containing particular alleles at loci elsewhere in the genome. As whole genome data become available for large, phenotypically variable isolate collections it will become easier to identify the gene networks that are involved in particular adaptive processes. This has the potential to enhance phylogenetic analysis of *Campylobacter*, and other bacteria, by directly linking the observed population genetic structure and the evolutionary forces that generated it.
8. References


Ecological Factors that Influence Genetic Structure in Campylobacter coli and Campylobacter jejuni


Thakur S, Gebreyes WA (2010) Phenotypic and genotypic heterogeneity of Campylobacter coli within individual pigs at farm and slaughter in the US. Zoonoses Public Health 57 Suppl 1, 100-106.


Mapping phylogenetics on geographical scales is one of the most important scientific aspects of bioscience research. Changes in the environment have evidently shaped the geographical distribution of organisms on land and in the oceans seen today. Overexploitation of key species has caused not only changes in the distribution and diversity of organisms and composition of the ecosystems, but is also leading to species extinction at accelerating rates. It is our duty as scientists to find ways of protecting the species endangered with extinction and preventing other species from entering the endangered stage. To manage this effectively, we need to map species distribution, understand life-history traits, define genetic variation within species and populations, identify lineages - especially at the molecular level - and correlate the historical, phylogenetic components with the spatial distributions of gene lineages. In this book, phylogenetics and phylogeography of a diverse range of organisms are reviewed: from microorganisms causing gastroenteritis in humans, fishes in the Southwest Atlantic Ocean and spiders of the western Indian Ocean, to mountain tapirs in South America and birch tree species of the Arctic tundra.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following: