Published in final edited form as:
Infect Genet Evol. 2013 March ; 14: 200–213. doi:10.1016/j.meegid.2012.12.010.

Comparative characterization of the virulence gene clusters (lipooligosaccharide [LOS] and capsular polysaccharide [CPS]) for *Campylobacter coli, Campylobacter jejuni* subsp. *jejuni* and related *Campylobacter* species

Vincent P. Richards[a], **Tristan Lefébure**[a,2], **Paulina D. Pavinski Bitar**[a], and **Michael J. Stanhope**[a,#]

[a]Department of Population Medicine and Diagnostic Sciences, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853 USA

Abstract

Campylobacter jejuni subsp. *jejuni* and *Campylobacter coli* are leading causes of gastroenteritis, with virulence linked to cell surface carbohydrate diversity. Although the associated gene clusters are well studied for *C. jejuni* subsp. *jejuni, C. coli* has been largely neglected. Here we provide comparative analysis of the lipooligosaccharide (LOS) and capsular polysaccharide (CPS) gene clusters, using genome and cluster sequence data for 36 *C. coli* strains, 67 *C. jejuni* subsp. *jejuni* strains and ten additional *Campylobacter* species. Similar to *C. jejuni* subsp. *jejuni, C. coli* showed high LOS/CPS gene diversity, with each cluster delineated into eight gene content classes. This diversity was predominantly due to extensive gene gain/loss, with the lateral transfer of genes likely occurring both within and between species and also between the LOS and CPS. Additional mechanisms responsible for LOS/CPS diversity included phase-variable homopolymeric repeats, gene duplication/inactivation, and possibly host environment selection pressure. Analyses also showed that (i) strains of *C. coli* and *Campylobacter upsaliensis* possessed genes homologous to the sialic acid genes implicated in the neurological disorder Guillain Barré syndrome (GBS), and (ii) *C. coli* LOS classes were differentiated between bovine and poultry hosts, potentially aiding post infection source tracking.

Keywords

Campylobacter; virulence gene clusters; lipooligosaccharide; capsular polysaccharide; lateral gene transfer; genomics

[#]Please address correspondence to: Michael J. Stanhope, Department of Population Medicine and Diagnostic Sciences, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853 USA, mjs297@cornell.edu, Telephone: 607-253-3859, Fax: 607-253-3083.
[2]Present address: Université de Lyon, F-69622, Lyon, France; Université Lyon 1, Villeurbanne, CNRS, UMR5023, Laboratoire d'Ecologie des Hydrosystèmes Naturels et Anthropisés, France.

1. Introduction

Campylobacter jejuni subsp. *jejuni* and *Campylobacter coli* are recognized as the leading causes of human bacterial gastroenteritis in the industrialized world (Alfredson and Korolik, 2007; Ketley, 1997; Moore et al., 2005), with reported incidences estimated to be between 27 and 880 cases per 100,000 individuals (Blumer et al., 2003; CDC, 2004; Friedman et al., 2000; Gallay et al., 2003; Takkinen et al., 2003; Unicomb et al., 2003; Unicomb et al., 2006). *Campylobacter* infections (*Campylobacter*iosis) can also lead to several complications including toxic mega-colon, hemolytic uremic syndrome, Reiter's syndrome, Miller Fisher syndrome, and Guillain Barré syndrome (GBS).

Although *C. coli* accounts for far fewer infections than *C. jejuni* subsp. *jejuni* (Alfredson and Korolik, 2007; Ketley, 1997; Moore et al., 2005), its impact is still considerable. For example, in Israel, the proportion of *C. coli* contained within *Campylobacter* isolates from diarrheal specimens is consistently 24–30% (Bersudsky et al., 2000), and a 2000 survey of *Campylobacter* infection within the UK showed that *C. coli* accounted for over 25,000 cases of gastroenteritis, with 11 subsequent deaths (Tam et al., 2003).

No suitable animal models (non-primate) of human *Campylobacter*iosis are available. Consequently, little is known of how *Campylobacter* cause disease. However, there are several *Campylobacter* gene clusters involved in human epithelial cell invasion and attachment and are consequently implicated in pathogenesis (Wassenaar and Blaser, 1999). For example, the capsular polysaccharides (CPS) of many bacterial pathogens are known to play an important role in host invasion and subsequent evasion of the host immune response (Roberts, 1996). In *C. jejuni* subsp. *jejuni*, Bacon et al. (2001) demonstrated a role for the CPS in serum resistance, epithelial cell invasion, and diarrhoeal disease. More recently, Jones et al. (2004) demonstrated a role for the CPS in gastrointestinal tract invasion. Lipooligosacharides (LOS), are found on the surface of many mucosal pathogens, and in *C. jejuni* subsp. *jejuni* have been shown to be important in adhesion to human intestinal cells, invasion, and protection from complement-mediated killing (Guerry et al., 2002; McSweegan and Walker, 1986). The LOS are capable of mimicking human antigens (Guerry and Szymanski, 2008), and it is this mimicry that is implicated in GBS and Miller Fisher syndrome (Ang et al., 2004; Willison and O'Hanlon, 1999).

Studies focusing on *C. jejuni* subsp. *jejuni* have revealed a remarkable diversity in gene content for CPS and LOS (Dorrell et al., 2001; Godschalk et al., 2004; Karlyshev et al., 2005; Parker et al., 2008; Pearson et al., 2003). Comparative sequence analysis of the CPS and particularly the LOS for *C. jejuni* subsp. *jejuni* has provided a characterization of these gene clusters and subsequent delineation of strains into numerous gene content classes. Eleven classes are reported for CPS (Karlyshev et al., 2005; Poly et al., 2011) and 18 for LOS (Gilbert et al., 2002; Godschalk et al., 2004; Parker et al., 2008). Importantly, Godschalk et al. (2004) determined that specific LOS classes containing sialylation genes were associated with GBS. Structural variation of the CPS and LOS may represent important *C. jejuni* subsp. *jejuni* strategies for evading the immune response and genetic characterization of *C. jejuni* subsp. *jejuni* CPS and LOS genes have suggested multiple mechanisms responsible for such variation, including (i) lateral gene transfer, (ii) gene inactivation, duplication, deletion, and fusion, and (iii) phase variable homopolymeric tracts (Gilbert et al., 2002; Godschalk et al., 2004; Karlyshev et al., 2005; Parker et al., 2008; Parker et al., 2005; Parkhill et al., 2000).

In contrast to *C. jejuni* subsp. *jejuni*, the genetic characterization of the gene clusters described above have been largely neglected for *C. coli*. An exception is the study of Lang et al. (2010) who presented comparative genomic hybridization data showing a pattern of

high gene content diversity for CPS and LOS, similar to that described for *C. jejuni* subsp. *jejuni*. Here we make use of an extensive *C. coli* genome sequence data set, generated as part of a previous study of ours that addressed bacteria species questions (Lefebure et al., 2010) (see below for details) to provide the first detailed characterization of the CPS and LOS gene clusters for this species. We make additional use of this earlier genome data set by providing new gene cluster data for strains of *C. jejuni* subsp. *jejuni* from multiple hosts, which we combine with data already available for ten additional *Campylobacter* species, to provide a comprehensive comparative perspective of the CPS and LOS gene clusters between *C. coli* and four thermophilic and seven non-thermophilic *Campylobacter* species.

2. Materials and methods

2.1. Strains, sequencing, and assembly

In the previous study mentioned above (Lefebure et al., 2010), Illumina GA II technology was used to sequence genomic DNA obtained from 42 *C. coli* strains isolated from human, turkey, chicken, swine, and bovine hosts, and 43 *C. jejuni* subsp. *jejuni* strains isolated from human, chicken, and bovine hosts (details of the sequencing procedure and original *de novo* assembly are provided therein). Annotation of these *de novo* assemblies was performed as part of this study, using the NCBI Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP). LOS and CPS genes were further annotated using Blast2GO (Gotz et al., 2008). These Whole Genome Shotgun projects have been deposited at DDBJ/EMBL/GenBank under the accessions AIMI00000000 - AIPO00000000. The versions described in this paper are the first versions, AIMI 01000000 - AIPO01000000.

In order to generate contiguous sequence data completely spanning a gene cluster, strains with a cluster contained in multiple contigs in the original Velvet (Zerbino and Birney, 2008) assembly were further assembled using Geneious v5.1.2 (Drummond et al., 2010). Accuracy of these Geneious assemblies was confirmed using Sanger sequencing. The gene content of the CPS gene cluster for two *C. coli* strains (H8 and 2553) was of particular interest (see Results and Discussion). These strains required additional Sanger sequencing to generate contiguous assemblies spanning the complete CPS gene cluster. Table 1 presents details for the 34 *C. coli* strains and 21 *C. jejuni* subsp. *jejuni* strains for which we were able to assemble contiguous LOS and/or CPS gene clusters. The table also shows details for 10 additional *C. coli* and *C. jejuni* subsp. *jejuni* strains for which genome sequences were available and were included in our analyses. Several studies have shown that *C. coli* STs are delineated into three major groups, with one of these groups typically restricted to human and domestic livestock isolates (Colles et al., 2011; Sheppard et al., 2010; Sopwith et al., 2010). The 36 *C. coli* isolates analyzed here were all isolated from humans or domestic livestock and represented 29 distinct STs. Of these, 26 were included in the studies mentioned above and as expected all of them fell within the human/domestic livestock group.

2.2. Orthology assignment

Prior to the genomic era, the identification of orthologs (homologous genes related through speciation from a common ancestral gene present in their last common ancestor) and paralogs (homologous genes related through duplication) was typically accomplished using phylogenetic techniques. However, the high computational demand of these approaches coupled with the need to analyze large genomic data sets has led to the development of complementary approaches not requiring phylogenetic analyses. Rather, these approaches rely on the simplifying assumption that homologs can be identified via reciprocal best BLAST search hits. Here, we combine an implementation of this approach that incorporates a Markov Cluster algorithm with DNA sequence alignments to delineate multiple

Campylobacter strains and species into gene content classes for CPS and LOS gene clusters. This approach allowed precise comparison of the gene content of these virulence gene clusters across multiple species and strains. DNA sequence alignments were performed using MAFFT v6.814b (Katoh et al., 2002) as implemented in Geneious v5.1.2, and OrthoMCL v2.0 (Li et al., 2003) was used to delineate orthologous and paralogous protein sequences. In addition to identifying orthologs, the program also attempts to differentiate between recent paralogs (in-paralogs) and those that pre-date the species split (out-paralogs) (the program groups orthologs with putative in-paralogs). We analyzed 114 *Campylobacter* strains as follows: (i) genome sequences from 36 C. *coli* strains (34 strains sequenced as part of this study plus two additional strain sequences obtained from NCBI), (ii) 29 genome sequences from *C. jejuni* subsp. *jejuni* strains (21 strains sequenced in this study plus eight additional strain sequences obtained from NCBI), (iii) genome sequences obtained from NCBI for three additional thermophilic *Campylobacter* species: *Campylobacter jejuni* subsp. *doylei* (n=1), *Campylobacter lari* (n=1) *Campylobacter upsaliensis* (n=2), (iv) single genome sequences obtained from NCBI for seven non-thermophilic *Campylobacter* species: *Campylobacter rectus, Campylobacter showae, Campylobacter concisus, Campylobacter curvus, Campylobacter fetus* subsp. *fetus, Campylobacter gracilis, Campylobacter hominis*, (v) NCBI sequences of the LOS gene cluster from 34 *C. jejuni* subsp. *jejuni* strains, and (vi) NCBI sequences of the CPS gene cluster from six *C. jejuni* subsp. *jejuni* strains. For *C. jejuni* subsp. *jejuni* strain 81–176, the analysis included two separate sequences. The first was a CPS sequence, and the second was a complete genome sequence. For *C. jejuni* subsp. *jejuni* strain RM1862, the analysis also included two separate sequences. The first was a LOS sequence, and the second was a complete genome sequence. See Tables 1, 2, 3, and 4 for Accession and strain ID numbers.

The first step in the OrthoMCL procedure was to perform a reciprocal BLASTp among protein sequences. The resulting e-values were then used to build a normalized similarity matrix, which was analyzed using a Markov Cluster algorithm to delineate proteins into "OrthoMCL groups" containing sets of orthologs and/or recent paralogs. Proteins were considered recent paralogs if they were more similar to each other than to any protein from another strain. Following Li et al. (2003), an e-value cut-off of 1e-5 was used in the BLASTp. Throughout the manuscript, the term ortholog refers to an OrthoMCL group/cluster.

2.3. Phylogenetics, gene content, recombination, and homopolymeric tracts

Gene content similarity across strains for each of the gene clusters, as well as for the remainder of the genome, was explored using the presence or absence of the orthologs delineated using the orthology assignment procedure. Presence/absence of orthologs was used to generate binary sequences that were used to construct a split network using the Neighbor-Net procedure (Bryant and Moulton, 2004), as implemented in the program SplitsTree4 v4.9.1 (Huson and Bryant, 2006). We tested for evidence of recombination for each of the LOS and CPS orthologs as follows. First, nucleotide sequences were aligned using Probalign v1.1 (Roshan and Livesay, 2006). We then tested for recombination using a combination of three methods: Pairwise Homoplasy Index (PHI), Neighbour Similarity Score (NSS), and Maximum χ^2, as implemented in the program PhiPack (Bruen et al., 2006). The first two methods (PHI and NSS) are compatibility methods that examine pairs of sites for homoplasy (Bruen et al., 2006; Jakobsen and Easteal, 1996). Maximum χ^2 is a substitution distribution method that searches for significant clustering of substitutions at putative recombination break points (Maynard Smith, 1992).

Evolutionary relationships among strains were compared to the distribution of LOS and CPS classes among strains. For *C. coli* and *C. jejuni* subsp. *jejuni* the number of core orthologs (orthologs seen in all strains of each species) were 832 and 774 respectively. The nucleotide

sequences for these orthologs were aligned using Probalign and then tested for evidence of recombination using the procedure outlined above. Orthologs showing evidence for recombination for one or more of the methods (*C. coli* = 433, *C. jejuni* subsp. *jejuni* = 568) were removed. Maximum Likelihood (ML) phylogenies (gene-trees) for each of the remaining orthologs (*C. coli* = 399, *C. jejuni* subsp. *jejuni* = 206) were constructed using PhyML v3.0. The GTR substitution model was employed. Using the two sets of gene trees, consensus phylogenies (species-trees) for each species were constructed using the Triple Construction Method as implemented in the program Triplec (Ewing et al., 2008). Species trees with four or more taxa can produce anomalous gene-trees (AGTs), which for a given data set, are more probable to observe than a tree that is congruent with the species-tree (Degnan and Rosenberg, 2006). Based on the observation that rooted three taxa trees (rooted triples) do dot exhibit AGTs, the Triple Construction Method searches all gene-trees for the most frequent of the three possible rooted triples for each set of three taxa. Once found, the set of rooted triples are joined to form the species-tree using the quartet puzzling heuristic (Strimmer and vonHaeseler, 1996). The procedure has been shown to be a statistically consistent estimator of the species-tree topology and to out perform majority-rule and greedy consensus methods (Degnan et al., 2009). The distribution of LOS and CPS classes among strains was overlain on each of the species-trees.

Open reading frames from single representatives of each class for each gene cluster and species were searched for homopolymeric sequence repeats (i.e. $A_n/T_n/G_n/C_n$) using the script Poly (Gotz et al., 2008). Fragment ORFs clearly resulting from frame shifts caused by homopolymeric sequence repeats were deemed transient due to phase variation and not regarded as valid orthologs.

3. Results and discussion

3.1. Gene content diversity

Similar to that described for *C. jejuni* subsp. *jejuni*, our orthology assignment revealed a high level of gene content diversity in *C. coli* for both the LOS and CPS gene clusters. For example, we detected 51 distinct orthologs occurring within *C. coli*'s LOS, with these 51 orthologs occurring in eight distinct combinations, which we designated classes I through VIII (Table 1, Fig. 1). Following Gilbert et al. (2002), the conserved LOS biosynthesis genes *waaC* and *waaF* were considered the first and last genes of the cluster respectively. Orthologs occurring between (and including) these genes were considered as part of the LOS cluster. For *C. jejuni* subsp. *jejuni*, we detected 40 LOS orthologs, 11 fewer than observed for *C. coli*. However, these orthologs occurred in more combinations producing 22 distinct classes, which included four new classes not previously described. The previously described 18 LOS classes (Gilbert et al., 2002; Godschalk et al., 2004; Parker et al., 2008) are designated A through S (excluding N for *C. jejuni* subsp. *doylei*). Here we continue this nomenclature and designate the four new classes as T through W (Table 1, Fig. 1).

For the CPS, we followed Karlyshev et al. (2005) and considered the conserved genes *kpsF* and *kpsC* as the first and last genes of the cluster respectively. The CPS for both *C. coli* and *C. jejuni* subsp. *jejuni* contained almost twice the number orthologs than observed within the LOS (95 and 93 orthologs respectively). However, these orthologs were delineated into the same number of classes for *C. coli* (eight) and fewer classes for *C. jejuni* subsp. *jejuni* (12). The eight CPS classes for *C. coli* were designated I through VIII (Table 1, Fig. 1). The 12 CPS classes detected for *C. jejuni* subsp. *jejuni* included seven new classes not previously described, which we designated as classes F through L following Karlyshev et al. (2005) (Table 1, Fig. 1). *C. jejuni* subsp. *doylei* possessed a distinctive class which we designated M.

Subsequent to completion of our analyses, additional CPS sequence data for eight strains of *C. jejuni* subsp. *jejuni* were deposited at NCBI (Poly et al., 2011) (human isolates). An additional strain (ATCC43457) with the Accession EU200439 is also in the database, however, there appears to be no associated publication regarding generation of these data. Although the absence of these additional sequences from our analyses did not significantly affect the major findings of our study, we nevertheless proceeded to determine whether these sequences represented additional CPS classes. We constructed a database comprising 12 contiguous sequences representing each of the 12 classes described above and then used discontinuous Megablast to search it for the nine new sequences. Two pairs of the new sequences had identical gene content and none of the nine sequences had contiguous matches along their entire length with any of the sequences in the database, suggesting that these sequences represented an additional seven CPS classes for *C. jejuni* subsp. *jejuni*. Typically, there were matches for contiguous stretches of four to seven genes. However, strain ATCC43442 was very similar to class F, with this latter class having four additional genes. Accession numbers representing the seven additional classes are as follows: HQ343267, HQ343269, HQ343270 HQ343271, HQ343272, HQ343274, and EU200439.

3.2. Sialic acid biosynthesis orthologs

Previous studies have demonstrated that ganglioside mimicry is an important factor contributing to *C. jejuni* subsp. *jejuni*'s ability to cause Guillain Barré syndrome (GBS) (Gilbert et al., 2000; Godschalk et al., 2004). Three classes of *C. jejuni* subsp. *jejuni* LOS (A, B, and C) have been shown to be the most frequently associated with GBS (Godschalk et al., 2004; Parker et al., 2005), and these three classes possess genes capable of synthesizing and transferring sialic acid, which is an essential component of gangliosides (Godschalk et al., 2004). Specifically, these genes (which cluster together) are three sialic acid biosynthesis genes (*neuBCA*) and a gene encoding a sialic acid transferase (*cst*). More recently, Parker et al. (2008) discovered two new classes (R and M) that also contain these genes and thus have potential to cause GBS. Here we present an additional class (V) that also possess these genes and therefore likely has a similar potential (Table 1, Fig. 1).

The ortholog identification numbers arising from the OrthoMCL analysis corresponding to each of the sialic acid genes were as follows: *cst* = 1501, *neuB* = 9, *neuC* = 1645, *neuA* = 58 (however, also see discussion below for an expanded delineation of ortholog 58). The discussion that follows makes use of ortholog numerical identifiers, partly to make it easier to follow along with Fig. 1, and partly because the LOS cluster in particular has a history of various gene names being applied to the different loci; Table S1 in the supplementary material should be consulted for the associated locus tags/protein IDs, annotations, and GO-terms. Table S1 also contains the previous gene identifiers of Godschalk et al. (2004) and Parker et al. (2008) so that they can be cross-referenced to our ortholog identification numbers. Three of the sialic acid orthologs (9, 1645, and 58) were also found in a contiguous cluster in two strains of *C. coli* (H8 and 2553). However, rather than occurring in the LOS, these orthologs were found imbedded in the CPS. Strains H8 and 2553 represent two distinct CPS classes for *C. coli* (VII and VIII respectively). The fourth ortholog seen in the sialic acid gene cluster of *C. jejuni* subsp. *jejuni* (1501) was absent in *C. coli*. However, in its place was another ortholog (2742), which for strains H8 and 2553 had very similar functional annotation to the CDS representing 1501. The PGAAP annotation for 1501 was alpha-2,3 sialyltransferase, whereas 2742 in strains H8 and 2553 was annotated as a hypothetical protein. However, the Blast2GO annotation for both orthologs was alpha sialyltransferase. Furthermore, both orthologs shared the same molecular function GO-term of “transferase activity, transferring glycosyl groups,” suggesting a role as a sialic acid transferase.

Although preliminary without further functional analyses, these findings nevertheless suggest that (i) certain strains of *C. coli* may have the potential to cause GBS, and (ii) rather than the LOS, the CPS might be involved in ganglioside mimicry. Furthermore, our findings appear concordant with previous studies examining the potential of *C. coli* to induce GBS. For example, Bersudsky et al. (2000) examined a *C. coli* strain isolated from a patient with GBS and found evidence for a ganglioside-like epitope within its lipopolysaccharide (LPS) [subsequently, Karlyshev et al. (2000) demonstrated that the *C. jejuni* subsp. *jejuni* LPS was in fact the CPS]. Whereas Funakoshi et al. (2006) and van Belkum et al. (2009) found no evidence for ganglioside mimicry in the LOS of *C. coli* strains isolated from GBS patients.

It is of note that for *C. coli* only strains H8 and 2553 contained *neuBCA* and a sialyltransferase contiguously. Most of the remaining *C. coli* strains contained only one or two of these genes and they were located external to the LOS or CPS (Table S2 supplementary material). For *C. jejuni* subsp. *jejuni,* only classes A, B, C, M, R, and V contained *neuBCA* and *cst* contiguously. Typically, these classes also contained an additional copy of *neuB* external to the LOS. For the remaining classes, the majority had only one or two of these genes typically within the LOS.

We also found the four sialic acid orthologs (1501, 9, 1645, and 58) within the genome sequence of *Campylobacter upsaliensis*, strain JV21. Although the loci were contiguous, they were clustered external of the LOS (254 CDS separated them). Interestingly, a second *C. upsaliensis* strain (RM3195), which had been associated with GBS (Goddard et al., 1997), lacked these four orthologs. The sensitivity of *C. upsaliensis* isolation methods has been questioned in the literature (Byrne et al., 2001; Lastovica and Le Roux, 2001). Consequently, it's possible that the GBS patient infected with strain RM3195 may have been carrying additional strain/s of *C. upsaliensis* that did possess the sialic acid genes. *C. upsaliensis* is an emerging human gastrointestinal pathogen with two previous reports of association with GBS (Bourke et al., 1998; Goddard et al., 1997; Ho et al., 1997). Furthermore, the pathogen is frequently isolated from domestic cats and dogs, with evidence of transmission to humans (Acke et al., 2009; Bourke et al., 1998; Goossens et al., 1991; Gurgan and Diker, 1994). Although *C. jejuni* subsp. *jejuni* is also associated with domestic pets, recent studies have shown *C. upsaliensis* to be the most frequently recovered species from dogs (results for cats were less conclusive) (Acke et al., 2009; Koene et al., 2009; Parsons et al., 2010; Westgarth et al., 2009). Our findings suggest that there is the genetic potential for some strains of *C. upsaliensis* to induce GBS, and this in turn suggests the potential importance of domestic dogs as a potential reservoir for pathogens linked to this sequela.

3.3. Gene duplication

CDS representing ortholog 58 occurred frequently among and within *C. jejuni* subsp. *jejuni* LOS classes. For example, in addition to occurring in half of all the classes, this gene (or perhaps more accurately "gene-group") also occurred twice in each of classes A and R and three times in class B. Furthermore, it also occurred occasionally within the CPS for both *C. jejuni* subsp. *jejuni* and *C. coli*. Using nucleotide sequences, we investigated the phylogenetic relationships of the members of this putative gene family. For *C. jejuni* subsp. *jejuni* LOS classes, CDS were sub-divided into six distinct groupings, which we designated 58-1a, 58-2, 58-3a, 58-3b, 58-4, and 58-5 (see Fig. S1 in the supplementary material). The family included CDS annotated as *neuA* (58-1a and 58-1b) and *ctgA* (58-2, 58-3a, 58-3b, and 58-5). Group 58-4 contained CDS that were an "in-frame fusion" of CDS from groups 58-1a and 58-2, as described previously by Gilbert et al. (2000). A single CDS designated 58-1b that was closely related to group 58-1a was found within the CPS of *C. jejuni* subsp. *jejuni* (class J). The LOS classes most frequently associated with GBS (A, B, and C) carried CDS from groups 58-1a and 58-4. However, classes R and M also contained CDS from

group 58-1a, while class V contained CDS from group 58-4. CDS representing ortholog 58 from *C. coli* and *C. upsaliensis* that were contiguous with the other sialic acid genes formed two groups in the network that were designated 58-6a and 58-6b respectively. These CDS were closely related to CDS within group 58-1a (75.4% average pairwise nucleotide sequence identity), supporting their possible role sialic acid biosynthesis and GBS.

This analysis also suggests that *neuA* and *ctgA* are paralogs resulting from historical gene duplication. Subsequent to the *neuA* – *ctgA* duplication, *ctgA* appears to have undergone multiple additional duplications, with the 58-3a – 58-3b duplication being the most recent. Indeed, gene duplication appears to be a common feature of the evolutionary history of LOS and CPS for *C. jejuni* subsp. *jejuni*, with duplications in seven LOS classes and seven CPS classes (Fig. 1). Gene duplication events were also apparent in *C. coli*, occurring once in a LOS class and in two separate CPS classes. It seems likely that the fewer number of gene duplications observed for *C. coli* compared to *C. jejuni* subsp. *jejuni* is simply a reflection of the fewer number of *C. coli* LOS and CPS classes. Of the remaining *Campylobacter* species, we also detected duplication in *C. lari*, where ortholog 6 occurred three times (Fig. 2). The presence of multiple copies of an ortholog within a strain could also be due to lateral gene transfer.

3.4. Gene gain/loss and recombination

Numerous studies have provided evidence for lateral gene transfer (LGT) between the LOS of distinct strains of *C. jejuni* subsp. *jejuni* (Gilbert et al., 2004; Parker et al., 2008; Phongsisay et al., 2006). The bacterial core genome species concept first introduced by Dykhuizen and Green (1991) and later refined by Lan and Reeves (Lan and Reeves, 2000; 2001) proposed that the dispensable genome would turn over rapidly due to frequent LGT, and recent genomic work has supported this proposal (Donati et al., 2010). More specifically, Lefébure et al. (2010) showed that inter-species recombination for *C. coli* and *C. jejuni* subsp. *jejuni* predominated within the dispensable component of the genome. For the strains analyzed here, we regarded genes as dispensable for each species if they were not shared among all strains for the respective species. Core genes were shared among all strains of a species. With the exception of the five LOS and two CPS core flanking genes, all LOS and CPS genes for both species belonged to the dispensable genome.

The principal forces that shape gene content diversity are gene gain/loss events within and among strains, and attempts at reconstructing evolutionary relationships using gene content are sensitive to these events (Kunin and Ouzounis, 2003). However, the number of genes shared by genomes has been shown to depend on evolutionary distance (Fitz-Gibbon and House, 1999; Snel et al., 1999; Tekaia et al., 1999). Therefore, if gene gain/loss events occur more frequently within the dispensable genome of strains of the same species, gene content analyses should group strains by species. We constructed a gene content network using genome wide orthologs not shared among all *C. coli* and *C. jejuni* subsp. *jejuni* strains (i. e. the collective dispensable genome) and *C. coli* and *C. jejuni* subsp. *jejuni* grouped strongly by species (Fig. S2 in the supplementary material). In contrast, gene content networks based exclusively on LOS and CPS orthologs showed the LOS and CPS classes to not group by species (Fig. 3). Two possible explanations for these observations are (i) frequent LGT between *C. coli* and *C. jejuni* subsp. *jejuni* for LOS and CPS genes, and (ii) gene loss events that are decreasing the gene content distinctiveness of each species (i.e. the loss of species-specific LOS/CPS dispensable genes). There were 188 distinct orthologs within the LOS and CPS for *C. coli* and *C. jejuni* subsp. *jejuni*, and 113 (60.1%) of these orthologs showed evidence for recombination for one or more of the recombination methods. This result supports frequent LGT for the LOS and CPS genes. It should be noted that our power to detect recombination for orthologs only occurring in a few strains was diminished due to the lower number of sequences in the respective ortholog alignment. The distribution of

recombinant orthologs provides support for inter-species LGT (Fig. 4). For example, 71% of LOS ortholgs and 65% of CPS orthologs shared between both species were recombinant. On average, these frequencies were approximately equal to those for recombinant LOS/CPS orthologs specific to each species (*C. coli* = 52% and *C. jejuni* subsp. *jejuni* = 71%).

The frequency of recombinant orthologs present in at least three of the following four possible locations (i) *C. coli* LOS, (ii) *C. coli* CPS, (iii) *C. jejuni* subsp. *jejuni* LOS, and (iv) *C. jejuni* subsp. *jejuni* CPS was 60%, providing support for LGT between the LOS and CPS. The sialic acid cluster of orthologs (1501-9-1645-58) provide a specific example of this type of exchange: these orthologs are common in the LOS of many strains of *C. jejuni* subsp. *jejuni*, absent in the LOS of *C. coli*, and yet three of them are present in the CPS of *C. coli*, suggesting that they have been exchanged inter-specifically between the LOS and CPS. The much higher frequency of these orthologs in *C. jejuni* subsp. *jejuni* suggests that they may have been transferred from this species to *C. coli*. Indeed, Sheppard et al. (2008) recently showed genetic exchange between these species to be strongly biased in the *C. jejuni* subsp. *jejuni* to *C. coli* direction. Furthermore, for both *C. coli* CPS classes VII and VIII, the sialic acid genes were part of a five-gene cassette (58, 1645, 9, 2742, and 2688) that was oriented in the opposite direction to all remaining CPS genes in all classes (Fig. 1), again suggesting that these genes were imported into the CPS. A previous comparison of *C. coli* (RM2228) and *C. jejuni* subsp. *jejuni* (RM1221) suggested that LGT between these two strains was biased towards certain biological functional categories. Here we highlight that LGT for LOS and CPS genes is an important component of genetic exchange both within and between *C. coli* and *C. jejuni* subsp. *jejuni.*

While LOS and CPS genes are fundamental virulence components for *C. coli* and *C. jejuni* subsp. *jejuni*, with the exception of the conserved flanking genes, they are components of a large collective pool of dispensable genes. It appears that many of these genes can be assembled into multiple combinations of LOS and CPS classes, providing both these species with the ability to adapt and evade patient defenses. The extent of this diversity is apparent when you compare the much higher LOS ortholog diversity within *C. coli* and *C. jejuni* subsp. *jejuni* to the diversity observed among all the non-thermophilic species combined (Fig. 3), which likely reflects the greater virulence potential of *C. coli* and *C. jejuni* subsp. *jejuni*. Furthermore, this shared dispensable gene pool may extend to other thermophilic species, but exclude the non-thermophilic species. For example, *C. jejuni* subsp. *doylei*, and *C. lari* shared at least 70% of their LOS and CPS orthologs with other thermophilic species, whereas the thermophilic and non-thermophilic species shared no LOS dispensable orthologs (Fig. 2). A phylogeny for the genus supports the non-thermophilic and thermophilic species as two distinct clades (Lefebure and Stanhope, 2009), suggesting the dispensable LOS loci first appeared on the branch leading to the thermophilic species. The LOS of *C. upsaliensis* was unusual among the thermophiles in that it only shared three dispensable orthologs with the other thermophilic species (Fig. 2). *C. upsaliensis* was particularly distinctive in that it also lacked a characteristic CPS (see below). However, despite the apparent lack of lateral exchange involving the LOS for *C. upsaliensis*, this species nevertheless may have obtained the sialic acid genes via LGT, as the absence of these genes from *C. upsaliensis* strain RM3195 suggests they were either gained via LGT or present in the ancestor and lost in RM3195.

3.5. Comparison to non-thermophilic Campylobacter

With the exception of two non-thermophilic species (*C. hominis* and *C. gracilis*), all *Campylobacter* species possessed LOS gene content types unique to each taxon. The diversity of LOS types is undoubtedly a reflection of the fact that there are many genes serving as potential LOS loci and thus many combinations are possible. For several of these taxa there is at present only a single genome sequence. However, it is of note that with the

LOS cluster sequenced for 33 *C. coli* and 60 *C. jejuni* subsp. *jejuni* strains, we found no evidence of interspecific overlap in LOS class, but plenty of intraspecific/interstrain overlap in LOS class, tending to support a biological distinction between these two taxa. Typically, genes *waaCM* (5' end) and *waaVF* (3' end) flank the LOS gene cluster. However, orthologs corresponding to *waaC*, and *waaF* were absent from *C. hominis*, and orthologs corresponding to *waaC, waaV*, and *waaF* were absent from *C. gracilis*, with both species lacking a characteristic cluster of LOS biosynthesis genes. For *C. hominis*, this may reflect the fact that this species is likely a non-pathogenic commensal of the human gastrointestinal tract (strains have been repeatedly isolated from human fecal material obtained from healthy individuals) (Lawson et al., 2001). In addition, *C. hominis* and *C. gracilis* also share an "unusual aflagellate rod-like cell structure" and high 16S rDNA sequence similarity (Lawson et al., 2001) suggesting a close relationship between these two species.

With the exception of *C. upsaliensis*, the CPS gene cluster of the thermophilic *Campylobacter* species was flanked by genes *kpsFDETM* (5' end) and *kpsCS* (3' end). Although *C. upsaliensis* possessed both of these sets of genes, they were separated by approximately 83kbp, and there were no genes characteristic of the CPS immediately downstream of *kpsFDETM* (e.g. capsular polysaccharide biosynthesis and transfer). None of the non-thermophilic species possessed the cluster of CPS biosynthesis genes typical of the thermophilic species. For example, with the exception of *kpsF* and *kpsT*, all orthologs corresponding to the *kps* genes were absent from the non-thermophilic species (the ortholog corresponding to *kpsT* was only present in *C. rectus, C. showae*, and *C. fetus* subsp. *fetus*).

3.6. Distribution of LOS and CPS classes among hosts

Previous studies for *C. jejuni* subsp. *jejuni* and *C. coli* have shown genetic partitioning among isolates with regard to source of isolation (Champion et al., 2005; Colles et al., 2011; Miller et al., 2006; Sheppard et al., 2010). Therefore, employing split networks depicting shared ortholog content, we investigated the possibility that LOS and CPS classes might have non-random distributions among hosts. The results revealed a non-random pattern for the LOS of *C. coli*. For example, all bovine strains were restricted to classes IV, V, and VI (Table 1), and strains possessing these LOS classes formed a tight cluster (I) in the split network (Fig. 3). This cluster also included the sole representative of class VII, a human sourced isolate. Our analysis included 33 *C. coli* strains isolated from four different host groups: human (n=12), poultry (turkey=6, chicken=3), bovine (n=7), and swine (n=5). Although there was overlap regarding human and swine sourced isolates between cluster I and all remaining strains, none of the poultry sourced isolates occurred in cluster I. Consequently, our results show strong partitioning between bovine and poultry sourced isolates, with bovine sourced isolates restricted to classes IV, V, and VI and poultry sourced isolates restricted to classes I, II, and III (Fisher exact test: $P = 0.0006$). In addition, the phylogenetic analysis (Fig. 5A) showed strong support for a grouping that contained all but one of the bovine sourced isolates.

Our findings are concordant with the previous genomic hybridization work of Lang et al. (2010) that showed a tendency for strains to cluster by isolation source and the study of Miller et al. (2006), which showed MLST alleles to be source associated. However, our findings additionally suggest a possibly more precise basis for this association, in that strains possessing a specific LOS structure might be more successful at colonizing certain species. More specifically, classes in cluster I possessed seven orthologs (8, 1742, 1743, 1790, 1959, 2066, and 2089) not seen in any other *C. coli* LOS class (see Table S1 in the supplementary material for annotations). Two of these orthologs (1742 [glucose-1-phosphate thymidylyltransferase] and 1743 [dTDP-glucose dehydratase]) were involved in glucose metabolism. With the exception of class VIII, none of the other *C. coli* classes possessed genes involved in glucose metabolism. Although class VIII also possessed glucose

metabolism genes, it was distinct from all other LOS classes (including *C. jejuni* subsp. *jejuni*) as it contained two genes for mannose metabolism. Concordant with Miller et al. (2006), we also detected less genetic diversity for bovine *C. coli*, suggesting a more recent colonization. The partitioning of *C. coli* strains by source, could lead to more efficient source tracking following human infection (Miller et al., 2006).

3.7. Convergence

Comparison of the evolutionary relationships among strains to the distribution of LOS classes among strains can be seen in Fig. 5. Despite low support for several sections of the *C. coli* phylogeny (species-tree), many groupings were supported by greater than 50% of the gene-trees. Focusing on these groupings, there were multiple examples of incongruence between evolutionary relationships among strains and the distribution of LOS classes. Specifically, there were examples of strains with different evolutionary histories converging on the same class for six of the eight classes (I, II, III, IV, V, and VI) It was impossible to assess convergence for class VII as it was only represented by a single strain. A possible explanation for this observation may be related, in part, to the source partitioning as described above. For example, classes I, II, and III have converged to poultry whereas classes IV, V, and VI have converged to bovine, with this convergence possibly driven by source environment selection. A similar pattern has been observed for mammalian gut bacterial communities converging on diet rather than the mammalian phylogeny (Ley et al., 2008; Muegge et al., 2011). Nucleotide sequence identities among *C. coli* LOS orthologs provide support for this hypothesis. For example, orthologs 6 (putative glycosyltransferase), 1541 (glycosyltransferase), and 1715 (hypothetical protein) were among those seen in both classes II and IV (Fig. 1). Average pairwise identities among strains within each class for each of these orthologs was very high (ranging between 99.8% and 100%). Whereas identities between the classes for these orthologs was considerably lower. For example, pairwise identity between two strains from each class was 81.7% (6), 80.4% (1541), and 76.5% (1715). For a particular ortholog seen in different classes, these observations suggest an older split between the classes, with a more recent proliferation within a class, possibly driven by selection. The distribution of CPS classes among *C. coli* strains was less conclusive with only two examples of convergence (classes III and VI) (Fig. S3 in the supplementary material). However, additional sampling will be required to more accurately assess the distribution of *C. coli* CPS classes among sources (see Table 1).

Overall, groupings within the *C. jejuni* subsp. *jejuni* phylogeny (species-tree) were well supported (Fig. 5). Seven classes were represented by multiple strains. Of these, five showed examples of convergence: A, B, E, H, and R. Sequence identities among *C. jejuni* subsp. *jejuni* LOS orthologs also showed a similar pattern to *C. coli*. For example, orthologs 6 (galactosyltransferase), 9 (*neuB*), and 1645 (*neuC*) were among those seen in classes A, B, C and R (Fig 1). Average pairwise identities among strains within each class for each of these orthologs was again high (ranging between 96.9% and 100%). Whereas identities between the classes for these orthologs was considerably lower. For example, pairwise identity between single strains from each class was 87.7% (6), 85.8% (9), and 84.3% (1645). Following the same logic as for *C. coli*, these observations again suggest recent selection for these classes. However, the factors responsible are less clear, as unlike *C. coli*, there is no apparent source partitioning for *C. jejuni* subsp. *jejuni*. For the *C. jejuni* subsp. *jejuni* CPS, there were only four strains that could be compared: two for class I, and two for class F. Interestingly, neither showed a pattern of convergence (phylogeny not shown). However, additional samples should be analyzed before convergence is rejected.

3.8. Additional mechanisms creating diversity

Phase variation due to homopolymeric tracts is often reported as an additional mechanism responsible for LOS and CPS structural variation in *C. jejuni* subsp. *jejuni* (Gilbert et al., 2002; Godschalk et al., 2004; Karlyshev et al., 2005; Parker et al., 2008; Parker et al., 2005; Parkhill et al., 2000). We assessed the potential for this mechanism to cause similar variation in *C. coli* and the other *Campylobacter* species included in this study by searching the nucleotide sequence of open reading frames (ORFs) from the LOS and CPS for homopolymeric sequence repeats. Results showed *C. coli* to have a comparable number of homopolymeric sequence repeats to *C. jejuni* subsp. *jejuni* (Table S3 in the supplementary material), suggesting that *C. coli* had a similar capacity for phase variation as that previously reported for *C. jejuni* subsp. *jejuni*. Indeed, for those *C. coli* classes containing multiple strains, sequence alignments revealed four clear instances where a homopolymeric tract appeared to have caused a reading frame shift (see Fig. 1). However, the possibility that sequencing error may have caused these shifts should also be considered.

The potential for phase variation was also detected for *C. upsaliensis*. Strain JV21 contained a homopolymeric C tract (C_{13}) at the 3' end of ortholog 1501 (one of the four orthologs associated with GBS). The homopolymeric tract appeared to have shifted the reading frame for this ortholog resulting in its premature termination and subsequent creation of a small 156bp ORF. Alignment to *C. jejuni* subsp. *jejuni* showed this small ORF to be a fragment of ortholog 1501, and the addition of one extra C to the C_{13} tract via manual editing extended the ORF to encompass the entire length of ortholog 1501. This suggests that *C. upsaliensis* might be able to phase modulate the genes implicated in GBS. Overall, the thermophilic species had more repeats than the non-thermophilic species, suggesting the potential for adaptive flexibility and increased virulence of the thermophilic *Campylobacter* species. Another mechanism likely responsible for LOS and CPS structural diversity is gene inactivation due to the disruption of reading frames by substitutions, insertions and deletions. An example of this process for *C. coli* involves class III where multiple indels have fragmented ortholog 1501 into several distinct ORFs.

4. Conclusion

Although mechanisms such as phase variation and gene duplication are important factors creating LOS and CPS structural diversity for both *C. coli* and *C. jejuni* subsp. *jejuni*, the most important factor appears to be that these genes comprise a highly dynamic component of the dispensable genome and that frequent LGT has likely contributed to their assembly into a highly diverse array of combinations. This genetic exchange might also extend to other thermophilic *Campylobacter* species such as *C. jejuni* subsp. *doylei* and *C. lari*. In addition to suggesting a close association among these thermophilic *Campylobacter* species, this also suggests repeated LGT within a shared gastrointestinal environment. Selection pressure of source environment might be an additional factor creating LOS gene content diversity for *C. coli*. The thermophilic species had in general much higher diversity than the non-thermophilic species likely reflecting increased virulence potential. In addition to evading patient defenses, the highly dynamic nature of the LOS/CPS gene content likely has important epidemiological implications. For example, recent bovine colonization by *C. coli* may have been facilitated by the capacity to rapidly produce a distinct combination of LOS genes, with the resulting source differentiation in turn providing a valuable aid for source tracking. Furthermore, repeated LGT of the dispensable LOS/CPS genes may have transferred genes implicated in GBS from *C. jejuni* subsp. *jejuni* to *C. coli*, with *C. upsaliensis* possibly involved in a similar type of exchange. Although preliminary without further functional analyses, these findings highlight the potential for GBS development following *Campylobacter* infection from swine, and domestic cat/dog sources.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This project was supported by the Cornell University Zoonotic Research Unit of the Food and Waterborne Diseases Integrated Research Network, which was funded by the National Institute of Allergy and Infectious Diseases, US National Institutes of Health, under contract number N01-AI-30054 (ZC003-05 and ZC010-09) awarded to M.J.S.

References

Acke E, McGill K, Golden O, Jones BR, Fanning S, Whyte P. Prevalence of thermophilic *Campylobacter* species in household cats and dogs in Ireland. Vet Rec. 2009; 164:44–47. [PubMed: 19136683]

Alfredson DA, Korolik V. Antibiotic resistance and resistance mechanisms in *Campylobacter* jejuni and *Campylobacter coli*. FEMS Microbiol Lett. 2007; 277:123–132. [PubMed: 18031331]

Ang CW, Jacobs BC, Laman JD. The Guillain-Barre syndrome: a true case of molecular mimicry. Trends Immunol. 2004; 25:61–66. [PubMed: 15102364]

Bacon DJ, Szymanski CM, Burr DH, Silver RP, Alm RA, Guerry P. A phase-variable capsule is involved in virulence of *Campylobacter jejuni* 81–176. Mol Microbiol. 2001; 40:769–777. [PubMed: 11359581]

Bersudsky M, Rosenberg P, Rudensky B, Wirguin I. Lipopolysaccharides of a *Campylobacter coli* isolate from a patient with Guillain-Barre syndrome display ganglioside mimicry. Neuromuscul Disord. 2000; 10:182–186. [PubMed: 10734265]

Blumer C, Roche P, Spencer J, Lin M, Milton A, Bunn C, Gidding H, Kaldor J, Kirk M, Hall R, Della-Porta T, Leader R, Wright P. Australia's notifiable diseases status, 2001: annual report of the National Notifiable Diseases Surveillance System. Commun Dis Intell. 2003; 27:1–78.

Bourke B, Chan VL, Sherman P. *Campylobacter upsaliensis* : waiting in the wings. Clin Microbiol Rev. 1998; 11:440–449. [PubMed: 9665977]

Bruen TC, Philippe H, Bryant D. A simple and robust statistical test for detecting the presence of recombination. Genetics. 2006; 172:2665–2681. [PubMed: 16489234]

Bryant D, Moulton V. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. Mol Biol Evol. 2004; 21:255–265. [PubMed: 14660700]

Byrne C, Doherty D, Mooney A, Byrne M, Woodward D, Johnson W, Rodgers F, Bourke B. Basis of the superiority of cefoperazone amphotericin teicoplanin for isolating *Campylobacter upsaliensis* from stools. J Clin Microbiol. 2001; 39:2713–2716. [PubMed: 11427603]

CDC. Preliminary FoodNet data on the incidence of infection with pathogens transmitted commonly through food – selected sites, United States. Morbidity and Mortality Weekly Report. 2004; 53:338–343. [PubMed: 15123986]

Champion OL, Gaunt MW, Gundogdu O, Elmi A, Witney AA, Hinds J, Dorrell N, Wren BW. Comparative phylogenomics of the food-borne pathogen *Campylobacter jejuni* reveals genetic markers predictive of infection source. Proc. Natl. Acad. Sci. U. S. A. 2005; 102:16043–16048. [PubMed: 16230626]

Colles FM, Ali JS, Sheppard SK, McCarthy ND, Maiden MC. *Campylobacter* populations in wild and domesticated Mallard ducks (Anas platyrhynchos). Environ Microbiol Rep. 2011; 3:574–580. [PubMed: 22164198]

Degnan JH, DeGiorgio M, Bryant D, Rosenberg NA. Properties of consensus methods for inferring species trees from gene trees. Syst Biol. 2009; 58:35–54. [PubMed: 20525567]

Degnan JH, Rosenberg NA. Discordance of species trees with their most likely gene trees. PLoS Genet. 2006; 2:e68. [PubMed: 16733550]

Donati C, Hiller NL, Tettelin H, Muzzi A, Croucher NJ, Angiuoli SV, Oggioni M, Dunning Hotopp JC, Hu FZ, Riley DR, Covacci A, Mitchell TJ, Bentley SD, Kilian M, Ehrlich GD, Rappuoli R, Moxon ER, Masignani V. Structure and dynamics of the pan-genome of S*treptococcus pneumoniae* and closely related species. Genome Biol. 2010; 11:R107. [PubMed: 21034474]

Dorrell N, Mangan JA, Laing KG, Hinds J, Linton D, Al-Ghusein H, Barrell BG, Parkhill J, Stoker NG, Karlyshev AV, Butcher PD, Wren BW. Whole genome comparison of *Campylobacter jejuni* human isolates using a low-cost microarray reveals extensive genetic diversity. Genome Res. 2001; 11:1706–1715. [PubMed: 11591647]

Drummond A, Ashton B, Cheung M, Heled J, Kearse M, Moir R, Stones-Havas S, Thierer T, Wilson A. Geneious (5.1.2 ed.). 2010

Dykhuizen DE, Green L. Recombination in *Escherichia coli* and the definition of biological species. J. Bacteriol. 1991; 173:7257–7268. [PubMed: 1938920]

Ewing GB, Ebersberger I, Schmidt HA, von Haeseler A. Rooted triple consensus and anomalous gene trees. BMC Evol Biol. 2008; 8:118. [PubMed: 18439266]

Fitz-Gibbon ST, House CH. Whole genome-based phylogenetic analysis of free-living microorganisms. Nucleic Acids Res. 1999; 27:4218–4222. [PubMed: 10518613]

Friedman, C.; Neimann, J.; Wegener, H.; Tauxe, R. Campylobacter. 2nd ed.. Washington, DC: ASM Press; 2000. Epidemiology of *Campylobacter jejuni* infections in the United States and other industrialized nations; p. 121-138.

Funakoshi K, Koga M, Takahashi M, Hirata K, Yuki N. *Campylobacter coli* enteritis and Guillain-Barre syndrome: no evidence of molecular mimicry and serological relationship. J Neurol Sci. 2006; 246:163–168. [PubMed: 16600306]

Gallay A, Simon F, Megraud F. Surveillance of human *Campylobacter* infections in France--part 1--which data? A study of microbiological laboratories, 2000. Euro Surveill. 2003; 8:213–217. [PubMed: 14684881]

Gilbert M, Brisson JR, Karwaski MF, Michniewicz J, Cunningham AM, Wu Y, Young NM, Wakarchuk WW. Biosynthesis of ganglioside mimics in *Campylobacter jejuni* OH4384. Identification of the glycosyltransferase genes, enzymatic synthesis of model compounds, and characterization of nanomole amounts by 600-mhz (1)h and (13)c NMR analysis. J Biol Chem. 2000; 275:3896–3906. [PubMed: 10660542]

Gilbert, M.; Godschalk, PC.; Parker, C.; Endtz, H.; Wakarchuk, WW. Genetic basis for the variation in the lipooligosaccharide outer core of *Campylobacter jejuni* and possible association of glycosyltransferase genes with post-infectious neuropathies. In: Ketley, JM.; Konkel, ME., editors. *Campylobacter* : molecular and cellular biology. Norfolk: Horizon Bioscience; 2004.

Gilbert M, Karwaski MF, Bernatchez S, Young NM, Taboada E, Michniewicz J, Cunningham AM, Wakarchuk WW. The genetic bases for the variation in the lipo-oligosaccharide of the mucosal pathogen, *Campylobacter jejuni* Biosynthesis of sialylated ganglioside mimics in the core oligosaccharide. J Biol Chem. 2002; 277:327–337. [PubMed: 11689567]

Goddard EA, Lastovica AJ, Argent AC. *Campylobacter* 0:41 isolation in Guillain-Barre syndrome. Arch Dis Child. 1997; 76:526–528. [PubMed: 9245852]

Godschalk PC, Heikema AP, Gilbert M, Komagamine T, Ang CW, Glerum J, Brochu D, Li J, Yuki N, Jacobs BC, van Belkum A, Endtz HP. The crucial role of *Campylobacter jejuni* genes in anti-ganglioside antibody induction in Guillain-Barre syndrome. J Clin Invest. 2004; 114:1659–1665. [PubMed: 15578098]

Goossens H, Vlaes L, Butzler JP, Adnet A, Hanicq P, N'Jufom S, Massart D, de Schrijver G, Blomme W. *Campylobacter upsaliensis* enteritis associated with canine infections. Lancet. 1991; 337:1486–1487. [PubMed: 1675356]

Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J, Conesa A. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 2008; 36:3420–3435. [PubMed: 18445632]

Guerry P, Szymanski CM. *Campylobacter* sugars sticking out. Trends Microbiol. 2008; 16:428–435. [PubMed: 18707886]

Guerry P, Szymanski CM, Prendergast MM, Hickey TE, Ewing CP, Pattarini DL, Moran AP. Phase variation of *Campylobacter jejuni* 81–176 lipooligosaccharide affects ganglioside mimicry and invasiveness in vitro. Infect. Immun. 2002; 70:787–793. [PubMed: 11796612]

Gurgan T, Diker KS. Abortion associated with *Campylobacter upsaliensis*. J Clin Microbiol. 1994; 32:3093–3094. [PubMed: 7883913]

Ho TW, Hsieh ST, Nachamkin I, Willison HJ, Sheikh K, Kiehlbauch J, Flanigan K, McArthur JC, Cornblath DR, McKhann GM, Griffin JW. Motor nerve terminal degeneration provides a potential mechanism for rapid recovery in acute motor axonal neuropathy after *Campylobacter* infection. Neurology. 1997; 48:717–724. [PubMed: 9065554]

Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 2006; 23:254–267. [PubMed: 16221896]

Jakobsen IB, Easteal S. A program for calculating and displaying compatibility matrices as an aid in determining reticulate evolution in molecular sequences. Comput. Appl. Biosci. 1996; 12:291–295. [PubMed: 8902355]

Jones MA, Marston KL, Woodall CA, Maskell DJ, Linton D, Karlyshev AV, Dorrell N, Wren BW, Barrow PA. Adaptation of *Campylobacter jejuni* NCTC11168 to high-level colonization of the avian gastrointestinal tract. Infect Immun. 2004; 72:3769–3776. [PubMed: 15213117]

Karlyshev AV, Champion OL, Churcher C, Brisson JR, Jarrell HC, Gilbert M, Brochu D, St Michael F, Li J, Wakarchuk WW, Goodhead I, Sanders M, Stevens K, White B, Parkhill J, Wren BW, Szymanski CM. Analysis of *Campylobacter jejuni* capsular loci reveals multiple mechanisms for the generation of structural diversity and the ability to form complex heptoses. Mol Microbiol. 2005; 55:90–103. [PubMed: 15612919]

Karlyshev AV, Linton D, Gregson NA, Lastovica AJ, Wren BW. Genetic and biochemical evidence of a *Campylobacter jejuni* capsular polysaccharide that accounts for Penner serotype specificity. Mol Microbiol. 2000; 35:529–541. [PubMed: 10672176]

Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002; 30:3059–3066. [PubMed: 12136088]

Ketley JM. Pathogenesis of enteric infection by *Campylobacter*. Microbiology. 1997; 143(Pt 1):5–21. [PubMed: 9025274]

Koene MG, Houwers DJ, Dijkstra JR, Duim B, Wagenaar JA. Strain variation within *Campylobacter* species in fecal samples from dogs and cats. Vet Microbiol. 2009; 133:199–205. [PubMed: 18678447]

Kunin V, Ouzounis CA. The balance of driving forces during genome evolution in prokaryotes. Genome Res. 2003; 13:1589–1594. [PubMed: 12840037]

Lan R, Reeves PR. Intraspecies variation in bacterial genomes: the need for a species genome concept. Trends Microbiol. 2000; 8:396–401. [PubMed: 10989306]

Lan R, Reeves PR. When does a clone deserve a name? A perspective on bacterial species based on population genetics. Trends Microbiol. 2001; 9:419–424. [PubMed: 11553453]

Lang P, Lefebure T, Wang W, Pavinski Bitar P, Meinersmann RJ, Kaya K, Stanhope MJ. Expanded multilocus sequence typing and comparative genomic hybridization of *Campylobacter coli* isolates from multiple hosts. Appl Environ Microbiol. 2010; 76:1913–1925. [PubMed: 20097815]

Lastovica AJ, Le Roux E. Efficient isolation of *Campylobacter upsaliensis* from stools. J Clin Microbiol. 2001; 39:4222–4223. [PubMed: 11712518]

Lawson AJ, On SL, Logan JM, Stanley J. *Campylobacter hominis* sp. nov., from the human gastrointestinal tract. Int J Syst Evol Microbiol. 2001; 51:651–660. [PubMed: 11321111]

Lefebure T, Bitar PD, Suzuki H, Stanhope MJ. Evolutionary dynamics of complete *Campylobacter* pan-genomes and the bacterial species concept. Genome Biol Evol. 2010; 2:646–655. [PubMed: 20688752]

Lefebure T, Stanhope MJ. Pervasive, genome-wide positive selection leading to functional divergence in the bacterial genus *Campylobacter*. Genome Res. 2009; 19:1224–1232. [PubMed: 19304960]

Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS, Schlegel ML, Tucker TA, Schrenzel MD, Knight R, Gordon JI. Evolution of mammals and their gut microbes. Science. 2008; 320:1647–1651. [PubMed: 18497261]

Li L, Stoeckert CJ Jr, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 2003; 13:2178–2189. [PubMed: 12952885]

Maynard Smith J. Analyzing the mosaic structure of genes. J. Mol. Evol. 1992; 34:126–129. [PubMed: 1556748]

McSweegan E, Walker RI. Identification and characterization of two *Campylobacter jejuni* adhesins for cellular and mucous substrates. Infect. Immun. 1986; 53:141–148. [PubMed: 2873103]

Miller WG, Englen MD, Kathariou S, Wesley IV, Wang G, Pittenger-Alley L, Siletz RM, Muraoka W, Fedorka-Cray PJ, Mandrell RE. Identification of host-associated alleles by multilocus sequence typing of *Campylobacter coli* strains from food animals. Microbiology. 2006; 152:245–255. [PubMed: 16385134]

Moore JE, Corcoran D, Dooley JS, Fanning S, Lucey B, Matsuda M, McDowell DA, Megraud F, Millar BC, O'Mahony R, O'Riordan L, O'Rourke M, Rao JR, Rooney PJ, Sails A, Whyte P. Campylobacter. Vet Res. 2005; 36:351–382. [PubMed: 15845230]

Muegge BD, Kuczynski J, Knights D, Clemente JC, Gonzalez A, Fontana L, Henrissat B, Knight R, Gordon JI. Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. Science. 2011; 332:970–974. [PubMed: 21596990]

Parker CT, Gilbert M, Yuki N, Endtz HP, Mandrell RE. Characterization of lipooligosaccharide-biosynthetic loci of *Campylobacter jejuni* reveals new lipooligosaccharide classes: evidence of mosaic organizations. J Bacteriol. 2008; 190:5681–5689. [PubMed: 18556784]

Parker CT, Horn ST, Gilbert M, Miller WG, Woodward DL, Mandrell RE. Comparison of *Campylobacter jejuni* lipooligosaccharide biosynthesis loci from a variety of sources. J Clin Microbiol. 2005; 43:2771–2781. [PubMed: 15956396]

Parkhill J, Wren BW, Mungall K, Ketley JM, Churcher C, Basham D, Chillingworth T, Davies RM, Feltwell T, Holroyd S, Jagels K, Karlyshev AV, Moule S, Pallen MJ, Penn CW, Quail MA, Rajandream MA, Rutherford KM, van Vliet AH, Whitehead S, Barrell BG. The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. Nature. 2000; 403:665–668. [PubMed: 10688204]

Parsons BN, Porter CJ, Ryvar R, Stavisky J, Williams NJ, Pinchbeck GL, Birtles RJ, Christley RM, German AJ, Radford AD, Hart CA, Gaskell RM, Dawson S. Prevalence of *Campylobacter* spp. in a cross-sectional study of dogs attending veterinary practices in the UK and risk indicators associated with shedding. Vet J. 2010; 184:66–70. [PubMed: 19223212]

Pearson BM, Pin C, Wright J, I'Anson K, Humphrey T, Wells JM. Comparative genome analysis of *Campylobacter jejuni* using whole genome DNA microarrays. FEBS Lett. 2003; 554:224–230. [PubMed: 14596944]

Phongsisay V, Perera VN, Fry BN. Exchange of lipooligosaccharide synthesis genes creates potential Guillain-Barre syndrome-inducible strains of *Campylobacter jejuni*. Infect Immun. 2006; 74:1368–1372. [PubMed: 16428786]

Poly F, Serichatalergs O, Schulman M, Ju J, Cates CN, Kanipes M, Mason C, Guerry P. Discrimination of major capsular types of *Campylobacter jejuni* by multiplex PCR. J. Clin. Microbiol. 2011; 49:1750–1757. [PubMed: 21411576]

Roberts IS. The biochemistry and genetics of capsular polysaccharide production in bacteria. Annu Rev Microbiol. 1996; 50:285–315. [PubMed: 8905082]

Roshan U, Livesay DR. Probalign: multiple sequence alignment using partition function posterior probabilities. Bioinformatics. 2006; 22:2715–2721. [PubMed: 16954142]

Sheppard SK, Dallas JF, Wilson DJ, Strachan NJ, McCarthy ND, Jolley KA, Colles FM, Rotariu O, Ogden ID, Forbes KJ, Maiden MC. Evolution of an agriculture-associated disease causing *Campylobacter coli* clade: evidence from national surveillance data in Scotland. PLoS One. 2010; 5:e15708. [PubMed: 21179537]

Sheppard SK, McCarthy ND, Falush D, Maiden MC. Convergence of *Campylobacter* species: implications for bacterial evolution. Science. 2008; 320:237–239. [PubMed: 18403712]

Snel B, Bork P, Huynen MA. Genome phylogeny based on gene content. Nat. Genet. 1999; 21:108–110. [PubMed: 9916801]

Sopwith W, Birtles A, Matthews M, Fox A, Gee S, James S, Kempster J, Painter M, Edwards-Jones V, Osborn K, Regan M, Syed Q, Bolton E. Investigation of food and environmental exposures relating to the epidemiology of *Campylobacter coli* in humans in Northwest England. Appl. Environ. Microbiol. 2010; 76:129–135. [PubMed: 19854914]

Strimmer K, vonHaeseler A. Quartet puzzling: A quartet maximum-likelihood method for reconstructing tree topologies. Mol. Biol. Evol. 1996; 13:964–969.

Takkinen J, Ammon A, Robstad O, Breuer T. European survey on *Campylobacter* surveillance and diagnosis 2001. Euro Surveill. 2003; 8:207–213. [PubMed: 14684880]

Tam CC, O'Brien SJ, Adak GK, Meakins SM, Frost JA. *Campylobacter coli* - an important foodborne pathogen. J Infect. 2003; 47:28–32. [PubMed: 12850159]

Tekaia F, Lazcano A, Dujon B. The genomic tree as revealed from whole proteome comparisons. Genome Res. 1999; 9:550–557. [PubMed: 10400922]

Unicomb L, Ferguson J, Riley TV, Collignon P. Fluoroquinolone resistance in *Campylobacter* absent from isolates, Australia. Emerg Infect Dis. 2003; 9:1482–1483. [PubMed: 14718099]

Unicomb LE, Ferguson J, Stafford RJ, Ashbolt R, Kirk MD, Becker NG, Patel MS, Gilbert GL, Valcanis M, Mickan L. Low-level fluoroquinolone resistance among *Campylobacter jejuni* isolates in Australia. Clin Infect Dis. 2006; 42:1368–1374. [PubMed: 16619147]

van Belkum A, Jacobs B, van Beek E, Louwen R, van Rijs W, Debruyne L, Gilbert M, Li J, Jansz A, Megraud F, Endtz H. Can *Campylobacter coli* induce Guillain-Barre syndrome? Eur J Clin Microbiol Infect Dis. 2009; 28:557–560. [PubMed: 19002726]

Wassenaar TM, Blaser MJ. Pathophysiology of *Campylobacter jejuni* infections of humans. Microbes Infect. 1999; 1:1023–1033. [PubMed: 10617934]

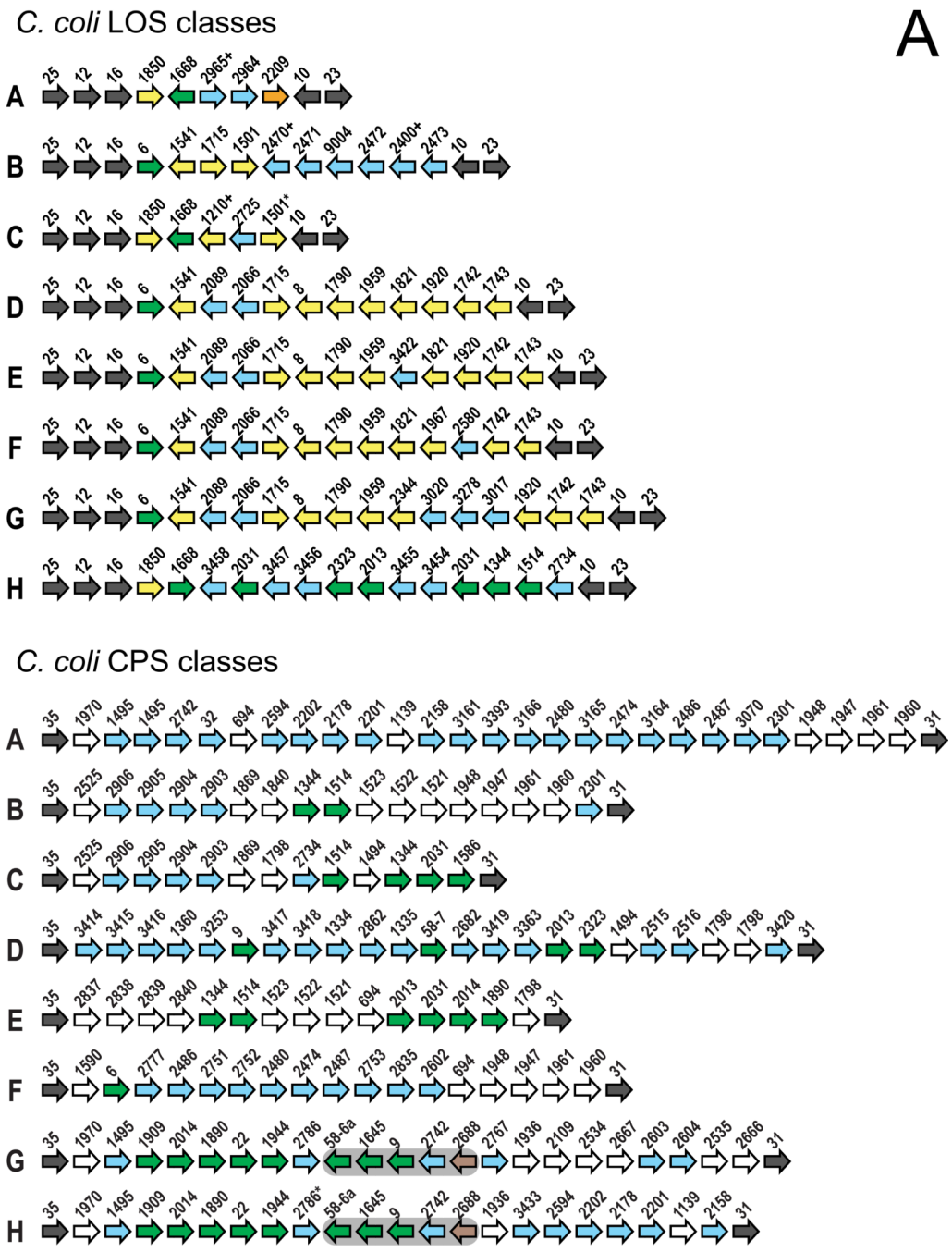
Westgarth C, Porter CJ, Nicolson L, Birtles RJ, Williams NJ, Hart CA, Pinchbeck GL, Gaskell RM, Christley RM, Dawson S. Risk factors for the carriage of *Campylobacter upsaliensis* by dogs in a community in Cheshire. Vet Rec. 2009; 165:526–530. [PubMed: 19880860]

Willison HJ, O'Hanlon GM. The immunopathogenesis of Miller Fisher syndrome. J Neuroimmunol. 1999; 100:3–12. [PubMed: 10695710]

Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008; 18:821–829. [PubMed: 18349386]

- High *C. coli* LOS/CPS gene diversity: each cluster delineated into eight gene content classes
- Diversity predominantly due to extensive gene gain/loss both within and between species and also between the LOS and CPS
- Additional mechanisms: phase-variation, gene duplication/inactivation, selection pressure
- Sialic acid biosynthesis homologs detected in *C. coli* and *C. upsaliensis*
- *C. coli* LOS classes differentiated between bovine and poultry hosts

A



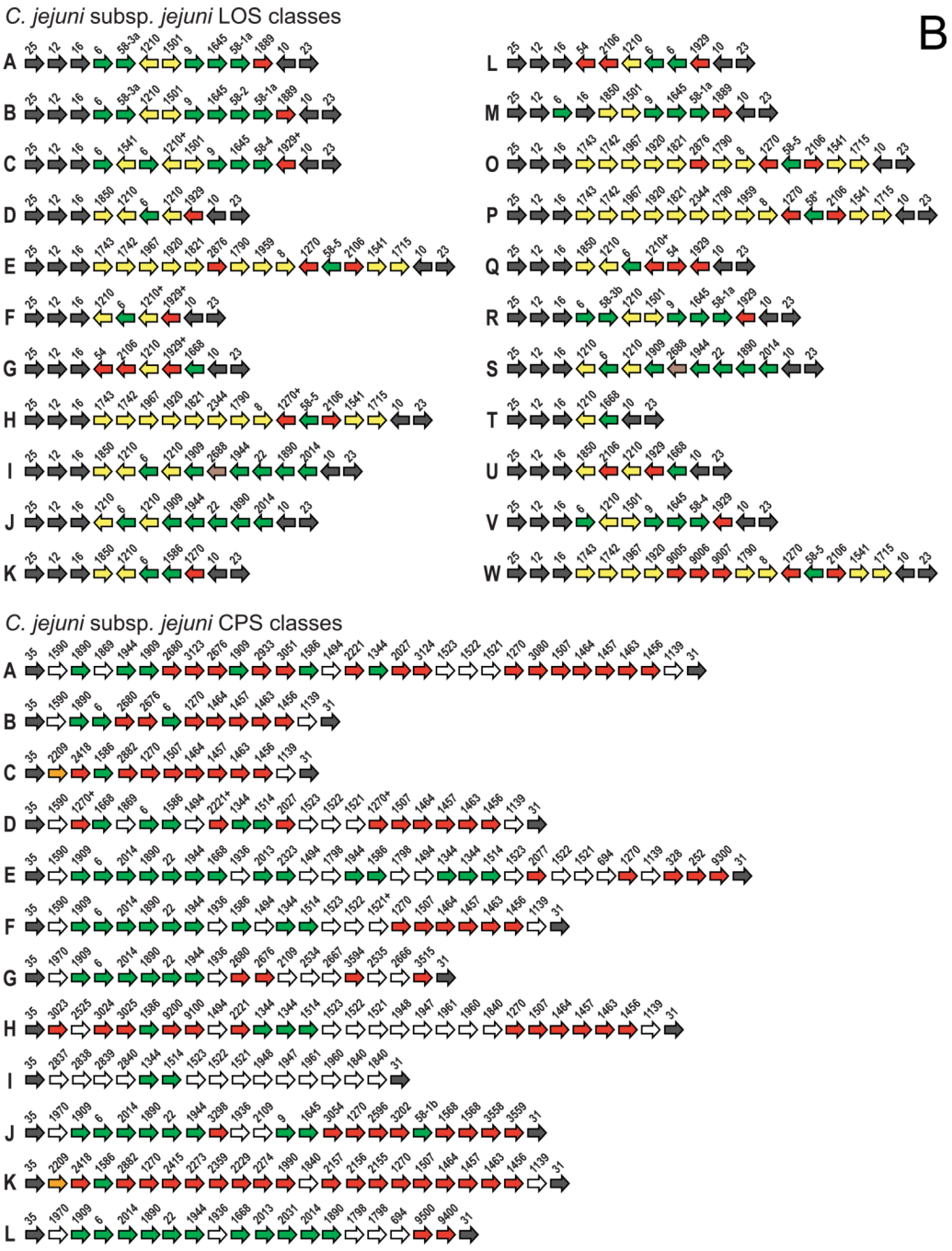


Fig. 1.

(A and B). *C. coli* and *C. jejuni* subsp. *jejuni* ortholog content for the LOS and CPS gene clusters. Arrows represent orthologs, and numbers are ortholog ID numbers (see Table S1 in the supplementary material for associated annotations). Grey arrows: orthologs belonging to the core genome (i.e. orthologs seen in all strains), with all remaining arrows representing orthologs belonging to the dispensable genome. Yellow arrows: orthologs present in the LOS of both species. White arrows: orthologs present in the CPS of both species. Green arrows: orthologs present in at least three of the following four possible locations (i) *C. coli* LOS, (ii) *C. coli* CPS, (iii) *C. jejuni* subsp. *jejuni* LOS, (iv) *C. jejuni* subsp. *jejuni* CPS. Only

ortholog 6 was present in all four locations. Blue arrows: *C. coli* orthologs not present in the LOS or CPS of *C. jejuni* subsp. *jejuni*. Red arrows: *C. jejuni* subsp. *jejuni* orthologs not present in the LOS or CPS of *C. coli*. Orange arrows: orthologs present in the LOS of *C. coli* and the CPS of *C. jejuni* subsp. *jejuni*. Brown arrows: orthologs present in the CPS of *C. coli* and the LOS of *C. jejuni* subsp. *jejuni*. For reference, names of the first and last genes within each cluster and associated ortholog IDs were as follows: 25=*waaC*, 23=*waaF*, 35=*kpsF*, 31=*kpsC*. CDS for orthologs marked with (i) an asterisk were fragmented (see Table S1 in the supplementary material for details), and (ii) a plus sign showed evidence of phase variation. Grey shading highlights sialic acid gene cassette discussed in text. Note: *C. jejuni* subsp. *doylei* was omitted from the figure due to space considerations. However, gene content information for this species can be found in Table S1 in the supplementary material.

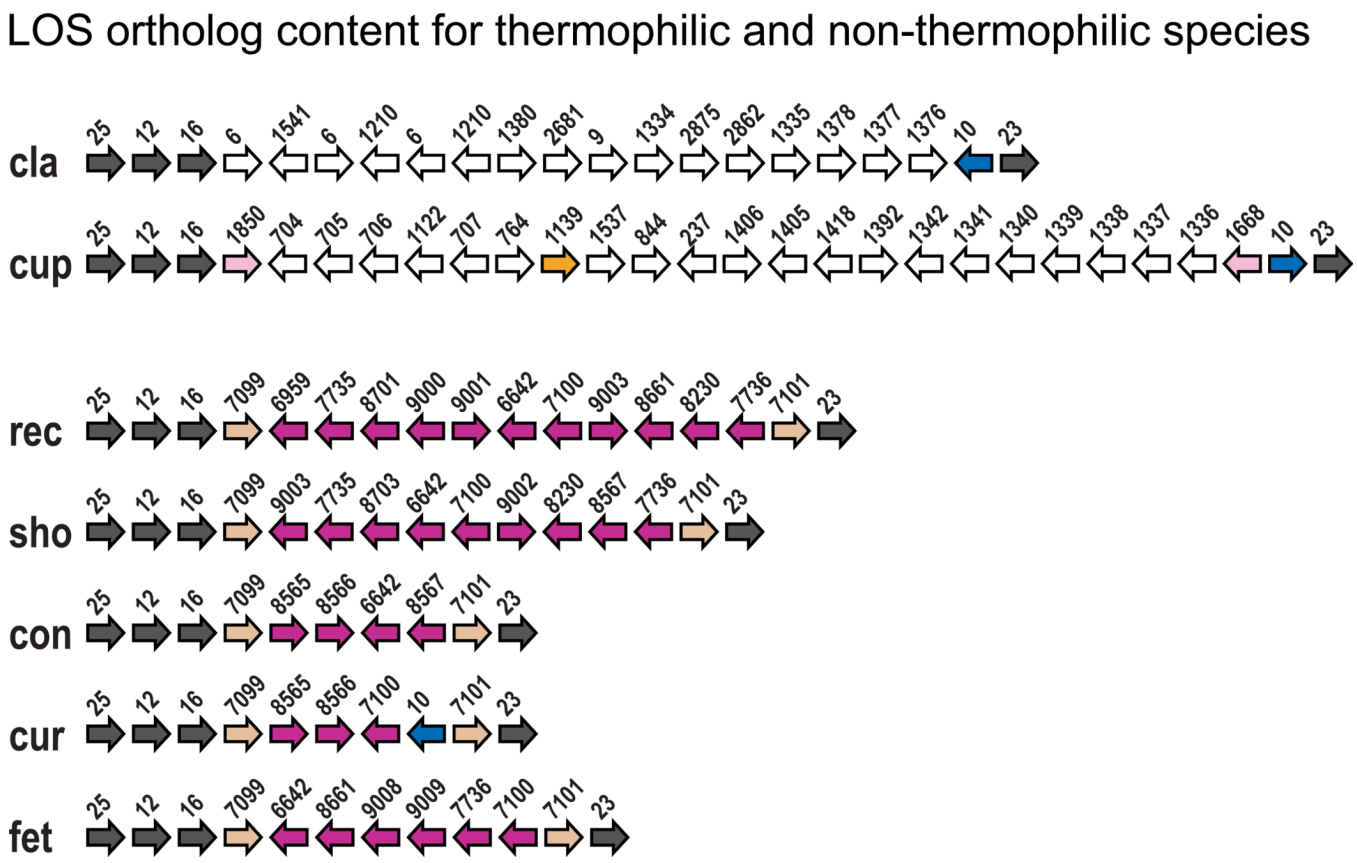


Fig. 2.

Ortholog content for the LOS gene cluster for thermophilic and non-thermophilic species other than *C. jejuni* subsp. *jejuni* and *C. coli*. Species abbreviations are as follows: cla=*C. lari*, cup=*C. upsaliensis*, rec=*C. rectus*, sho=*C. showae*, con=*C. concisus*, cur=*C. curvus*, fet=*C. fetus* subsp. *fetus*. Arrows represent orthologs, and numbers are ortholog ID numbers (see Table S1 for associated annotations). Grey arrows: with the exception of *C. hominis* and *C. gracilis* (see text for discussion), orthologs seen in all species. Blue arrows: ortholog is seen in all thermophilic species and two non-thermophilic species (*C. curvus* and *C. hominis*). Brown arrows: orthologs seen in all non-thermophilic species except *C. hominis* and *C. gracilis.* Purple arrows: dispensable non-thermophilic orthologs, none of which are seen in the thermophilic species. For *C. upsaliensis*, the pink and orange arrows show three dispensable orthologs shared with other thermophilic species. Pink orthologs occurred in the LOS of other species and the orange ortholog occurred in the CPS of other species. For reference, ortholog 25=*waaC* and ortholog 23=*waaF*.

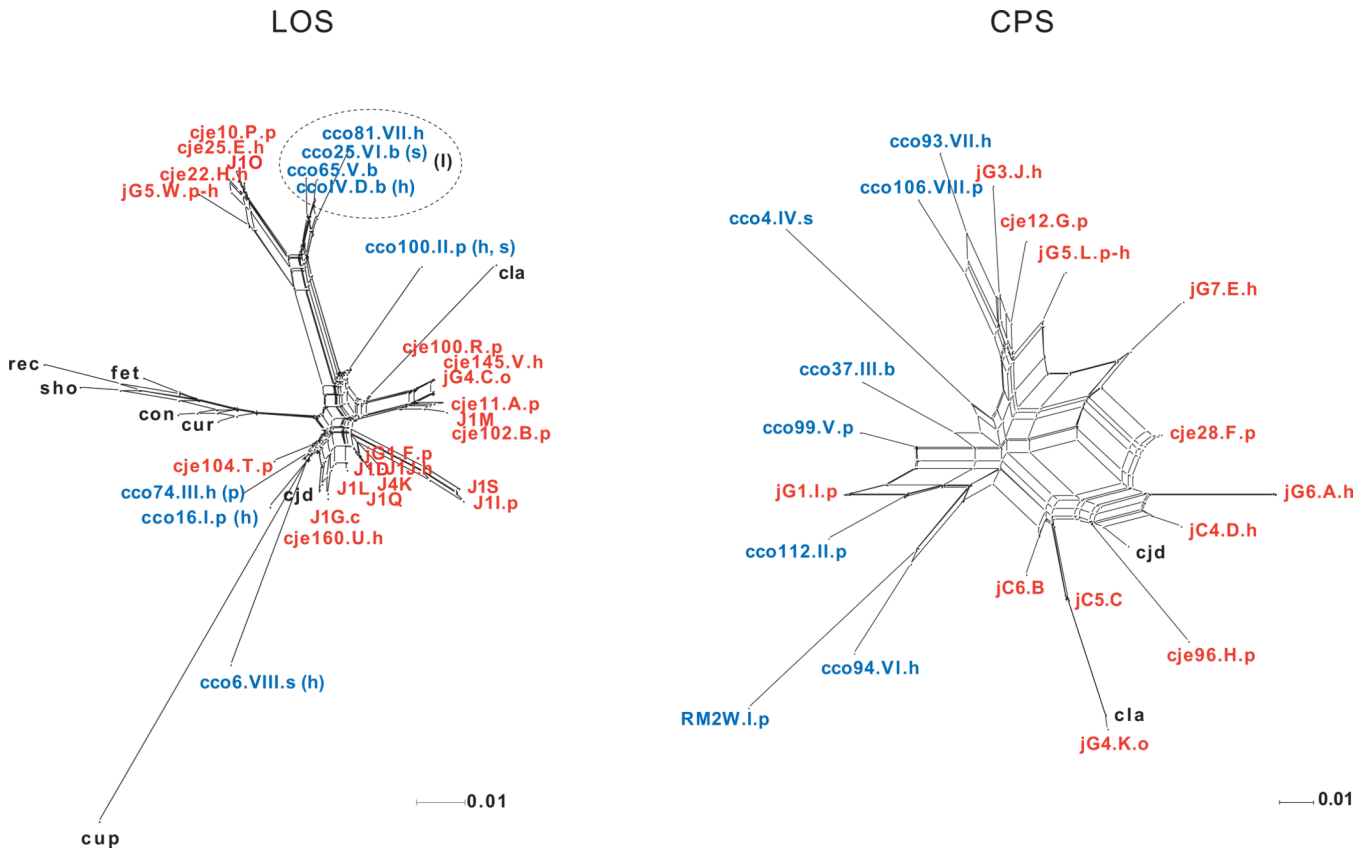


Fig. 3.

Split networks based on presence/absence of orthologs depicting gene content similarities among LOS and CPS gene cluster classes. Networks contain single representative *C. coli* and *C. jejuni* subsp. *jejuni* strains for each class (class IDs are shown in upper case following strain ID). *C. coli* strains are shown in blue and *C. jejuni* subsp. *jejuni* strains in red. Remaining taxa codes are as follows: cjd=*C. jejuni* subsp. *doylei*, cla=*C. lari*, cup=*C. upsaliensis*, rec=*C. rectus*, sho=*C. showae*, con=*C. concisus*, cur=*C. curvus*, fet=*C. fetus* subsp. *fetus* (also see Table 1 for additional strain information). Where present, lower case letters following class IDs for *C. coli* and *C. jejuni* subsp. *jejuni* show strain isolation source: h=human, p=poultry, b=bovine, s=swine, o=ovine, c=caprine. Many classes were isolated from multiple sources. For *C. coli* additional sources for each LOS class are shown in parentheses (see Table 1 for complete source distribution). The cluster of *C. coli* LOS classes (I) restricted to human, bovine, and swine sources is indicated with a dashed oval.

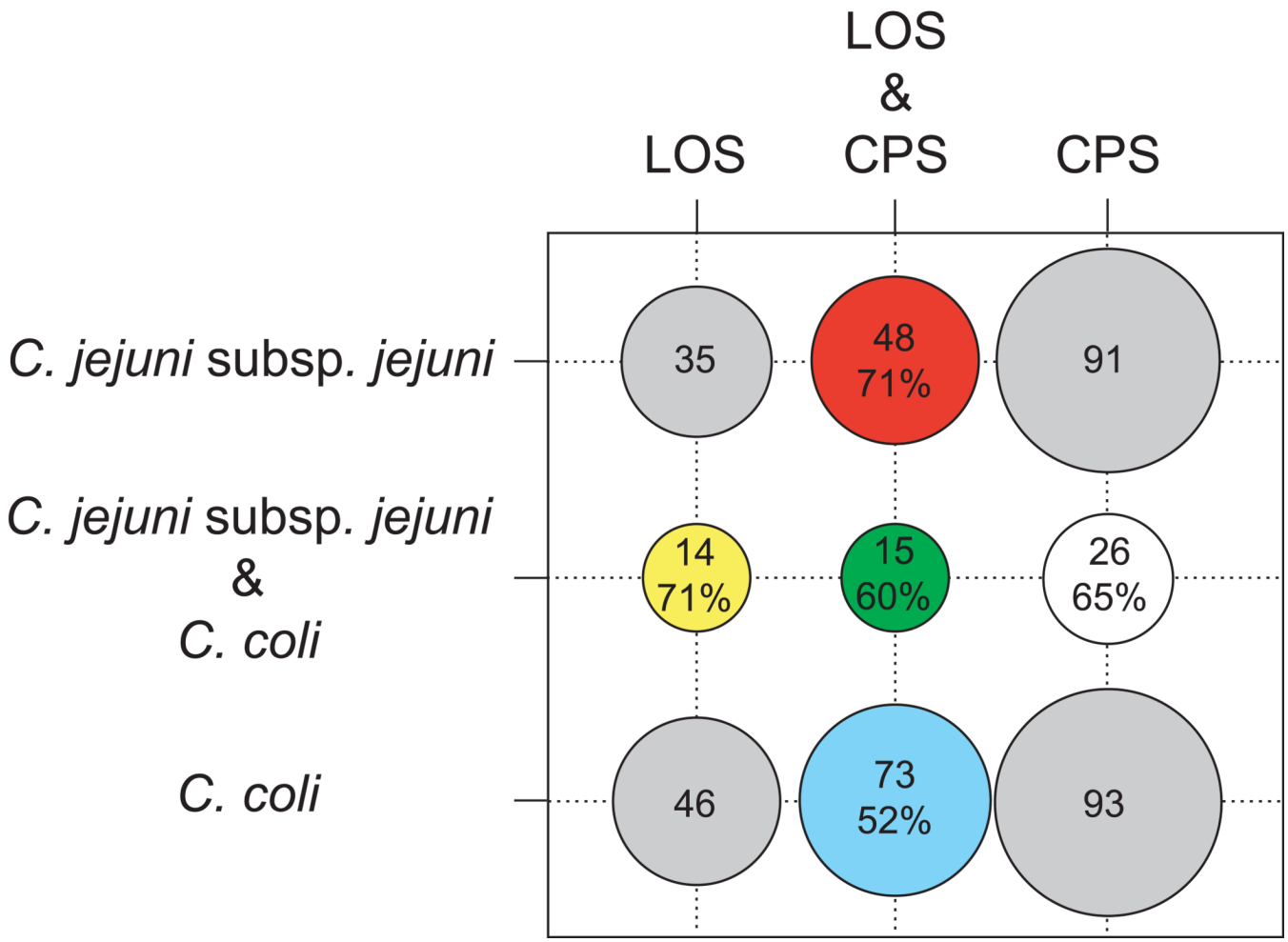
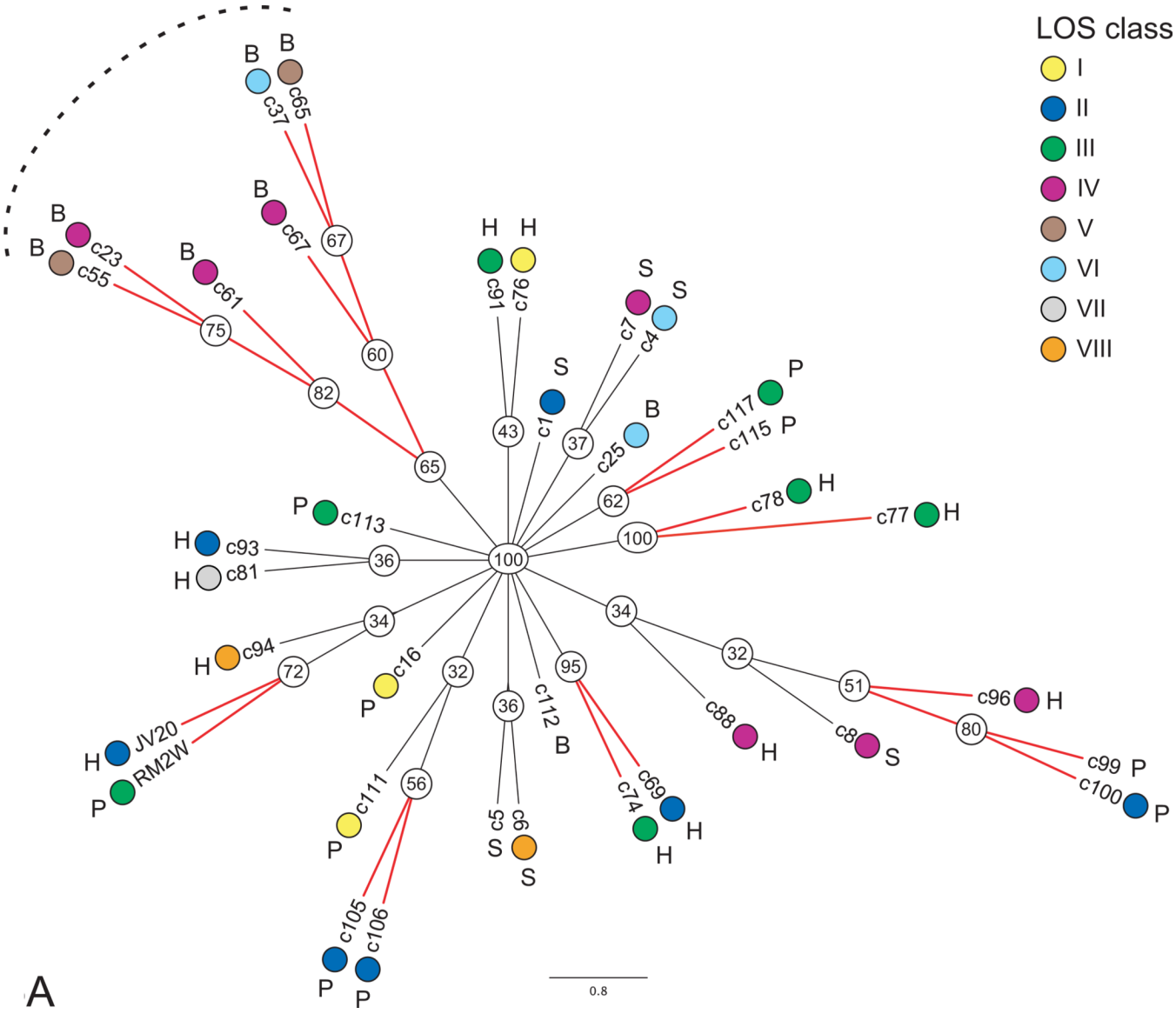


Fig. 4.

Distribution and overlap of the dispensable orthologs present in the LOS and CPS of *C. coli* and *C. jejuni* subsp. *jejuni* (the five core LOS and two core CPS orthologs are excluded). Circles shaded grey show the number of distinct orthologs observed in each gene cluster for each species. Circles shaded red, yellow, green, white, and blue correspond to the arrows (orthologs) in Fig.1 and show the number of distinct orthologs in each category (see Fig. 1 caption for a description of the categories). The proportion of orthologs in each of these categories showing evidence of recombination is also shown.



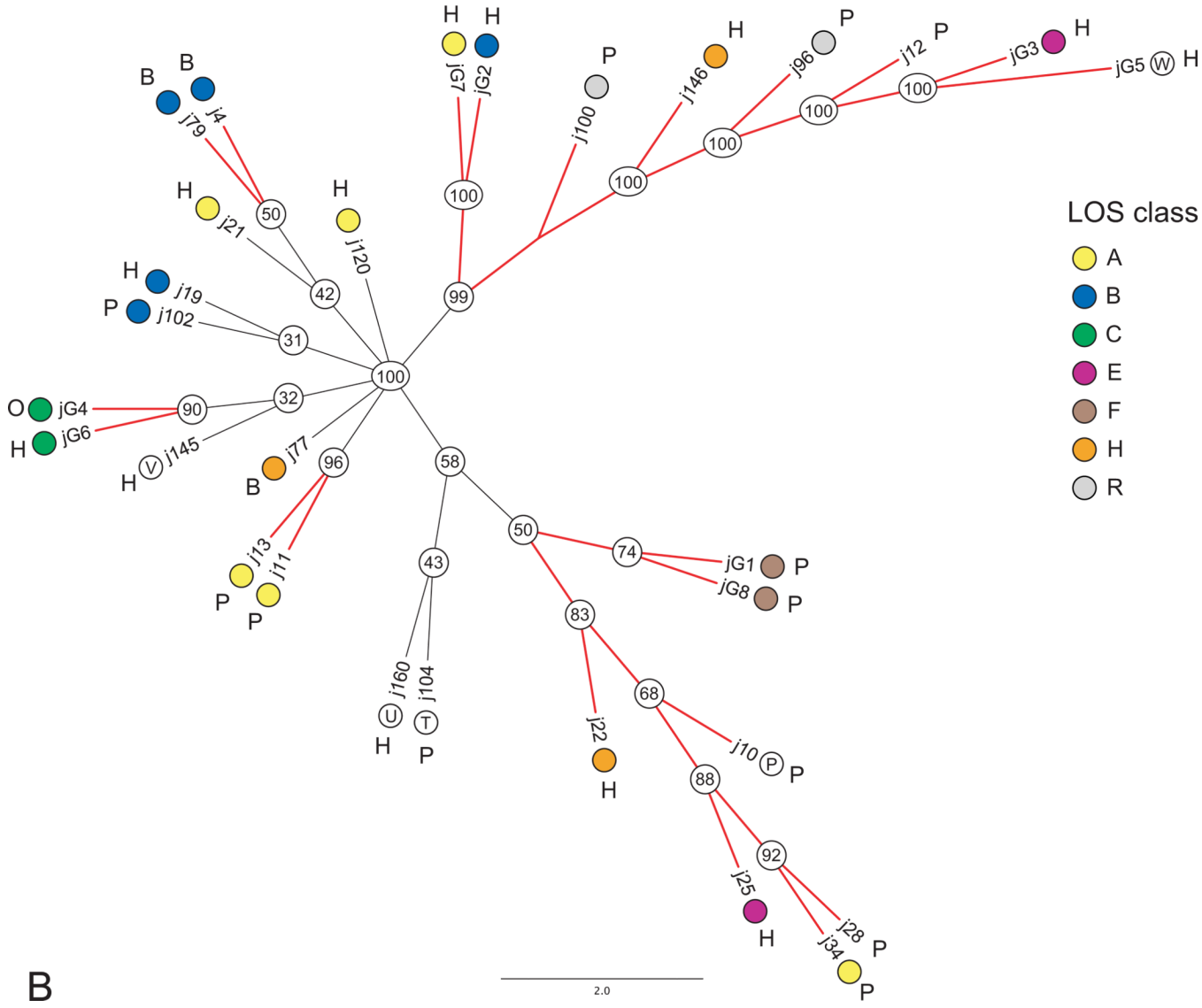


Fig. 5.

(A and B). Phylogenies showing the consensus of 399 and 206 gene-trees for *C. coli* (A) and *C. jejuni* subsp. *jejuni* (B) respectively. Numbers in circles show the proportion of gene-trees that supported a particular grouping. Branches for groupings with greater than 50% support are shown in red. Colored circles identify the LOS class possessed by a strain. Strain isolation source is shown with a single letter code (H=human, B=bovine, S=swine, and P=poultry). A well supported grouping of bovine sourced isolates for *C. coli* is indicated with a dashed line.

Table 1

A. *C. coli* strain information

Strain	Strain ID	Source	ST	LOS	CPS	Accession#
LMG 23336	cco76	Human	3868	I	-	AINM00000000
2680	cco111	Turkey	3872	I	-	AIMN00000000
2692	cco115	Turkey	860	I	-	AIMQ00000000
86119	cco16	Chicken	825	I	-	AIMU00000000
202/04	cco69	Human	1585	II	-	AINH00000000
Z163	cco100	Chicken	3336	II	V	AIMK00000000
H8	cco93	Human	901	II	VII	AINU00000000
111-3	cco1	Swine	1467	II		AIMI00000000
2548	cco105	Turkey	1167	II	VI	AIML00000000
2553	cco106	Turkey	825	II	VIII	AIMM00000000
JV20	JV20	Human	860	II	-	AEER01000001
LMG 23341	cco77	Human	855	III	-	AINN00000000
37/05	cco74	Human	1191	III	-	AINK00000000
LMG 23342	cco78	Human	855	III	-	AINO00000000
H6	cco91	Human	3020	III	-	AINT00000000
2688	cco113	Turkey	1017	III	III	AIMP00000000
2698	cco117	Turkey	829	III	-	AIMR00000000
RM2228	RM2W	Chicken	1063	III	I	AAFL01000001
1948	cco61	Bovine	1104	IV	-	AINE00000000
1961	cco67	Bovine	1104	IV	-	AING00000000
LMG 9860	cco88	Human	900	IV	-	AINS00000000
H56	cco96	Human	1096	IV	-	AINW00000000
1098	cco23	Bovine	1104	IV	-	AIMW00000000
67-8	cco7	Swine	1061	IV	IV	AINI00000000
151-9	cco8	Swine	1102	IV	-	AINQ00000000
1909	cco55	Bovine	1104	V	-	AINC00000000
1957	cco65	Bovine	2698	V	-	AINF00000000
1148	cco25	Bovine	1068	VI	-	AIMX00000000

A. *C. coli* strain information

Strain	Strain ID	Source	ST	LOS	CPS	Accession#
1417	cco37	Bovine	1436	VI	III	AIMY00000000
7--1	cco4	Swine	3860	VI	IV	AIMZ00000000
LMG 9853	cco81	Human	3869	VII	-	AINR00000000
H9	cco94	Human	825	VIII	VI	AINV00000000
59-2	cco6	Swine	890	VIII	-	AIND00000000
Z156	cco99	Chicken	854	-	V	AINX00000000
2685	cco112	Turkey	1082	-	II	AIMO00000000
132-6	cco5	Swine	3861	-	III	AINA00000000

B. *C. jejuni* subsp. *jejuni* and *C. jejuni* subsp. *doylei* strain information

Strain	Strain ID	Source	ST	LOS	CPS	Accession#
60004	cje11	Chicken	4836	A	-	AIOE00000000
1997-7	cje21	Human	93	A	-	AIOX00000000
87459	cje34	Chicken	6224	A	-	AIPE00000000
86605	cje13	Chicken	6219	A	-	AIOJ00000000
LMG 9879	cje120	Human	47	A	-	AIOI00000000
ICDCCJ07001	jG7	Human	2993	A	E	CP002029
81–176	jG2	Human	604	B	D	CP000538
1997-4	cje19	Human	475	B	-	AIOW00000000
LMG 23218	cje102	Chicken	48	B	-	AIOB00000000
1893	cje79	Bovine	38	B	-	AIPK00000000
140-16	cje4	Bovine	6217	B	-	AIPF00000000
IA3902	jG4	Sheep	8	C	K*	CP001876
NCTC 11168 (RM1862)	jG6	Human	43	C	A	AL111168
LMG 23216	cje100	Chicken	4835	R	-	AIOA00000000
LMG 23211	cje96	Chicken	220	R	H*	AIPO00000000
1997-14	cje25	Human	5159	E	F*	AIPA00000000
81116	jG3	Human	267	E	J*	CP000814
RM1221	jG1	Chicken	354	F	I*	CP000025

Infect Genet Evol. Author manuscript; available in PMC 2014 March 01.

B. *C. jejuni* subsp. *jejuni* and *C. jejuni* subsp. *doylei* strain information

Strain	Strain ID	Source	ST	LOS	CPS	Accession#
S3	jG8	Chicken	354	F	I*	CP001960
1997-10	cje22	Human	6222	H	-	AIOY00000000
1854	cje77	Bovine	922	H	-	AIPJ00000000
2008-894	cje146	Human	1962	H	-	AIOQ00000000
51494	cje10	Chicken	4834	P	-	AINZ00000000
LMG 23223	cje104	Chicken	791	T*	-	AIOC00000000
2008-979	cje160	Human	2274	U*	-	AIOU00000000
2008-1025	cje145	Human	50	V*	-	AIOP00000000
M1	jG5	Human	137	W*	L*	CP001900
51037	cje28	Chicken	939	-	F*	AIPB00000000
55037	cje12	Chicken	2223	-	G*	AIOH00000000
269-97 (C. j. doylei)	cjd	Human		N	M*	CP000768

Dash (-) refers to strains lacking contiguous sequence for a particular gene cluster.

Asterisk in Table 1B denotes new LOS or CPS classes. For one *C. coli* strain and five *C. jejuni* subsp. *jejuni* strains (omitted from table), none of the gene clusters could be assembled into contiguous sequence.

ST=multi locus sequence typing (MLST) sequence type.

Table 2

Strain ID, accession numbers, and gene cluster information for additional thermophilic and non-thermophilic *Campylobacter* species

Species	Strain	Species ID	Accession#
C. lari	RM2100	cla	CP000932
C. upsaliensis	JV21	cup1	AEPU01000001
C. upsaliensis	RM3195	cup2	AAFJ01000001
C. rectus *	RM3267	rec	ACFU01000001
C. showae *	RM3277	sho	ACVQ01000001
C. concisus *	13826	con	CP000792
C. curvus *	525.92	cur	CP000767
C. fetus subsp. *fetus* *	82-40	fet	CP000487
C. gracilis *	RM3268	gra	ACYG01000001
C. hominis *	ATCC BAA-381	hom	CP000776

All species were human sourced isolates.

Asterisk shows non-thermophilic species.

Table 3

C. jejuni subsp. *jejuni* strain information for previously sequenced CPS classes

Strain	Strain ID	Source	CPS	Accession#
NCTC 12517	jC6	human	B	BX545860
G1	jC5	human	C	BX545859
81–176	jC4	human	D	BX545858
176.83	jC3	human	E	BX545857
ATCC 43456	jC2	human	D	AY332624
CCUG 10954	jC1	undetermined	D	AY332625

Table 4

C. jejuni subsp. *jejuni* strain information for previously sequenced LOS classes

Strain	Strain ID	Source	LOS	Accession#
RM1048	J1A	human	A	AF215659
RM1556 (ATCC 43438)	J2A	human	A	AF400048
ATCC 43446	J3A	human	A	AF167344
OH4384	J4A	human	A	AF130984
RM1052	J1B	human	B	AF401528
RM1050	J2B	human	B	AF401529
RM1046	J1C	bovine	C	AF400047
RM1862 (NCTC 11168)	J2C	human	C	AL139077
RM1045	J3C	human	C	AY044156
LIO87	J1D	undetermined	D	AF400669
RM3418	J2D	undetermined	D	EU404109
RM1863 (81116)	J1E	human	E	AJ131360
RM1552	J2E	human	E	EU404105
GB15	J1F	human	F	AY423554
RM1170	J2F	chicken	F	AY434498
RM3415	J3F	undetermined	F	EU404108
RM1555 (ATCC 43437)	J1G	goat	G	AY436358
RM1047 (ATCC 43431)	J1H	human	H	AY800272
RM1553	J2H	human	H	EU404106
RM1850	J1I	chicken	I	EU404107
RM1508	J1J	human	J	EU404104
GB24	J1K	human	K	AY573819
RM2227	J2K	chicken	K	EF143353
RM2229	J3K	undetermined	K	EF143354
RM1861	J4K	undetermined	K	EU410350
RM3435	J1L	undetermined	L	EU404111
RM1503	J1M	undetermined	M	EF140720
RM2095	J1N	undetermined	N	AY816330
RM3423	J1O	undetermined	O	EF143352
GB4	J1P	human	P	AY943308
RM3437	J1Q	undetermined	Q	EU404112
GC149	J1R	undetermined	R	AY962325
RM3419	J1S	undetermined	S	EU404110
RM2095	J1N	undetermined	N	AY816330