Comparative Genomic Analysis Identifies a *Campylobacter* Clade Deficient in Selenium Metabolism

William G. Miller[1,*], Emma Yee[1], Bruno S. Lopes[2], Mary H. Chapman[1], Steven Huynh[1], James L. Bono[3], Craig T. Parker[1], Norval J.C. Strachan[2], and Ken J. Forbes[2]

[1]Produce Safety and Microbiology Research Unit, Agricultural Research Service, U.S. Department of Agriculture, Albany, CA

[2]School of Medicine, Medical Sciences and Nutrition, University of Aberdeen, United Kingdom

[3]Meat Safety and Quality Research Unit, Agricultural Research Service, U.S. Department of Agriculture, Clay Center, NE

*Corresponding author: E-mail: william.miller@ars.usda.gov.

Accepted: May 9, 2017

Data deposition: All genome sequencing data has been deposited in GenBank (accession numbers provided in table 1*A* and supplementary table S2, Supplementary Material online).

Abstract

The nonthermotolerant *Campylobacter* species *C. fetus*, *C. hyointestinalis*, *C. iguaniorum*, and *C. lanienae* form a distinct phylogenetic cluster within the genus. These species are primarily isolated from foraging (swine) or grazing (e.g., cattle, sheep) animals and cause sporadic and infrequent human illness. Previous typing studies identified three putative novel *C. lanienae*-related taxa, based on either MLST or *atpA* sequence data. To further characterize these putative novel taxa and the *C. fetus* group as a whole, 76 genomes were sequenced, either to completion or to draft level. These genomes represent 26 *C. lanienae* strains and 50 strains of the three novel taxa. *C. fetus*, *C. hyointestinalis* and *C. iguaniorum* genomes were previously sequenced to completion; therefore, a comparative genomic analysis across the entire *C. fetus* group was conducted (including average nucleotide identity analysis) that supports the initial identification of these three novel *Campylobacter* species. Furthermore, *C. lanienae* and the three putative novel species form a discrete clade within the *C. fetus* group, which we have termed the *C. lanienae* clade. This clade is distinguished from other members of the *C. fetus* group by a reduced genome size and distinct CRISPR/Cas systems. Moreover, there are two signature characteristics of the *C. lanienae* clade. *C. lanienae* clade genomes carry four to ten unlinked and similar, but nonidentical, flagellin genes. Additionally, all 76 *C. lanienae* clade genomes sequenced demonstrate a complete absence of genes related to selenium metabolism, including genes encoding the selenocysteine insertion machinery, selenoproteins, and the selenocysteinyl tRNA.

Key words: *Campylobacter*, *C. lanienae*, *C. fetus*, selenium, flagella.

Introduction

Campylobacter fetus (subspp. *fetus* and *venerealis*), the type species of the genus, is primarily isolated from cattle and sheep (Debruyne et al. 2008; On 2005). *C. fetus* subsp. *fetus* is isolated from fecal samples or the bovine/ovine intestinal tract and is an abortifacient in both cattle and sheep; however, *C. fetus* subsp. *venerealis* is more restricted to bovine/ovine venereal tissue and is the causal agent of bovine genital campylobacteriosis (Blaser et al. 2008; Debruyne et al. 2008; On 2005). Species related to *C. fetus* include: *C. hyointestinalis* (subspp. *hyointestinalis* [Gebhart et al. 1985] and *lawsonii* [On et al. 1995]) isolated from swine (Sasaki et al. 2013), cattle (Guevremont et al. 2014; Hakkinen et al. 2007; Inglis et al. 2004; Oporto and Hurtado 2011; Salihu et al. 2009; Serraino et al. 2013), and sheep (Oporto and Hurtado 2011); and *C. lanienae* (Logan et al. 2000), isolated from swine (Carbonero et al. 2014; Jay-Russell et al. 2012; Navarro-Gonzalez et al. 2014; Sasaki et al. 2013; Schweitzer et al. 2011), cattle (Inglis et al. 2004, 2003), sheep (Carbonero et al. 2014; Schweitzer et al. 2011), and chinchillas (Turowski et al. 2014). *C. fetus*-related taxa also include *C. fetus* subsp. *testudinum* (Fitzgerald et al. 2014) and *C. iguaniorum* (Gilbert et al. 2015), the only *Campylobacter* taxa isolated from reptiles. *C. fetus*-related organisms are primarily a veterinary or agricultural concern (Blaser et al. 2008), and infrequently or sporadically cause disease in humans (Blaser et al. 2008; Gorkiewicz et al. 2002; Kim do et al. 2015; Miller et al. 2012; Patrick et al. 2013; Samosornsuk et al. 2015).

Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution 2017. This work is written by US Government employees and is in the public domain in the US.

Phylogenetic analysis of 16S rDNA sequences clearly place *C. fetus* and the *C. fetus*-related taxa into a distinct cluster within the genus (see supplementary fig. S1, Supplementary Material online). Molecular typing methods suitable for typing members of the *C. fetus* group have been described. These include novel MLST typing methods (Miller et al. 2012; van Bergen et al. 2005) and an *atpA* typing method (Miller et al. 2014b). During the development of these typing methods it became clear that additional taxa may exist within the *C. fetus* group. MLST analysis identified two strains, RM6137 and NCTC 13003 (ST-4 and ST-1, respectively, in fig. 2 from Miller et al. [2012]), that were most closely related to, but distinguishable from, *C. lanienae*. RM6137 was isolated from a feral pig in California; NCTC 13003 was deposited as a *C. lanienae*, but no epidemiological data were available for this strain. *atpA* typing analysis also identified a *C. lanienae*-related cluster (fig. 3, cluster 4 in Miller et al. [2014b]) composed of eight strains. These strains were isolated in California during 2009–2010 from alpacas, goats and cattle. Additional MLST typing indicated that these eight strains were taxonomically related to, but distinct from, *C. lanienae*, RM6137 and NCTC 13003 (data not shown). Thus, the novel typing methods identified three putative novel *C. lanienae*-related taxa within the *C. fetus* group.

These data strongly suggested that these ten strains were representatives of three novel *Campylobacter* species. However, assignment of these strains to novel species were based on limited molecular data and a very small strain set (two putative taxa being represented by a single strain each). Nevertheless, additional strains within each of the three putative novel taxa began to be isolated from food animals, primarily swine. To clarify the taxonomic structure of the *C. fetus* group and further characterize these putative novel taxa, the genomes of five strains, including the *C. lanienae* type strain NCTC 13004 and representatives of all three putative novel taxa, were sequenced to completion; additional genomes were sequenced to draft level. The genomes of all other described taxa within the *C. fetus* group have been sequenced to completion (Gilbert et al. 2013, 2014; Miller et al. 2016a, 2016b; Oliveira et al. 2016; Stynen et al. 2011; van der Graaf-van Bloois et al. 2014a, 2014b); thus, a full comparative genomic analysis of the *C. fetus* group could be accomplished with the new genomic data.

In this study, we present the comparative analysis of 87 *C. fetus* group genomes: the genomes of eleven strains representing *C. fetus*, *C. hyointestinalis* and *C. iguaniorum* were sequenced previously; and the genomes of 76 additional strains representing *C. lanienae* (1 complete, 25 draft) or the three novel taxa (4 complete, 46 draft) were sequenced here. The genomic data presented in this study further support the initial identification of three novel taxa within the *C. fetus* group. Additionally, we demonstrate that together *C. lanienae* and the three novel taxa form a discrete clade within the *C. fetus* group with distinct genomic and phenotypic properties.

Materials and Methods

Reagents and Chemicals

PCR enzymes and reagents were purchased from New England Biolabs (Beverly, MA) or Epicentre (Madison, WI). All chemicals were purchased from Sigma-Aldrich Chemicals (St. Louis, MO) or Thermo Fisher Scientific (Waltham, MA). DNA sequencing reagents and capillaries were purchased from Applied Biosystems (Foster City, CA), Roche Life Science (Indianapolis, IN), Illumina Inc. (San Diego, CA) or Pacific Biosciences (Menlo Park, CA).

Isolation of *Campylobacter* Strains

California feral swine, goat, cow and alpaca fecal samples were collected from farms and ranches in the central California region (Cooley et al. 2013). A swab of fecal material was suspended in 6 ml 1× ABB (Anaerobe Basal Broth; Oxoid, Thermo Fisher Scientific) + Preston supplement (amphotericin B (10 μg/ml), rifampicin (10 μg/ml), trimethoprim lactate (10 μg/ml), polymyxin B (5 IU/ml); Oxoid) in a 6-well plate (Corning, Corning, NY). Plates were placed inside plastic Ziploc freezer bags containing 1–2% O_2 + Bioblend gas (10% CO_2, 10% H_2 and 80% N_2; Praxair, Danbury, CT) and incubated for 24 h at 37 °C and 40 rpm for enrichment. After incubation, a 10 μl loop of enriched sample was struck onto an ABA plate (Anaerobe Basal Agar; Oxoid) amended w/ 15% lysed horse blood (Innovative Research, Novi, MI) and CAT supplement (cefoperazone (8 μg/ml), amphotericin B (10 μg/ml), teicoplanin (4 μg/ml); Oxoid) for 24 h at 37 °C. Plates were incubated in microaerobic gas jars (AnaeroJar 3.5L System, Oxoid) at 37 °C under 1–2% O_2 + Bioblend gas for 24–48 h. All positive cultures were examined under a 1,000× phase-contrast microscope. Cultures positive for *Campylobacter* were then filtered through 0.6 μ mixed-cellulose filters (Whatman, Thermo Fisher Scientific) onto ABA plates. After growth for 24 h, single colonies were picked onto a new ABA plate and incubated in microaerobic gas jars for 24–48 h, as above. Pure cultures were stored at −80 °C on Microbank beads (Pro-Lab Diagnostics, Round Rock, TX).

Scotland sheep and pig fecal samples were collected from farms across North East Scotland, Scotland, UK by Food Standards Scotland (FSS) Operations staff at the Portlethen and Brechin abattoirs on a monthly basis. The samples were stored in a refrigerator and couriered on the next day in cool boxes with cooling gel packs to the lab at the University of Aberdeen. Bacterial strains of *Campylobacter* were isolated from pig ($n = 87$) and sheep feces ($n = 1$). Animal fecal specimens (25 g) were homogenized in 225 ml of *Campylobacter* enrichment broth and incubated at ambient temperature in enrichment broth for 1 h with occasional agitation. Enrichment incubations were performed in a modular atmosphere-controlled system cabinet (5% H_2, 10% CO_2, and 85% N_2). 100 μl of broth was plated directly onto modified

charcoal–cefoperazone–deoxycholate agar (mCCDA) plates, with the remainder incubated at 37 °C for 48 h in a modular atmosphere-controlled system cabinet (5% H_2, 10% CO_2, and 85% N_2) for enrichment growth. If direct plating yielded no colonies, then 100 µl of enrichment growth was plated onto mCCDA plates, and the plates incubated under the same conditions as described above for a further 48 h. A single well-isolated colony was picked from the mCCDA plate following incubation and latex sero-agglutination for identification of *Campylobacter* spp. was carried out using the Microgen kit (Surrey, UK). The agglutinate-positive colony was then plated onto a blood agar plate (E&O, Bonnybridge, UK) and incubated for a further 48 h at the same optimum temperature and conditions. Upon incubation, a 1 µl loopful of growth was harvested and used for genomic DNA extraction. Colonies from the remainder of the plate were stored in glycerol broth at −80 °C.

Information (MLST sequence type, source and location/date of isolation) on the *C. fetus* group strains from this study is listed in supplementary table S1, Supplementary Material online.

Genomic DNA Extraction

Prior to freezing, DNA was extracted from a 1–3 µl loop of pure culture using the Wizard Genomic DNA Purification Kit (Promega, Madison, WI), as per the manufacturer's guidelines. The concentration of DNA was quality checked by running it on a 1.5% agarose gel with 10 µl of SafeView dye (NBS Biologicals, Cambridgeshire, UK). Samples were stored at −80 °C until needed.

Polymerase Chain Reactions

Standard amplifications were performed on a Tetrad thermocycler (Bio-Rad, Hercules, CA) with the following settings: 94 °C for 30 s, 53 °C for 30 s, and 72 °C for 2 min (30 cycles). Each amplification mixture contained 50 ng genomic DNA, 1× PCR buffer (Epicentre), 1× PCR enhancer (Epicentre), 2.5 mM $MgCl_2$, 250 µM each dNTP, 50 pmol each primer, and 1 U polymerase (New England Biolabs). Amplicons were purified using ExoSAP-IT (Affymetrix, Santa Clara, CA). Sequencing and PCR oligonucleotides were designed using Primer Premier (v 5.0, Premier Biosoft, Palo Alto, CA) and purchased from Eurofins Genomics (Louisville, KY).

Sanger Sequencing

Sanger cycle sequencing reactions were performed on a 96-well Tetrad thermocycler (Bio-Rad, Hercules, CA) using the ABI PRISM BigDye terminator cycle sequencing kit (version 3.1) and standard protocols. Cycle sequencing extension products were purified using BigDye XTerminator (Applied Biosystems). DNA sequencing was performed on an ABI PRISM 3730 DNA Analyzer (Applied Biosystems), using POP-7 polymer and ABI PRISM Genetic Analyzer Data Collection and ABI PRISM Genetic Analyzer Sequencing Analysis software. Sequences were trimmed, assembled and analyzed in SeqMan (v 8.0.2; DNASTAR, Madison, WI).

Next-Generation Sequencing

Shotgun and paired-end (8–12 kb) 454 reads were obtained on a Roche 454 GS-FLX+ Genome Sequencer with Titanium chemistry using standard protocols. Illumina MiSeq libraries were prepared as described previously (Miller et al. 2014a) with the following changes: DNA was sheared using an M220 instrument (Covaris, Woburn, MA) in 50 µl screw-cap microtubes at 50 peak power, 20 duty factor, 20 °C, 200 cycles per burst and 25 s duration. Adapter-ligated fragments were size-selected to 700–800 bp. PCR was reduced to four cycles to minimize amplification bias. Pooled libraries were sequenced on a MiSeq instrument (Illumina) at 13.5 pM, based on ddPCR quantification, using 2 × 250 bp paired end v2 kits, following manufacturer's protocols. Illumina HiSeq sequencing was carried out at the Wellcome Trust Sanger Institute in Hinxton, Cambridge, using an Illumina HiSeq 2000 sequencer with 100 base paired-end sequencing. Illumina HiSeq reads were also obtained from SeqWright (Houston, TX). Single Molecule, Real-Time (SMRT) sequencing was performed on the Pacific Biosciences (PacBio) RSII sequencing platform using 20-kb SMRTbell libraries as described (Miller et al. 2014a) with the following changes: P6/C4 sequencing chemistry was used with the 360 min data collection protocols. The SMRTbell libraries were prepared from 10 mg of bacterial genomic DNA, using the standard protocol from Pacific Biosciences, as described in the Procedure-Checklist-20-kb-Template-Preparation-Using-BluePippin-Size-Selection procedure, and processed for sequencing as recommended by the supplier. A FASTQ file was generated from the PacBio reads using SMRTanalysis (ver. 2.3.0).

Genome Assembly

The five genomes completed in this study (table 1*A*) were sequenced using three next-generation platforms. First, for each genome, shotgun and paired-end 454 reads were assembled using the Roche Newbler assembler (v 2.6) into a single chromosomal scaffold. Each scaffold was closed into one contiguous sequence, using the 454 repeat contigs and the Perl script contig_extender3 ([Merga et al. 2013]; see also supplementary File S1, Supplementary Material online); contig gaps/junctions within each scaffold were also validated using PCR amplification and Sanger sequencing. Second, Illumina HiSeq or MiSeq reads were assembled de novo within Newbler; these Illumina contigs were assembled into the 454 assembly within Seqman (v. 8.0.2) to validate all 454 base calls. The presence/absence of single nucleotide polymorphisms (SNPs) within the repeat contigs was assessed using Geneious (v. 8.1.2, Biomatters, Auckland, NZ) and the Illumina reads or by using the 454 paired-end reads to link

Table 1

(A) *C. lanienae* and *C. lanienae*-Related Genomes Completed in This Study

Strain	Clade	Isolation source	Location	Optical map	Coverage (×)				Accession no(s).[a]
					454	Illumina	PacBio	Total	
C. lanienae NCTC 13004[T]	N/A	Human, fecal	Switzerland	Y (*Afl*II)	97	1396[H]	350	1843	CP015578
Campylobacter sp. NCTC 13003	1	Unknown	Unknown	Y (*Afl*II)	93	236[M]	351	680	CP018788 (chr)
Campylobacter sp. RM6137	2	Wild pig, fecal	USA: California	Y (*Spe*I)	58	208[M]	297	563	CP018789 (chr) CP018790 (pSUIS6137)
Campylobacter sp. RM8964	3	Goat, fecal	USA: California	Y (*Nco*I)	42	398[M]	429	869	CP018791 (chr) CP018792 (pVIC8964)
Campylobacter sp. RM12175	3	Cow, fecal	USA: California	Y (*Nco*I)	66	516[M]	445	1027	CP018793 (chr) CP018794 (pVIC12175-1) CP018795 (pVIC12175-2)

(B) Other Complete *C. fetus* Group Genomes Utilized in This Study

Strain	Isolation source	Location	Accession number
C. fetus fetus 04/554	Bovine fetus	Argentina	CP008808
C. fetus fetus 82-40	Human, blood	USA	CP000487
C. fetus venerealis 97/608	Bovine, placenta	Argentina	CP008810
C. fetus venerealis intermedius 03/293	Aborted bovine fetus, lung	Argentina	CP006999
C. fetus testudinum 03-427[T]	Human, blood	USA:New York	CP006833
C. fetus testudinum SP3	Western hog-nose snake	UK	CP010953
C. hyointestinalis hyointestinalis LMG 9260	Human, fecal	Belgium	CP015575
C. hyointestinalis lawsonii LMG 15993	Pig, stomach	Sweden	CP015576
C. iguaniorum 1485E[T]	Bearded dragon	The Netherlands	CP009043
C. iguaniorum 2463D	Green iguana	The Netherlands	CP010995
C. iguaniorum RM11343	Alpaca, fecal	USA:California	CP015577

[a]chr: chromosome; [H]: HiSeq; [M]: MiSeq.

SNPs to adjacent unique contigs. Third, all five genomes were also sequenced using a PacBio *RS* sequencer (Pacific Biosciences, Menlo Park, CA) to address repeat regions that could not be resolved using 454/Illumina/Sanger sequencing. Finally, each assembly was verified using a bacterial optical restriction map (OpGen, Gaithersburg, MD).

Seventy-one draft genomes were also sequenced for this study. All draft genomes were obtained by assembling Illumina MiSeq reads de novo within Newbler (v 2.6) or by using Velvet Optimiser (v. 2.2.5). Assembly information for the draft genomes is listed in supplementary table S2, Supplementary Material online.

Genome Annotation

Putative coding sequences (CDSs) were identified using GeneMark (Besemer and Borodovsky 2005) for the completed genomes or Prodigal (Hyatt et al. 2010) for the draft genomes. Annotation was accomplished by comparing the predicted proteins to the proteomes of other members of the *C. fetus* group (e.g., *C. fetus* subsp. *fetus* strain 82-40 (CP006833.1), *C. iguaniorum* 1485E (CP009043.1), and *C. hyointestinalis* subsp. *hyointestinalis* LMG 9260 (CP015575.1)) and to the NCBI nonredundant (nr) database using BLASTP; positive matches had an identity of $\geq 45\%$, and an alignment length of $\geq 75\%$ across both the query and match sequences. Final annotation, including manual start codon curation, determination of homopolymeric G:C tract variability and identification of rRNA- and tRNA-coding genes and pseudogenes, was performed as previously described (Miller et al. 2014a). Additionally, selenocysteine tRNAs were identified using ARAGORN (Laslett and Canback 2004) and CRISPR elements were identified using CRISPRFinder (Grissa et al. 2007).

Comparative Genome Analysis

The predicted proteomes of the completed (five strains; table 1*A*) and draft *C. lanienae* clade genomes (71 strains; supplementary table S2, Supplementary Material online) sequenced in this study, and the other completed *C. fetus* group genomes sequenced previously (11 strains; table 1*B*), were compiled to create a composite proteome representing 87 strains and containing 155,825 proteins. Comparative genomic analysis was performed through a pairwise BLASTP analysis of the *C. fetus* group proteome against itself. The custom Perl script BlastPTrimmer14 (supplementary File S2, Supplementary Material online) was used to identify the top

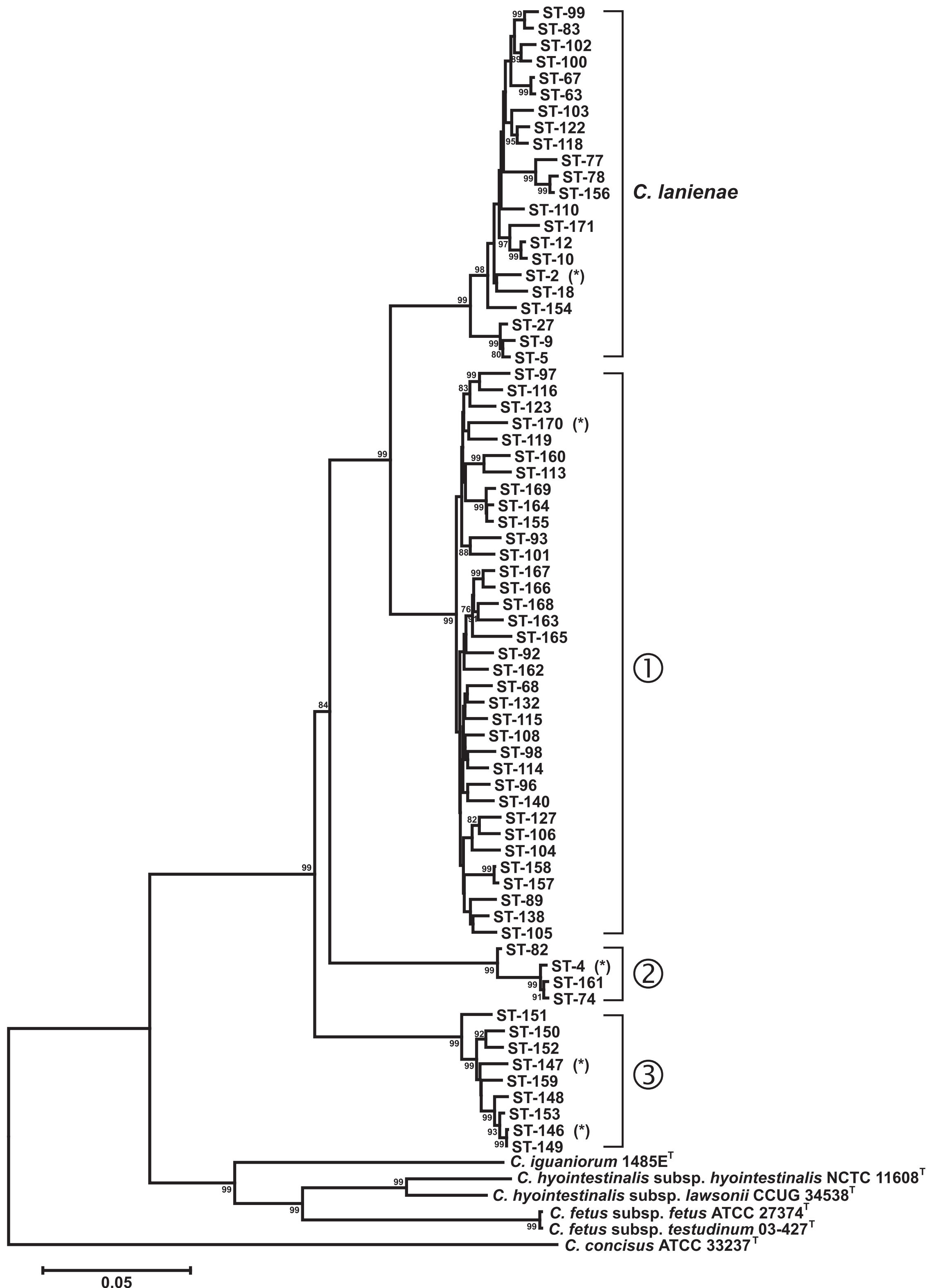

Fig. 1.—Phylogeny of *C. lanienae* clade MLST profiles. Concatenated MLST alleles (concatenated in the order *aspA-atpA-glnA-gltA-glyA-pgm-tkt*) representing profiles within the *C. lanienae* MLST scheme were exported from PubMLST and aligned. The dendrogram was constructed using the neighbor-joining algorithm and the Kimura 2-parameter distance correction model. Bootstrap values of ≥75%, generated from 500 replicates, are shown at the nodes. Putative novel taxa are numbered (e.g., "1" = Clade 1). *C. lanienae* clade sequence types represented by complete genomes are indicated (*). The concatenated profile sequences for the *C. fetus* (subspp. *fetus* and *testudinum*), *C. iguaniorum*, and *C. hyointestinalis* (both subspecies) type strains are included for comparison. The concatenated profile sequence for *C. concisus* strain ATCC 33237^{T} was included in the alignment to root the tree. Scale bar represents substitutions per site.

match for each protein within the other proteomes, where present, using the match parameters described above. Due to the large size of the BLASTP output (~19 million lines), BlastPTrimmer14 was run on a workstation with two six-core Xeon X5690 processors (24 total processors) and 96 GB of RAM. The extra RAM was necessary, because BlastPTrimmer14 loads all of the input files directly into memory. To facilitate identification of genes and genetic loci present in the various *C. fetus* group taxa, the BlastPTrimmer14 output file Genesimmatrix.txt was imported into Excel, such that the orthologs were displayed in a tabular format with the genes identified in the completed genomes along the Y axis and their respective orthologs (where present) identified in the 87 genomes along the X axis (clustered by taxon). Further comparative analysis was performed using JSpecies (v. 1.2.1; [Richter and Rossello-Mora 2009]), using default parameters, to determine average nucleotide identity (ANI) values.

Phylogenetic Analysis

Sequence alignments were performed using CLUSTALX (ver. 2.1). Dendrograms were constructed using the neighbor-joining method and Poisson correction. Bootstraps were conducted with 500 replicates. Phylogenetic analyses were performed using MEGA version 6.05 (Tamura et al. 2013).

Accession Numbers

The nucleotide sequences and annotations of the genomes completed in this study (table 1*A*) and the contig sequences for the draft genomes (supplementary table S2, Supplementary Material online) were deposited in GenBank.

Results and Discussion

Identification of the *C. lanienae* Clade within *Campylobacter*

As described above, putative new taxa within the *Campylobacter fetus* group were identified through MLST and *atpA* typing (Miller et al. 2012, 2014b). Within the non-*jejuni/coli Campylobacter* PubMLST database (http://pubmlst.org/campylobacter/), 171 *C. lanienae* MLST profiles have been deposited (accessed Jan 24, 2017). Concatenated sequences representing these 171 profiles were exported from PubMLST. Phylogenetic analysis of a subset (70/171) of these *C. lanienae* profile sequences clearly indicated that they form a clade distinct from the other species of the *C. fetus* group, that is, *C. fetus*, *C. hyointestinalis*, and *C. iguaniorum* (fig. 1). Additionally, this clade (referred to hereafter as the "*C. lanienae* clade") is composed of four smaller, discrete clades: *C. lanienae* and three clades representing putative novel species (labeled clades 1–3 in fig. 1). Clade 1 contains strain NCTC 13003 (ST-170; deposited in the NCTC as *C. lanienae*; ST-1 described earlier by Miller et al. [2012], clade 2 contains strain RM6137 (ST-4 described earlier by Miller et al. [2012]) and clade 3 contains strains RM8964 and RM12175 (cluster 4 in fig. 3 from Miller et al. [2014b]). The genomes of these four strains, as well as the genome of the *C. lanienae* type strain NCTC 13004, were sequenced to completion (table 1*A*). The genomes of 71 strains representing the remaining *C. lanienae* clade MLST profiles illustrated in figure 1 were sequenced to draft level (supplementary table S2, Supplementary Material online).

The amino acid sequences of 20 proteins conserved among the 76 *C. lanienae* clade genomes were extracted and concatenated for each strain. Phylogenetic analysis of these concatenated protein sequences reiterated the results of the MLST-based phylogeny (fig. 2). Again, four distinct clades were observed that were clearly distinguishable from the other taxa within the *C. fetus* group. Comparative analyses using the full set of 1,051 proteins common to the *C. fetus* group gave similar results (supplementary fig. S2, Supplementary Material online). Average nucleotide identity (ANI) analysis of the completed and draft genome sequences also supported the above results (table 2). It is widely held that ANI values of ~95% correspond to DNA–DNA hybridization values of 70% (Goris et al. 2007; Richter and Rossello-Mora 2009). Thus, ANI values below 95% would indicate that the two strains are representatives of different bacterial species. However, the ANI values between the *C. hyointestinalis* subspecies are 94% and the ANI values between *C. fetus* subsp. *testudinum* and either of the other two *C. fetus* subspecies is 92% (table 2), suggesting that the ANI species boundary in *Campylobacter* is somewhat lower than 95%. Nevertheless, average ANI values between clades 1, 2, or 3 and the other taxa (putative or previously described) within the *C. fetus* group are between 71% and 87% (table 2). This is consistent with ANI comparisons between other *Campylobacter* species (e.g., *C. jejuni* and *C. coli* have ANI values of ~84% [data not shown]). Thus, even with the potential of a discriminatory ANI value in *Campylobacter* that is lower than 95%, it is likely that clades 1–3 represent novel species within the *C. fetus* group and *C. lanienae* clade.

General Features of the Closed *C. lanienae* Clade Genomes

The genomes of the *C. lanienae* clade genomes range in size from 1.58 to 1.73 Mb ($\bar{x} = 1.63$ Mb; table 3). Therefore, these genomes are generally smaller than other genomes within the *C. fetus* group (1.68–1.94 Mb; $\bar{x} = 1.79$ Mb; table 3). However, the %G + C content is relatively consistent across the entire *C. fetus* group, ranging from 32.1% to 35.9% GC. With the exception of *C. lanienae* strain NCTC 13004, all of the *C. lanienae* clade strains contain at least one plasmid, with the two clade 3 strains each containing a highly-similar 25 kb plasmid.

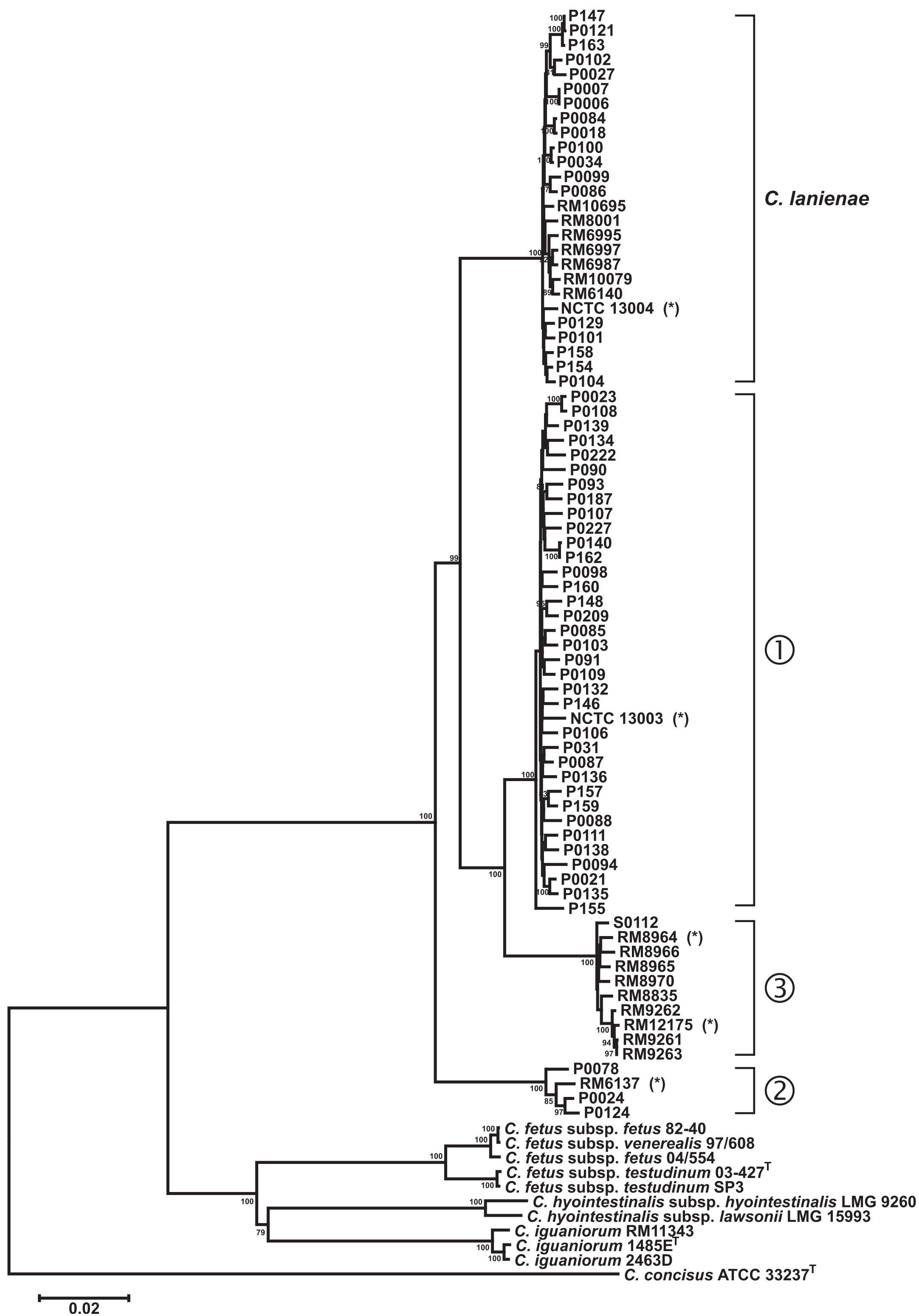

Fig. 2.—Phylogenetic relationships within the *C. fetus* group. The amino acid sequences of 20 core proteins (DnaN, NrdB, Tkt, Eno, QueD, FabH, Mrp, IleS, GroEL, SpeA, GlnA, HobA, GlyA, SpoT, Pnp, MiaB, AtpA, HemB, PrfA, and Frr), extracted from completed or draft genomes, were concatenated and aligned. The dendrogram was constructed using the neighbor-joining algorithm and the Poisson distance correction model. Bootstrap values of ≥75%, generated from 500 replicates, are shown at the nodes. Putative novel taxa within the group are numbered (e.g., "1" = Clade 1). Complete genomes within the *C. lanienae* clade are indicated (*). The concatenated 20 core protein set of *C. concisus* strain ATCC 33237^{T} was included in the alignment to root the tree. Scale bar represents substitutions per site.

Table 2

Average Nucleotide Identity (ANI) Values for Taxa within the *C. fetus* Group

	N	*Cf*[a]	*Cft*	*Cig*	*Chh*	*Chl*	*Clan*	Clade 1	Clade 2	Clade 3
C. fetus[a]	4	100	92	75	78	77	72	72	72	72
C. fetus testudinum	2	92	99	76	79	78	72	72	72	72
C. iguaniorum	3	75	76	99	76	75	72	72	72	72
C. hyointestinalis hyointestinalis	1	78	79	76	98	94	72	72	72	72
C. hyointestinalis lawsonii	1	77	78	75	94	99	74	73	73	72
C. lanienae	26	72	72	71	71	74	98	87	86	82
Clade 1	36	72	72	71	71	73	87	98	83	86
Clade 2	4	72	72	72	72	73	86	83	98	83
Clade 3	10	72	72	72	71	72	82	86	83	98

Note.—Values represent averages (in %) of all pairwise combinations within each taxon pair. Values ≥ 95% are shaded in dark grey; values between 90 and 95% are shaded in light grey.

[a]Includes both subspecies *fetus* and *venerealis* and the intermedius biovar.

Differences in the sizes of the *C. lanienae* clade genomes, with respect to the *C. fetus* group as a whole, cannot be attributed solely to the presence of genomic islands and integrated prophage, because such elements are found in both clades within the *C. fetus* group. Subtracting the gene content of these genomic elements, the *C. lanienae* clade genomes encode 1,507 genes on average, compared with 1,652 genes encoded on average by other members of the *C. fetus* group (table 3). During annotation, coding sequences were assigned to one of three categories (table 3): assigned function (i.e., genes with a gene name and a specific function, e.g., *rpoA*); general function (e.g., methyltransferase); and hypothetical. Although the proportion of hypothetical genes is roughly equivalent across both clades of the *C. fetus* group ($\bar{x} = 21\%$ [*lanienae*] versus 20% [*fetus*]), genomes of the *C. lanienae* clade strains encode fewer genes of general function ($\bar{x} = 399$ [*lanienae*] versus 504 [*fetus*]; table 3). The *C. lanienae* clade strains also encode fewer genes of assigned function ($\bar{x} = 849$ versus 872); however, the proportion of assigned genes is higher in the *C. lanienae* clade strains ($\bar{x} = 54\%$ versus 51%; table 3).

As illustrated in table 3, this reduction in gene content within the *C. lanienae* clade is reflected in part in the smaller number of genes that encode signal transduction (excluding *che*/*mot* genes) and transcriptional regulatory proteins. Signal transduction proteins possessing multiple, N-terminal sensory domains were not observed within the *C. lanienae* clade proteome. Nevertheless, it is possible that the potential deficit in signal transduction capacity within the *C. lanienae* clade is compensated for to some extent by increasing the number and range of input domains on one or more of the *C. lanienae* clade signal transduction proteins. Limitations in the suite of signal transduction and regulatory proteins might imply a more restricted lifestyle for members of the *C. lanienae* clade, whereby these organisms could utilize, respond and adapt to fewer nutrient sources and environmental conditions. One consequence might be differences in host range between the two groups of strains (Baumler and Fang 2013). However, members of both clades within the *C. fetus* group have been isolated from both foraging (swine) and grazing (cows, sheep, goats, alpaca) animals. Nevertheless, little data exist on the true host range of many of these taxa beyond farmed animals, especially because many of the taxa described here are potentially novel or recently-described. A reduction in signaling pathways may indicate that these organisms are less able to grow and survive outside of a host. *C. hyointestinalis* has been isolated from water samples by our lab (data not shown), whereas *C. lanienae* clade organisms have not been isolated from environmental samples. However, here again critical sampling data are lacking and a definitive conclusion about host range or host restriction cannot be reached. Finally, although these organisms are occasionally isolated from human clinical samples, little is known about the virulence of many of the *C. fetus* group taxa; thus, it is unknown what effect the observed genomic reduction would have on the pathogenicity of the *C. lanienae* clade organisms.

CRISPR Loci

Clustered regularly interspaced short palindromic repeats (CRISPR) systems are prevalent in prokaryotes, with each CRISPR array consisting of a series of direct repeats separated by unique spacers. These arrays are generally adjacent to several *cas* (CRISPR-associated) genes (Jansen et al. 2002). The Cas proteins are responsible for both the integration of short stretches of viral or plasmid DNA into the CRISPR array and the cleavage of incoming foreign DNA or RNA (Makarova et al. 2011). Thus, CRISPR/Cas systems provide a form of adaptive immunity against invading genetic elements (Barrangou et al. 2007; Garneau et al. 2010). CRISPR/Cas systems are divided into two main classes: in Class 1 systems, foreign nucleic acids are cleaved by a multi-protein complex, whereas in Class 2 systems cleavage is performed by a single protein (e.g., Cas9 or Cpf1 [Hille and Charpentier 2016]). These two classes are further divided into several types and subtypes (Chylinski et al. 2013; Hille and Charpentier 2016;

Makarova et al. 2011), with types I, III and IV comprising Class 1 systems and types II, V and VI comprising Class 2 systems (Hille and Charpentier 2016). Each type and subtype is composed of a defined suite of *cas* genes and alleles; however, *cas1* and *cas2* have been identified in nearly all CRISPR subtypes (Hille and Charpentier 2016).

In general, *C. lanienae* clade strains contain type II-B or type II-C CRISPR/Cas systems, whereas the remaining *C. fetus* group taxa contain eight gene (*cas1-8*) type I-B CRISPR/Cas systems, if present (supplementary fig. S3, Supplementary Material online). *C. hyointestinalis* subsp. *hyointestinalis* strain LMG 9260 contains a type II-C CRISPR/Cas system in addition to the type I-B system, but the *cas9* gene is presumably nonfunctional. *cas9* is also a putative pseudogene in *C. lanienae* strain NCTC 13004 and *Campylobacter* sp. strain NCTC 13003. *Campylobacter* sp. strain RM12175 also contains a type III-A cluster adjacent to the type II-B cluster; however, this type III cluster does not contain a *cas1* or *cas2* gene. *cas1* and *cas2* are absent in some type III clusters (Makarova et al. 2011) and, similar to strain RM12175, these organisms contain additional CRISPR loci, suggesting that the missing type III Cas1/Cas2 functionality might be provided in trans by one of the other CRISPR systems (Makarova et al. 2011).

CRISPR systems have been identified in multiple *Campylobacter* taxa ([Ali et al. 2012; Deshpande et al. 2011; Pearson et al. 2015; Tasaki et al. 2012); see also supplementary fig. S3, Supplementary Material online). *C. jejuni* and other members of the "thermotolerant" campylobacters contain Class 2 type II-C CRISPR systems, whereas "nonthermotolerant" *Campylobacter* species, such as *C. concisus* and *C. rectus*, possess Class 1 type I-B CRISPR systems. Therefore, the *C. lanienae* clade taxa are unusual: they do not contain the Class 1 CRISPR systems identified in other nonthermotolerant *Campylobacter* spp, but rather carry CRISPR types (i.e., type II-B and type III-A) not observed previously in *Campylobacter*. Although these type II-B and type III-A systems likely function in bacterial immunity in a fashion similar to the type II-C and type I-B systems, any potential roles of these CRISPR systems in other aspects of *Campylobacter* biology, such as gene regulation and/or virulence remain to be determined.

Flagellar Genes within the *C. lanienae* Clade

The genomes of the motile campylobacters typically encode one to three flagellin subunits (Gilbert et al. 2013, 2014; Miller and Yee 2015; Miller et al. 2014a, 2016a; Parkhill et al. 2000). However, analysis of the closed *C. lanienae* clade genomes indicated that they encode between four (*Campylobacter* sp. strain RM12175) and ten (*Campylobacter* sp. strain NCTC 13003) flagellin subunits (table 3). Unlike *C. jejuni*, in which the *flaA* and *flaB* genes are adjacent in the chromosome, these flagellin genes are not in tandem but are scattered throughout the chromosome. Additionally, hypervariable GC tracts are present immediately upstream of several of these flagellin genes. Because these tracts are located upstream of the *fla* coding regions, they would not affect *fla* translation; however, it is possible that these tracts could regulate *fla* transcription.

Phylogenetic analysis identified a substantial amount of variety among the *C. lanienae* clade Fla alleles. Very few alleles were found twice in the same chromosome (fig. 3), and 29 alleles were identified in the five completed genomes for the clade, with an average amino-acid similarity of 76.6–83.5% between the alleles for each taxon. An additional 197 complete *fla* genes (156 Fla alleles) were identified within the draft proteomes (supplementary table S3, Supplementary Material online). Moreover, 315 *fla* gene fragments were situated at contig ends within the draft genomes (supplementary table S3, Supplementary Material online), and were not included in the count above. Thus, the gene total listed above represents a minimum value. Although the true number of *fla* genes per strain would require closure of each draft genome, it is likely, based on these data, that the presence of four or more flagellin genes per strain is a conserved feature within the *C. lanienae* group. Similar to the completed genomes, very few allele duplications were observed within each draft genome.

Campylobacter flagella are modified through the addition of the nine carbon sugars pseudaminic acid and/or legionaminic acid (Logan et al. 2008; Schoenhofen et al. 2009). Orthologs of genes in the pseudaminic and legionaminic acid biosynthetic pathways were identified in the completed and draft *C. lanienae* clade genomes, suggesting that these flagellin subunits are glycosylated in a fashion similar to that observed in *C. jejuni*. Additionally, three motility accessory factor/sugar transferase genes were identified in each closed genome, in identical chromosomal locations (linked to *nfo*, *nth* or *ubiX*).

The existence of multiple flagellin subunits raises intriguing questions about the *C. lanienae* clade flagella. It is not known whether all of the putative flagellin subunits are expressed at the same time in any given cell: if they are all constitutively expressed or are only expressed under certain environmental conditions. Furthermore, it is also unknown what effect, if any, the upstream hypervariable GC tracts have on flagellin expression in these organisms. The presence of up to ten putative flagellin genes with multiple alleles, often in close proximity to hypervariable GC tracts, and encoding flagellar subunits that are modified with at least two nine carbon sugars would strongly indicate that the *C. lanienae* clade taxa possess a dynamic and variable flagellar structure, which may reflect more extensive interactions with animal host immune systems than is the case in other *Campylobacter* species. Dissection of this flagellar structure would certainly be warranted and will require further study.

Selenium Metabolism within the *C. fetus* Group

In *Campylobacter*, selenium is utilized in two ways: 1) through modification of tRNA-Glu$_{UUC}$, tRNA-Gln$_{UUG}$, or tRNA-Lys$_{UUU}$

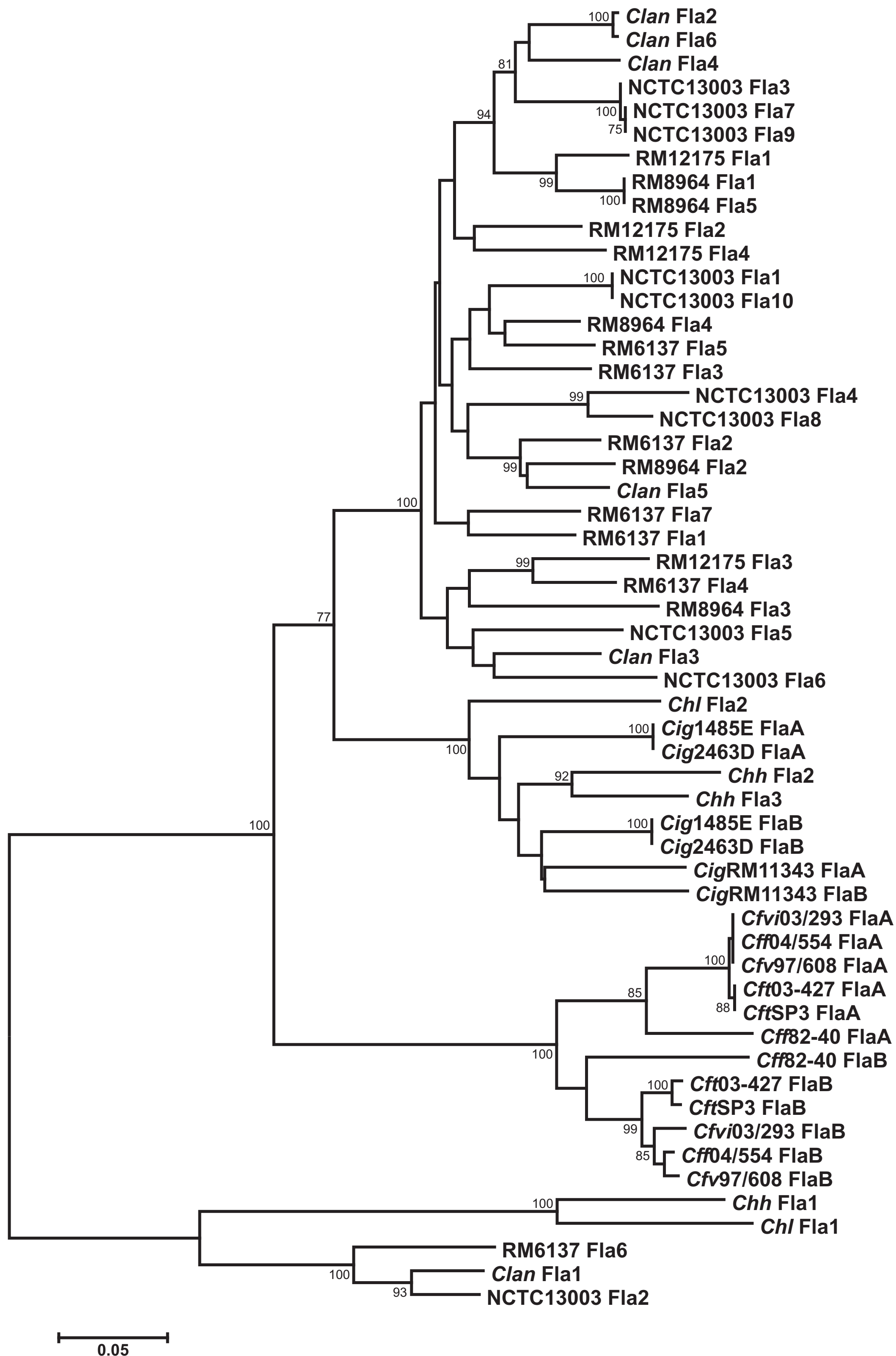

Fig. 3.—Diversity of *C. fetus* group Fla alleles. Flagellin protein sequences were extracted from the completed *C. fetus* group genomes and aligned. The dendrogram was constructed using the neighbor-joining algorithm and the Poisson distance correction model. Bootstrap values of ≥75%, generated from 500 replicates, are shown at the nodes. Scale bar represents substitutions per site.

Table 3
General Features of the Closed *C. fetus* Group Genomes

	C. lanienae NCTC 13004[T]	*Campylobacter* sp. NCTC 13003 (Clade 1)	*Campylobacter* sp. RM6137 (Clade 2)	*Campylobacter* sp. RM8964 (Clade 3)	*Campylobacter* sp. RM12175 (Clade 3)	*C. fetus fetus* 82-40	*C. fetus venerealis* 97/608	*C. fetus testudinum* SP3	*C. iguaniorum* 1485E[T]	*C. h. hyointestinalis* LMG 9260	*C. h. lawsonii* LMG 15993
Chromosome											
Size (kbp)	1,594.55	1,593.99	1,634.52	1,729.34	1,583.17	1,773.61	1,935.03	1,819.05	1,684.61	1,753.39	1,753.28
% G+C content	34.6	33.6	34.2	32.1	32.5	33.3	33.3	33.1	35.9	34.0	33.6
CDS numbers[a]	1561	1510	1583	1692	1568	1687	1865	1766	1646	1678	1711
Assigned func. (% CDS)	854 (55)	850 (56)	858 (54)	844 (50)	841 (54)	901 (53)	885 (47)	894 (51)	856 (52)	857 (51)	839 (49)
Pseudogenes	24	48	27	27	28	26	55	9	18	59	55
General func. (% CDS)	397 (25)	374 (25)	389 (25)	437 (26)	399 (25)	474 (28)	533 (29)	520 (29)	510 (31)	498 (30)	487 (28)
Hypothetical (% CDS)	310 (20)	286 (19)	336 (21)	411 (24)	328 (21)	312 (18)	447 (24)	353 (20)	280 (17)	323 (19)	385 (23)
Prophage/genetic islands	0	1	2	6	3	2	5	2	0	2	3
Zot islands	0	0	0	0	1	0	0	0	0	1	1
CDS not in genetic islands	1561	1493	1513	1480	1489	1666	1649	1697	1646	1660	1596
CRISPR/Cas loci	1 (type II)	1 (type II)	1 (type II)	1 (type II)	2 (type II, type II:III)	1 (type I)	0	1 (type I)	1 (type I)	2 (type I, type II[b])	0
G:C tracts ≥8 nt											
Total/HV (#intergenic)	50/47 (25)	42/29 (13)	56/51 (27)	38/28 (10)	49/45 (14)	30/30 (2)	34/26 (2)	35/26 (5)	34/23 (3)	53/52 (25)	49/42 (17)
Plasmids (size in kb.)	0	27	6	25	25; 4	0	38; 27	0	70	0	0
Signal transduction											
Che/Mot proteins	9	9	9	9	9	8	9	9	9	9	9
MCPs	7	8	6	9	8	14	12	16	7	6	5
2CS response regulators	8	8	8	8	8	12	12	12	14	13	10
2CS histidine kinases	7	8	7	7	8	11	10	11	13	12	9
Other	9	10	9	10	10	13	13	13	14	13	14
Motility											
Fla genes	6	10	7	5	4	2	2	2	2	3	2
R/M systems											
Type I (*hsd*)	2[c]	3[c]	1[c]	0[c]	1	1	1	0	0	1	1
Type II/IIS/III	2[c]	2	5	1	3	1	1	2	3	2	5
DNA methylases	3	2	2	4	4	1[c]	5	1	3	1	2
Transcription											
Regulatory proteins[d]	16	14	13	14	15	22	25	26	26	22	25
Transposition											
Transposase genes/IS elements; degenerate	3; 0	0	0	0	0; 2	0	13[e]; 0	0	0	5; 2	1; 1

CRISPR: clustered regularly interspaced short palindromic repeats; HV: hypervariable; Che/Mot: chemotaxis/motility; MCP: methyl-accepting chemotaxis protein; 2CS: two component system; R/M: restriction/modification; Zot: zonula occludens toxin.

[a]Number does not include pseudogenes.

[b]*cas9* degenerate in the typeII CRISPR locus.

[c]Strain also contains degenerate R/M systems/DNA methylases.

[d]Total does not include two-component system response regulators or sigma factors.

[e]Includes three plasmid-borne IS elements.

Table 4
Selenium Metabolism Genes and Genes Encoding Selenium-Dependent Enzymes

Gene	Function	*C. lanienae* (*N* = 26)	Clade 1 (*N* = 36)	Clade 2 (*N* = 4)	Clade 3 (*N* = 10)	*C. fetus fetus*	*C. fetus venerealis*	*C. fetus testudinum*	*C. iguaniorum*	*C. hyointestinalis*	*C. jejuni*	*C. coli*	*C. upsaliensis*	*C. lari lari*	*C. lari concheus*	*C. peloridis*	*C. subantarcticus*	*C. volucris*	*C. insulaenigrae*	*C. concisus*	*C. curvus*	*C. mucosalis*	*C. showae*	*C. rectus*	*C. gracilis*	*C. ureolyticus*
selA	selenocysteine synthase	0	0	0	0	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
selB	selenocysteine-specific elongation factor	0	0	0	0	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
selC	selenocysteinyl-tRNA	0	0	0	0	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
selD	selenophosphate synthetase	0	0	0	0	Y_s	Y_s	Y_c	Y_c	Y_c	Y_s	Y_s	Y_s	Y_s	Y_s	Y_s	Y_s	Y_s	Y_s	Y_s	Y_s	Y_s	Y_s	Y_s	Y_s	Y_s
	selT/W/H family selenoprotein	0	0	0	0	Y_1	Y_1	Y_2	Y_3	Y_4	Y_4	0	Y_4	Y_4	Y_4	Y_4	Y_4	Y_4	Y_4	Y_4	Y_4	Y_4	Y_4	Y_4	Y_4	Y_4
	DUF466 domain protein (CJ0916c)	Y_c	Y_c	Y_c	Y_c	Y_s	Y_s	Y_s	Y_s	Y_s	Y_s	Y_s	Y_s	Y_s	Y_s	Y_s	0	0	0	0	0	0	0	0	Y_s	0
yedE	putative selenium metabolism protein	0	0	0	0	Y	Y	Y	Y	Y	0	0	0	0	0	0	0	0	0	Y	Y	Y	Y	Y	Y	0
yedF	selenium metabolism protein	0	0	0	0	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
mnmH	tRNA 2-selenouridine synthase (*ybbB*; *selU*)	0	0	0	0	0	0	0	0	0	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	0	0	0	0
fdhA	formate dehydrogenase N, SeC-containing α subunit	0	0	0	0	0	0	0	0	0	Y	Y	Y	0	0	0	0	0	0	Y	0	Y	Y	Y	0	Y
fdhB	formate dehydrogenase N, Fe-S cluster subunit β	0	0	0	0	0	0	0	0	0	Y	Y	Y	0	0	0	0	0	0	Y	0	Y	Y	Y	0	Y
fdhD	formate dehydrogenase accessory protein	0	0	0	0	0	0	0	0	0	Y	Y	Y	0	0	0	0	0	0	Y	0	Y	Y	Y	0	Y
fdhM	formate dehydrogenase-associated chaperone	0	0	0	0	0	0	0	0	0	Y	Y	Y	0	0	0	0	0	0	Y	0	Y	Y	Y	0	Y
fdhF	formate dehydrogenase N, SeC-containing α subunit	0	0	0	0	Y	Y	Y	Y	Y	0	0	0	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
fdhG	formate dehydrogenase N, Fe-S cluster subunit β	0	0	0	0	Y	Y	Y	Y	Y	0	0	0	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
fdhL	formate dehydrogenase-associated chaperone	0	0	0	0	Y	Y	Y	Y	Y	0	0	0	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
fdhC	formate dehydrogenase N, cytochrome$_{b556}$ subunit	0	0	0	0	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
fdhH	formatedehydrogenase H, SeC-containing	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Y	0	0	Y	0	Y	0
fdhI	formate dehydrogenase H, SeC-containing	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Y	Y	0	Y	0	Y	0
fdhT	formate dehydrogenase biogenesis protein	0	0	0	0	Y	Y	Y	Y	Y	Y	Y	0	Y	0	0	Y	0	0	0	0	0	0	0	0	Y
fdhU	formate dehydrogenase biogenesis protein	0	0	0	0	Y	Y	Y	Y	Y	Y	Y	0	Y	0	0	Y	0	0	0	0	0	0	0	0	Y

Y_c: nonselenoprotein (cysteine-containing); Y_s: selenoprotein; Y_1: tandem copies of this protein (upstream: SeC-containing, downstream: possible pseudogene); Y_2: tandem copies of this protein (upstream: cysteine-containing, downstream: possible pseudogene); Y_3: in the reptile-associated strains 1485E and 2463D tandem copies of this protein (upstream: cysteine-containing, downstream: SeC-containing) and in the mammal-associated strain RM11343 a single SeC-containing protein; Y_4: single SeC-containing protein.

by methylaminomethyl-2-selenouridine (mnm^5Se^2U) (Wittwer 1983; Wolfe et al. 2004) and 2) translational incorporation of selenocysteine (Sec) into selenoproteins (Hatfield and Gladyshev 2002; Zinoni et al. 1987, 1990). mnm^5Se^2U-modified tRNAs are synthesized through conversion of 2-thiouridine to 2-selenouridine at the wobble position in the anticodon. 2-thiouridine conversion requires both the tRNA 2-selenouridine synthase MnmH (alternatively annotated as YbbB or SelU) and the selenophosphate synthetase SelD (Wolfe et al. 2004). Selenocysteine is incorporated into proteins through the recoding of UGA stop codons. Selenocysteine insertion requires: an in-frame UGA stop codon; SelD; selenocysteine synthase (SelA); a Sec-specific translation elongation factor (SelB); tRNA-Sec_{UCA} (*selC*); and a Sec insertion sequence (SECIS) element, a hairpin-containing sequence in the selenoprotein mRNA immediately downstream from the UGA stop codon (Hatfield and Gladyshev 2002). The *selC* tRNA-Sec_{UCA} is initially charged with serine by SerS, but is converted from seryl-$tRNA^{Sec}$ to selenocysteyl-$tRNA^{Sec}$ by SelA (Forchhammer et al. 1991). Other proteins proposed to be involved in selenium metabolism are the putative transporter YedE and the putative redox protein YedF (Lin et al. 2015).

The main class of selenoproteins in *Campylobacter* includes the formate dehydrogenases (Fdh). Four types of formate dehydrogenases are present in *Campylobacter* (Miller et al., unpublished data): two three-subunit Fdh-N complexes (FdhABC and Fdh FGC) and two single subunit Fdh-H enzymes, FdhH and FdhI. Other Fdh-associated proteins encoded by *Campylobacter* include the Fdh-associated chaperones FdhL and FdhM, the Fdh-accessory protein FdhD and the Fdh biogenesis proteins FdhT and FdhU. Additional small selenoproteins observed in *Campylobacter* include a SelT/SelW/SelH family selenoprotein and a DUF466 domain protein (Cravedi et al. 2015).

No genes encoding selenoproteins (e.g., *fdhA*, *fdhF* and *fdhH*), selenoprotein-associated proteins (e.g., *fdhB* and *fdhC*) or proteins involved in selenium metabolism (e.g., *selABD* and *mnmH*) were identified in the five completed *C. lanienae* clade genomes (table 4). Similar results were observed when the 71 draft genomes were analyzed (table 4). Additionally, the *selC* tRNA-Sec_{UCA} was not identified in any of the 76 *C. lanienae* clade genomes. tRNA-Sec_{UCA} is routinely missed by some web servers (such as tRNAscan-SE); however, parallel analyses using BLASTN or ARAGORN successfully identified *selC* within related genomes of the *C. fetus* group but could not detect *selC* within the *C. lanienae* clade. The DUF466 domain selenoprotein was identified in a subset of the *Campylobacter* genomes (table 4). Unlike other selenoproteins, where the UGA stop codon is towards the N-terminal end, the UGA codon in these proteins is at the C terminus (Cravedi et al. 2015). Thus, the last two amino acids in these proteins are cysteine and selenocysteine. DUF466 domain proteins were observed also in the *C. lanienae* clade genomes; however, these proteins are terminated by two cysteines and are therefore not selenoproteins, consistent with the absence of selenium metabolism in this clade. SelD is generally also a selenoprotein in *Campylobacter* (table 4), containing a selenocysteine residue at position 17. However, in *C. fetus* subsp. *testudinum*, *C. iguaniorum*, and *C. hyointestinalis*, SelD contains a cysteine at position 17. Nevertheless, *C. lanienae* clade organisms contain neither the cysteine or selenocysteine alleles of SelD.

Absence of genes involved in selenium metabolism would help in part to explain the lower gene content observed in members of the *C. lanienae* clade. It is unknown what effect the complete absence of selenium metabolism would have, because the repertoire of *Campylobacter* selenoproteins may be underestimated. Selenoprotein genes, including formate dehydrogenase subunits, are often incorrectly truncated, labeled as pseudogenes or missed altogether. For instance, the SelT/SelW/SelH family selenoprotein was not annotated in *C. jejuni* strain NCTC 11168. Thus, the *Campylobacter* selenoproteome could be substantially larger, especially if small (<75 aa) selenoproteins are overlooked. Querying genomic sequences for SECIS elements would improve the identification of putative selenoproteins. However, such a search is labor intensive and not always foolproof: some putative *Campylobacter* selenoproteins have very-low-scoring SECIS elements. Nevertheless, the function of some selenoproteins can be determined and the effect of their loss ascertained. Absence of formate dehydrogenases would likely lead to changes in bacterial respiration. Indeed, in *Helicobacter pylori*, that encodes only SelA, it has been proposed that a depleted selenoproteome would confer increased oxygen tolerance (Cravedi et al. 2015). Additionally, DUF466 domain proteins contain several C-terminal cysteine residues and these proteins might confer redox activity (Cravedi et al. 2015; Fomenko et al. 2007). In *Campylobacter*, these genes are always linked to *cstA* (encoding carbon storage protein A); DUF466 domain genes are absent in genomes in which CstA is not encoded or is truncated.

The absence of selenium metabolism in the *C. lanienae* clade raises questions regarding the evolution of the selenoproteome. Within the *C. fetus* clade, 13 selenium-associated genes (excluding the DUF466 domain protein from table 4) are located at seven chromosomal sites. As described above, the selenoproteome genes are "all or none" within the *C. fetus* group: all 13 genes are present in all sequenced strains of the *C. fetus* clade and all 13 genes are absent from the 76 sequenced *C. lanienae* clade genomes. Thus, there are two opposing scenarios concerning the origin and disposition of the selenoproteome genes within the *C. fetus* group: 1) the *C. fetus* group ancestral strain possessed a full selenoproteome which was subsequently lost, following seven independent deletion events, in the progenitor of the *C. lanienae* clade and 2) the *C. fetus* group ancestral strain lacked a selenoproteome and the 13 selenium-associated genes were introduced into the *C. fetus* clade via one or more lateral transfer events, with a potential subsequent movement of

the genes around the chromosome via genomic rearrangement. Although the latter scenario is more likely, because fewer initial recombination events would theoretically be required, additional research will be necessary to determine which of these scenarios accurately describes the evolutionary history of the *C. fetus* group.

Within some *Campylobacter* groups there is also notable variation in some selenoproteins. One example, identified here, is the SelT/SelW/SelH family selenoprotein gene in the *C. fetus* clade. Most campylobacters, with the exception of *C. coli* and the *C. lanienae* clade harbor one copy of this gene (table 4). However, *C. fetus* contains two copies, with the downstream copy potentially a pseudogene. The noteworthy difference at this locus is observed when the source of the organism is included. In the protein encoded by the upstream gene, the mammal-associated *C. fetus* subspecies (i.e., ssp. *fetus* and *venerealis*) contain a selenocysteine at position 11, whereas the reptile-associated *C. fetus* subsp. *testudinum* contains a cysteine at that position. Furthermore, in *C. iguaniorum*, only the reptile-associated strains contain a cysteine at position 11 in that protein (the mammal-associated *C. iguaniorum* contains a single SelT/SelW/SelH family selenoprotein). Thus, only strains of reptile origin harbor a nonselenoprotein ortholog. It is not known whether this selenoprotein plays a role in host association; nevertheless, its allelic linkage to source is intriguing. Further investigation of this locus might provide clues regarding not only the evolution of the selenoproteome but also the function of the selenoproteome in *Campylobacter*.

Conclusions

Previous studies suggested that additional taxa related to *C. lanienae* exist within *Campylobacter*, and taken together these organisms may form a discrete group within the genus. Data presented here putatively identify three novel *C. lanienae*-related species, termed clades 1–3. All three groups were isolated from either food animals or livestock: clade 1 and clade 2 strains were isolated from domestic pigs or feral swine; and clade 3 strains were isolated from cows, sheep, goats and alpacas. These sources are consistent with the isolation patterns of *C. lanienae*, which has also been isolated from cattle, sheep and swine (Inglis et al. 2004; Jay-Russell et al. 2012; Sasaki et al. 2013; Schweitzer et al. 2011). Molecular analysis indicates that *C. lanienae* and the clade1–3 taxa form a distinct cluster within the *C. fetus* group which we have termed the *C. lanienae* clade.

Strains representing two of the three new groups (clades 2 and 3) were isolated in both North America and Europe. This is generally true for all of the *C. fetus* group members, where strains have been isolated from both Old World and New World animals, even though the hosts for this group are primarily large grazing or foraging animals, such as cattle and pigs. These isolation patterns would suggest an ancient ancestry for the *C. lanienae* clade lineage. Furthermore, MLST typing data suggests that the strains within this clade do not segregate according to geographic origin. Because frequent movement of these organisms between North America and Europe is unlikely, unless transmission is mediated through different, more mobile, host animals, these data would imply a level of genomic stability. Analysis of the evolution of the *C. fetus* group, and characterization of the hosts and reservoirs of these organisms, will require further research.

The clade 1 strain *Campylobacter* sp. NCTC 13003 was deposited as a *C. lanienae*. Furthermore, using 16S typing data, several of the strains characterized in this study were initially misidentified as *C. lanienae*. Based on these results, it is possible that some of the strains originally reported to be *C. lanienae* in previous studies may have been misidentified and are actually either one of the three putative novel species or a member of another *C. lanienae*-related species. 16S typing is likely not sufficient to unequivocally assign strains to one of the *C. lanienae* clade taxa. Simple MLST typing of these organisms would also not distinguish *C. lanienae* from the *C. lanienae*-related taxa, because the alleles and profiles form part of the same MLST method. However, a phylogenetic analysis of the in-frame concatenated MLST alleles, as illustrated in figure 1, can be used to readily identify novel organisms. Thus, the molecular and genomic data presented in this study can be used to clarify the taxonomic relationships within the *C. lanienae* clade.

Members of the *C. lanienae* clade can be distinguished from other members of the *C. fetus* group by at least four characteristics. The first two are a reduced gene content and maintenance of a different class of CRISPR/Cas loci. Furthermore, the two significant differences are an expanded suite of flagellin genes and the complete absence of an encoded selenoproteome. *C. lanienae* clade genomes encode between four and ten flagellin genes with a wide variety of alleles identified within the cluster. However, even though a larger number of flagellin genes are contained with each genome, no additional or novel flagellar or flagellar modification genes were identified in these genomes. Therefore, even though the flagellin subunits may be more diverse, both their glycosylation and the flagellar machinery, per se, is expected to be highly similar to the flagella characterized previously in *Campylobacter*. The absence of an encoded selenoproteome is definitely a unique feature of the *C. lanienae* clade. All aspects of selenium metabolism, including SeC coding, selenoproteins, and SeC modification of tRNA, are missing from all 76 completed or draft *C. lanienae* clade genomes. The evolution of such selenium-negative taxa cannot be readily explained, but further reinforces the common ancestry of the *C. lanienae* cluster. The effect of any or all of these differences on the biology of the *C. lanienae* clade organisms is unknown. Putative strain misidentification and the generally sparse amount of epidemiological and molecular data present at

this time would confound any definitive conclusions. However, the data presented here can be used for future investigations into this unique *Campylobacter* taxonomic cluster.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

This work was funded by the United States Department of Agriculture, Agricultural Research Service, CRIS projects 2030-42000-230-047 and 2030-42000-230-051, and Food Standards Scotland project S14054.

Literature Cited

Ali A, et al. 2012. *Campylobacter fetus* subspecies: comparative genomics and prediction of potential virulence targets. Gene 508:145–156.

Barrangou R, et al. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science 315:1709–1712.

Baumler A, Fang FC. 2013. Host specificity of bacterial pathogens. Cold Spring Harb Perspect Med 3:a010041.

Besemer J, Borodovsky M. 2005. GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses. Nucleic Acids Res. 33:W451–W454.

Blaser MJ, Newell DG, Thompson SA, Zechner EL. 2008. Pathogenesis of *Campylobacter fetus*. In: Nachamkin I, Szymanski CM, Blaser MJ, editors. Campylobacter. Washington (DC): ASM Press. p. 401–428.

Carbonero A, et al. 2014. *Campylobacter* infection in wild artiodactyl species from southern Spain: occurrence, risk factors and antimicrobial susceptibility. Comp Immunol Microbiol Infect Dis. 37:115–121.

Chylinski K, Le Rhun A, Charpentier E. 2013. The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. RNA Biol. 10:726–737.

Cooley MB, et al. 2013. Development of a robust method for isolation of shiga toxin-positive *Escherichia coli* (STEC) from fecal, plant, soil and water samples from a leafy greens production region in California. PLoS One 8:e65716.

Cravedi P, Mori G, Fischer F, Percudani R. 2015. Evolution of the selenoproteome in *Helicobacter pylori* and Epsilonproteobacteria. Genome Biol Evol. 7:2692–2704.

Debruyne L, Gevers D, Vandamme P. 2008. Taxonomy of the family *Campylobacteraceae*. In: Nachamkin I, Szymanski CM, Blaser MJ, editors. Campylobacter. Washington (DC): ASM Press. p. 3–25.

Deshpande NP, et al. 2011. Sequencing and validation of the genome of a *Campylobacter concisus* reveals intra-species diversity. PLoS One 6:e22170.

Fitzgerald C, et al. 2014. *Campylobacter fetus* subsp. *testudinum* subsp. nov., isolated from humans and reptiles. Int J Syst Evol Microbiol. 64:2944–2948.

Fomenko DE, Xing W, Adair BM, Thomas DJ, Gladyshev VN. 2007. High-throughput identification of catalytic redox-active cysteine residues. Science 315:387–389.

Forchhammer K, Boesmiller K, Bock A. 1991. The function of selenocysteine synthase and SELB in the synthesis and incorporation of selenocysteine. Biochimie 73:1481–1486.

Garneau JE, et al. 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 468:67–71.

Gebhart CJ, Edmonds P, Ward GE, Kurtz HJ, Brenner DJ. 1985. "*Campylobacter hyointestinalis*" sp. nov.: a new species of *Campylobacter* found in the intestines of pigs and other animals. J Clin Microbiol. 21:715–720.

Gilbert MJ, Kik M, Miller WG, Duim B, Wagenaar JA. 2015. *Campylobacter iguaniorum* sp. nov., isolated from reptiles. Int J Syst Evol Microbiol. 65:975–982.

Gilbert MJ, et al. 2013. Complete genome sequence of *Campylobacter fetus* subsp. *testudinum* Strain 03-427T. Genome Announc. 1:e01002-13.

Gilbert MJ, et al. 2014. Complete genome sequence of *Campylobacter iguaniorum* strain 1485ET, isolated from a Bearded Dragon (*Pogona vitticeps*). Genome Announc. 2: e00844-14.

Goris J, et al. 2007. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol. 57:81–91.

Gorkiewicz G, Feierl G, Zechner R, Zechner EL. 2002. Transmission of *Campylobacter hyointestinalis* from a pig to a human. J Clin Microbiol. 40:2601–2605.

Grissa I, Vergnaud G, Pourcel C. 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res. 35:W52–W57.

Guevremont E, Lamoureux L, Loubier CB, Villeneuve S, Dubuc J. 2014. Detection and characterization of *Campylobacter* spp. from 40 dairy cattle herds in Quebec, Canada. Foodborne Pathog Dis. 11:388–394.

Hakkinen M, Heiska H, Hanninen ML. 2007. Prevalence of *Campylobacter* spp. in cattle in Finland and antimicrobial susceptibilities of bovine *Campylobacter jejuni* strains. Appl Environ Microbiol. 73:3232–3238.

Hatfield DL, Gladyshev VN. 2002. How selenium has altered our understanding of the genetic code. Mol Cell Biol. 22:3565–3576.

Hille F, Charpentier E. 2016. CRISPR-Cas: biology, mechanisms and relevance. Philos Trans R Soc Lond B Biol Sci. 371:20150496.

Hyatt D, et al. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119.

Inglis GD, Kalischuk LD, Busz HW. 2004. Chronic shedding of *Campylobacter* species in beef cattle. J Appl Microbiol. 97:410–420.

Inglis GD, Kalischuk LD, Busz HW. 2003. A survey of *Campylobacter* species shed in faeces of beef cattle using polymerase chain reaction. Can J Microbiol. 49:655–661.

Jansen R, Embden JD, Gaastra W, Schouls LM. 2002. Identification of genes that are associated with DNA repeats in prokaryotes. Mol Microbiol. 43:1565–1575.

Jay-Russell MT, Bates A, Harden L, Miller WG, Mandrell RE. 2012. Isolation of *Campylobacter* from feral swine (*Sus scrofa*) on the ranch associated with the 2006 *Escherichia coli* O157:H7 spinach outbreak investigation in California. Zoonoses Public Health 59:314–319.

Kim do K, et al. 2015. *Campylobacter hyointestinalis* isolated from a human stool specimen. Ann Lab Med. 35:657–659.

Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 32:11–16.

Lin J, et al. 2015. Comparative genomics reveals new candidate genes involved in selenium metabolism in prokaryotes. Genome Biol Evol. 7:664–676.

Logan JM, Burnens A, Linton D, Lawson AJ, Stanley J. 2000. *Campylobacter lanienae* sp. *nov*., a new species isolated from workers in an abattoir. Int J Syst Evol Microbiol. 50 Pt 2:865–872.

Logan SM, Schoenhofen IC, Guerry P. 2008. O-linked flagellar glycosylation in *Campylobacter*. In: Nachamkin I, Szymanski CM, Blaser MJ, editors. Campylobacter. Washington (DC): ASM Press. p. 471–481.

Makarova KS, et al. 2011. Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol. 9:467–477.

Merga JY, Winstanley C, Williams NJ, Yee E, Miller WG. 2013. Complete genome sequence of the *Arcobacter butzleri* cattle isolate 7h1h. Genome Announc. 1:e00655-13.

Miller WG, et al. 2012. Multilocus sequence typing methods for the emerging *Campylobacter* species *C. hyointestinalis*, *C. lanienae*, *C. sputorum*, *C. concisus*, and *C. curvus*. Front Cell Infect Microbiol. 2:45.

Miller WG, Yee E. 2015. Complete genome sequence of *Campylobacter gracilis* ATCC 33236T. Genome Announc. 3:e01087-15.

Miller WG, Yee E, Chapman MH. 2016a. Complete genome sequences of *Campylobacter hyointestinalis* subsp. *hyointestinalis* strain LMG 9260 and *C. hyointestinalis* subsp. *lawsonii* strain LMG 15993. Genome Announc. 4:e00665-16.

Miller WG, et al. 2014a. Comparative genomics of the *Campylobacter lari* group. Genome Biol Evol. 6:3252–3266.

Miller WG, Yee E, Huynh S, Chapman MH, Parker CT. 2016b. Complete genome sequence of *Campylobacter iguaniorum* strain RM11343, isolated from an alpaca. Genome Announc. 4:e00646-16.

Miller WG, Yee E, Jolley KA, Chapman MH. 2014b. Use of an improved *atpA* amplification and sequencing method to identify members of the *Campylobacteraceae* and *Helicobacteraceae*. Lett Appl Microbiol. 58:582–590.

Navarro-Gonzalez N, et al. 2014. *Campylobacter* shared between free-ranging cattle and sympatric wild ungulates in a natural environment (NE Spain). Ecohealth 11:333–342.

Oliveira LM, et al. 2016. Complete genome sequence of type strain *Campylobacter fetus* subsp. *fetus* ATCC 27374. Genome Announc. 4:e01344-16.

On SL, Bloch B, Holmes B, Hoste B, Vandamme P. 1995. *Campylobacter hyointestinalis* subsp. lawsonii subsp. nov., isolated from the porcine stomach, and an emended description of *Campylobacter hyointestinalis*. Int J Syst Bacteriol. 45:767–774.

On SLW. 2005. Taxonomy, phylogeny, and methods for the identification of *Campylobacter* species. In: Ketley JM, Konkel ME, editors. Campylobacter: Molecular and Cellular Biology. Norfolk (UK): Horizon Scientific Press. p. 13–42.

Oporto B, Hurtado A. 2011. Emerging thermotolerant *Campylobacter* species in healthy ruminants and swine. Foodborne Pathog Dis. 8:807–813.

Parkhill J, et al. 2000. The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. Nature 403:665–668.

Patrick ME, et al. 2013. Human infections with new subspecies of *Campylobacter fetus*. Emerg Infect Dis. 19:1678–1680.

Pearson BM, Louwen R, van Baarlen P, van Vliet AH. 2015. Differential distribution of Type II CRISPR-Cas systems in agricultural and nonagricultural *Campylobacter coli* and *Campylobacter jejuni* isolates correlates with lack of shared environments. Genome Biol Evol. 7:2663–2679.

Richter M, Rossello-Mora R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci U S A. 106:19126–19131.

Salihu MD, et al. 2009. Isolation and prevalence of *Campylobacter* species in cattle from Sokoto state, Nigeria. Vet Ital. 45:501–505.

Samosornsuk W, et al. 2015. Isolation and characterization of *Campylobacter* strains from diarrheal patients in central and suburban Bangkok, Thailand. Jpn J Infect Dis. 68:209–215.

Sasaki Y, et al. 2013. Prevalence and antimicrobial susceptibility of food-borne bacteria in wild boars (*Sus scrofa*) and wild deer (*Cervus nippon*) in Japan. Foodborne Pathog Dis. 10:985–991.

Schoenhofen IC, Vinogradov E, Whitfield DM, Brisson JR, Logan SM. 2009. The CMP-legionaminic acid pathway in *Campylobacter*: biosynthesis involving novel GDP-linked precursors. Glycobiology 7:715–725.

Schweitzer N, et al. 2011. Molecular characterization of *Campylobacter lanienae* strains isolated from food-producing animals. Foodborne Pathog Dis. 8:615–621.

Serraino A, et al. 2013. Presence of *Campylobacter* and *Arcobacter* species in in-line milk filters of farms authorized to produce and sell raw milk and of a water buffalo dairy farm in Italy. J Dairy Sci. 96:2801–2807.

Stynen AP, et al. 2011. Complete genome sequence of type strain *Campylobacter fetus* subsp. *venerealis* NCTC 10354T. J Bacteriol. 193:5871–5872.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol. 30:2725–2729.

Tasaki E, et al. 2012. Molecular identification and characterization of clustered regularly interspaced short palindromic repeats (CRISPRs) in a urease-positive thermophilic *Campylobacter* sp. (UPTC). World J Microbiol Biotechnol. 28:713–720.

Turowski EE, et al. 2014. Isolation of a *Campylobacter lanienae*-like bacterium from laboratory chinchillas (*Chinchilla laniger*). Zoonoses Public Health 61:571–580.

van Bergen MA, et al. 2005. Clonal nature of *Campylobacter fetus* as defined by multilocus sequence typing. J Clin Microbiol. 43:5888–5898.

van der Graaf-van Bloois L, et al. 2014a. First closed genome sequence of *Campylobacter fetus* subsp. *venerealis* bv. intermedius. Genome Announc. 2:e01246-13.

van der Graaf-van Bloois L, et al. 2014b. Inconsistency of phenotypic and genomic characteristics of *Campylobacter fetus* subspecies requires reevaluation of current diagnostics. J Clin Microbiol. 52:4183–4188.

Wittwer AJ. 1983. Specific incorporation of selenium into lysine- and glutamate-accepting tRNAs from *Escherichia coli*. J Biol Chem. 258:8637–8641.

Wolfe MD, et al. 2004. Functional diversity of the rhodanese homology domain: the *Escherichia coli ybbB* gene encodes a selenophosphate-dependent tRNA 2-selenouridine synthase. J Biol Chem. 279:1801–1809.

Zinoni F, Birkmann A, Leinfelder W, Bock A. 1987. Cotranslational insertion of selenocysteine into formate dehydrogenase from *Escherichia coli* directed by a UGA codon. Proc Natl Acad Sci U S A. 84:3156–3160.

Zinoni F, Heider J, Bock A. 1990. Features of the formate dehydrogenase mRNA necessary for decoding of the UGA codon as selenocysteine. Proc Natl Acad Sci U S A. 87:4660–4664.

Associate editor: John McCutcheon