Genomic Evidence for the Emergence and Evolution of Pathogenicity and Niche Preferences in the Genus *Campylobacter*

Gregorio Iraola[1,2,*], Ruben Pérez[2], Hugo Naya[1,3], Fernando Paolicchi[4], Eugenia Pastor[1], Sebastián Valenzuela[1], Lucía Calleros[2], Alejandra Velilla[4], Martín Hernández[2], and Claudia Morsella[4]

[1]Unidad de Bioinformática, Institut Pasteur Montevideo, Montevideo, Uruguay

[2]Sección Genética Evolutiva, Facultad de Ciencias, Montevideo, Uruguay

[3]Departamento de Producción Animal y Pasturas, Facultad de Agronomía, Montevideo, Uruguay

[4]Laboratorio de Bacteriología, Unidad Integrada INTA-Universidad Nacional de Mar del Plata, Balcarce, Argentina

*Corresponding author: E-mail: giraola@pasteur.edu.uy.

Accepted: September 1, 2014

Data deposition: *Campylobacter sputorum* bv. sputorum strain INTA08/209 genomic sequences have been deposited at GenBank/DDBJ/EMBL under the accession JMTI00000000.

Abstract

The genus *Campylobacter* includes some of the most relevant pathogens for human and animal health; the continuous effort in their characterization has also revealed new species putatively involved in different kind of infections. Nowadays, the available genomic data for the genus comprise a wide variety of species with different pathogenic potential and niche preferences. In this work, we contribute to enlarge this available information presenting the first genome for the species *Campylobacter sputorum* bv. sputorum and use this and the already sequenced organisms to analyze the emergence and evolution of pathogenicity and niche preferences among *Campylobacter* species. We found that campylobacters can be unequivocally distinguished in established and putative pathogens depending on their repertory of virulence genes, which have been horizontally acquired from other bacteria because the nonpathogenic *Campylobacter* ancestor emerged, and posteriorly interchanged between some members of the genus. Additionally, we demonstrated the role of both horizontal gene transfers and diversifying evolution in niche preferences, being able to distinguish genetic features associated to the tropism for oral, genital, and gastrointestinal tissues. In particular, we highlight the role of nonsynonymous evolution of disulphide bond proteins, the invasion antigen B (CiaB), and other secreted proteins in the determination of niche preferences. Our results arise from assessing the previously unmet goal of considering the whole available *Campylobacter* diversity for genome comparisons, unveiling notorious genetic features that could explain particular phenotypes and set the basis for future research in *Campylobacter* biology.

Key words: *Campylobacter*, pathogenicity, niche preferences, *Campylobacter sputorum*, comparative genomics.

Introduction

Members of the genus *Campylobacter* are ecologically diverse and naturally inhabit a wide variety of mammals, birds, and reptiles; and some species have an outstanding role in human and animal health (Nachamkin et al. 2008). The renewed effort in bacterial characterization driven by the advent and consolidation of next-generation sequencing technologies has brought back historically underestimated *Campylobacter* species that nowadays pose an emerging risk of zoonotic transmission as they are found in companion, farm and wild animals, and can also contaminate food (Wagenaar et al. 2014).

The first *Campylobacter* species whose genome was completely sequenced was the foodborne pathogen *Campylobacter jejuni*, the leading cause of gastrointestinal illness in developed and developing countries. This species is subdivided in *C. jejuni* subsp. *jejuni* and *C. jejuni* subsp. *doylei*, which differ in their genomes and pathogenic characteristics (Parker et al. 2007). Posterior efforts achieved the sequencing of *Campylobacter coli*, another established

pathogen that accounts for 1–25% of *Campylobacter*-related diarrheal diseases (Gurtler et al. 2005). Both *C. jejuni* and *C. coli* are probably the most studied species of the genus and have been subject of continuous research focused on the elucidation of their pathogenic and ecological features from a genomics perspective (Sheppard, Didelot, Jolley, et al. 2013; Sheppard, Didelot, Meric, et al. 2013; Skarp-de Haan et al. 2014). The species *Campylobacter upsaliensis* and *Campylobacter lari* complete a list of fully sequenced, gastrointestinal campylobacters that are established human pathogens (Labarca et al. 2002; Hayashi et al. 2012). Even though most of the species described above historically deserved great attention due to their importance in human health, when the genus *Campylobacter* was first proposed in 1963 it only included two species: *Campylobacter fetus* and "*Campylobacter bubulus*", both isolated from genital tissues of cattle (Sebald and Véron 1963). The species *C. fetus* is now recognized as an established pathogen with a well-known incidence in the reproductive health of cattle and sheep, causing significant socioeconomic burden worldwide (Mshelia et al. 2010). Furthermore, *C. fetus* is subdivided in two subspecies, whose genomes have been also sequenced: *C. fetus* subsp. *venerealis*, a bovine-exclusive pathogen that colonizes the genital tract causing infertility and abortion (van Bergen et al. 2005); and *C. fetus* subsp. *fetus*, which not only causes genital infections in cattle and sheep but also is associated to bacteremia in humans (Wagenaar et al. 2014). By the contrary, "*C. bubulus*" has deserved less attention because it is just a commensal bacterium typically found in the genital tissues of healthy cattle (Sebald and Véron 1963; Skirrow 1994), although some authors have remarked its role as a putative pathogen causing sporadic infections in humans (On et al. 1992; Tee et al. 1998). To date, *C. bubulus* strains have not been fully sequenced and released. Because the taxonomic structure of the genus has changed extensively, *C. bubulus* was renamed as *Campylobacter sputorum* and divided intraspecifically in three biovars (bv. sputorum, bv. fecalis, and bv. paraureolyticus), in accordance with their biochemical behavior (On et al. 1998).

Six additional species fill the list of fully sequenced organisms: *Campylobacter concisus*, *Campylobacter curvus*, *Campylobacter rectus*, *Campylobacter hominis*, *Campylobacter showae,* and *Campylobacter gracilis*. As well as *C. sputorum*, these species are remarked as putative pathogens as their prevalences in human or animal infections are widely variable and their clinical presentations are less clear; all of them have been reported at least once causing different kinds of infections (Man 2011). In particular, *C. rectus* and *C. gracilis* have been associated with periodontal diseases and infections in the oral cavity; however, despite a pathogenic role can be suspected for these species robust evidence of causality is still scarce (Siqueira and Rocas 2002). The species *C. concisus*, *C. curvus*, and *C. showae* have been related to the oral cavity too; nevertheless their role in clinical cases is even less clear.

In summary, the species belonging to the genus *Campylobacter* present a wide phenotypic variability. Based on their pathogenic potential and clinical presentations they can be clearly divided in established and putative pathogens and, although most species can be isolated from different hosts and tissues (niches), they can be grouped in gastrointestinal, oral, and genital depending on their main source of isolation (Man 2011). Nowadays, the number of publicly available genomes for *Campylobacter* species and subspecies are a good representative of the explained phenotypic diversity, allowing to study the relationship between genomic variability, pathogenic potential, and niche preferences. In this work, we report the first completed genome and characterization for *C. sputorum* bv. sputorum strain INTA 08/209, a natural isolation from semen of a healthy bull. We then use this and a representative set of publicly available *Campylobacter* genomes to analyze the emergence and evolution of pathogenicity in this genus using genome-wide comparative analyses and phylogenetics. Furthermore, we focused on the comparison of *Campylobacter* species that are able to colonize different tissues, in order to determine genetic features associated with niche preferences. Our findings suggest that the emergence of pathogenicity can be correlated to the acquisition of virulence genes through horizontal gene transfers from other bacteria and posterior interchange between some members of the genus. Niche preferences can be mostly explained by nonsynonymous evolution of DSB (disulphide bond) proteins and the invasin CiaB, as well as global compositional differences in GC content, genomes size, and secretomes. In this article, we provide the first comparative genomics analysis of a representative sample of *Campylobacter* taxonomic diversity, pointing the essentiality of sequencing nonclassical organisms to obtain information about the evolutionary mechanisms governing bacterial genomes.

Materials and Methods

Bacterial Strains, Sequencing, and Assembly

The strain INTA 08/209 was isolated from the semen of a healthy bull in Argentina in 2008. Samples were inoculated in Skirrow agar and the isolated colonies were classified as *Campylobacter sputorum* bv. sputorum because they were unable to produce catalase and urease, and they were unable to grow in a 1% glycine broth and did produce H_2S. For further confirmation, a fragment of rRNA 16S gene was polymerase chain reaction-amplified using previously described primers and conditions (Linton et al. 1996) and compared with the 16S ribosomal RNA sequences available in GenBank using BLASTN.

Genomic DNA was isolated with the Wizard Genomic DNA purification kit (Promega), sequencing was performed on an Illumina Hi-Seq 2000 platform and generated 9.617.780 paired-end reads (2 × 100 cycles). The resulting reads were first corrected using ALLPATHS-LG (Butler et al. 2008) and then assembled with Velvet software (Zerbino and Birney 2008), PAGIT toolkit (Swain et al. 2012) was used for postassembly improvement, and the final assembly quality was evaluated with ALE (Clark et al. 2013). The resulting contigs were automatically annotated with RAST (Aziz et al. 2008) and manually curated with Pfam and BLASTP over the nr database.

The genomes and available annotations for *C. concisus* 13826, *C. curvus* 525-92, *C. fetus* subsp. *fetus* 82-40, *C. fetus* subsp. *venerealis* NCTC10354, *C. hominis* ATCC BAA-381, *C. jejuni* subsp. *jejuni* RM1221, *C. jejuni* subsp. *jejuni* 55037, *C. jejuni* subsp. *jejuni* 129-258, *C. jejuni* subsp. *jejuni* 51494, *C. jejuni* subsp. *jejuni* LMG9879, *C. jejuni* subsp. *jejuni* LMG9217, *C. jejuni* subsp. *jejuni* LMG23218, *C. jejuni* subsp. *jejuni* 2008-872, *C. jejuni* subsp. *doylei* 269-97, *C. lari* RM2100, *C. gracilis* RM32668, *C. showae* RM3277, *C. rectus* RM3267, *C. upsaliensis* RM3195, *C. coli* 76339, *C. coli* BIGS0010, and *C. coli* RM2228 were retrieved from the NCBI. For *C. coli* we considered intraspecific diversity by analyzing representatives from clades 1, 2, and 3 described in Sheppard, Didelot, Jolley, et al. (2013). For *C. jejuni* we included representative genomes from the main clonal complexes described in Sheppard, Didelot X, Meric, et al. (2013). When not available, annotations were generated with RAST. All plots, graphics and data analysis were generated using in-house R scripts.

Analysis of Orthologous Groups, Virulence Factors, and Gene Ontologies

The best reciprocal hit approach using BLASTP was implemented to recover shared genes for each pair of genomes, for each reciprocal hit with query coverage >95% and identity >50%. This analysis was complemented running OrthoMCL (Li et al. 2003) with default parameters. The set of virulence-associated genes of each genome was recovered using BLASTP against an in-house database created from the virulence factors database (Chen et al. 2005) and the nr database, and Fisher's exact test was conducted to determine which virulence genes had significant differences among established and putative pathogens ($P < 0.01$).

For functional annotations, each proteome was analyzed with BLAST2GO (Conesa et al. 2005) and gene ontology (GO) terms (belonging to Biological Process) frequency distributions were used to implement the nonparametric Kruskal–Wallis test of variance ($P < 0.01$) in order to identify enriched gene functions among oral versus nonoral, genital versus nongenital, and established versus putative. These groups were defined based on the predominant source of isolation for each species and their pathogenic characteristics (table 1).

Phylogenetic Inferences, Ancestral Reconstruction, and Selection Analyses

The consensus phylogeny for *Campylobacter* genomes was obtained from 16S genes aligned with MUSCLE (Edgar 2004) and using Neighbor-Joining and maximum-likelihood (ML) methods; and from the full proteome, using an alignment-free method based on protein feature frequency profiles (Jun et al. 2010). In order to check if inferred phylogenetic relationships were product of bias in taxonomic sampling, a tree was constructed with 16S genes for all *Campylobacter* species available in SILVA (Pruesse et al. 2007) (as author request). In all cases, bootstrap analysis was performed with 1,000 resampled data sets. To infer phylogenetic relationships from genes selected for further discussion (i.e., *dsbA*), protein sequences were aligned with MUSCLE and ML method was used to build the trees under the Jones–Taylor–Thornton substitutions model. When required, sequences from *Arcobacter butzleri* RM4018, *Arcobacter nitrofigilis* DSM 7299, *Sulfurospirillum barnessi* SES3, and *Sulfurospirillum deleyianum* DSM 6946 were used as outgroups. For each virulence gene, the presence (1) or absence (0) was established as a discrete state in each genome. This information was used to reconstruct the internal nodes states over the consensus phylogeny using the ML method implemented in APE (Paradis et al. 2004).

Screening for diversifying selection over alignment positions was implemented with mixed effects model of evolution with default parameters, due to its ability of detecting episodic selection (Murrell et al. 2012). For finding conserved positions in alignments among organisms sharing a particular niche preference, we generated in-house R scripts for counting shared positions between them which differ in the rest and created a null distribution counting the same for all possible random groups of genomes.

Secretome Analysis

Secreted proteins were predicted for all proteomes using the default settings for Gram-negative bacteria on the SignalP Server 4.1 (Petersen et al. 2011). Nonclassical secreted proteins for the SignalP were predicted using SecretomeP 2.0 Server (Bendtsen et al. 2005). The correspondence analysis between amino acids usage and niche preferences was performed using "seqinr" (Charif and Lobry 2007) and "ca" (Nenadic and Greenacre 2007) packages in R.

Results and Discussion

Campylobacter sputorum Genome Overview

The complete chromosome of *C. sputorum* bv. sputorum strain INTA 08/209 was assembled into 34 contigs of 150-fold in average coverage and 52.394 bp in average length (maximum contig length was 407.694 bp). The estimated chromosome size resulted to be 1.781.420 bp. with an

Table 1
Description of Analyzed Genomes

Species	Clade/CC	Accession	Size	GC	Genes	Pathogenicity	Niche
Campylobacter coli RM2228	Clade 1	AAFL01	1.86	31.1	1,967	Established	Gastrointestinal
C. coli BIGS0010	Clade 2	ANGU00	1.66	31.5	1,665	Established	Gastrointestinal
C. coli 76339	Clade 3	HG326877	1.58	32	1,556	Established	Gastrointestinal
Campylobacter jejuni 2008-872	61	AIOR00	1.6	30.4	1,702	Established	Gastrointestinal
C. jejuni LMG23218	48	AIOB01	1.6	30.4	1,734	Established	Gastrointestinal
C. jejuni 51494	353	AINZ00	1.8	30.2	1,975	Established	Gastrointestinal
C. jejuni LMG9217	443	AIOO01	1.6	30.3	1754	Established	Gastrointestinal
C. jejuni LMG9879	21	AIOI01	1.6	30.4	1,734	Established	Gastrointestinal
C. jejuni 129-258	42	AINY01	1.6	30.5	1,679	Established	Gastrointestinal
C. jejuni 55037	45	AIOH01	1.59	30.5	1,666	Established	Gastrointestinal
C. jejuni jejuni RM1221	354	NC_003912	1.78	30.5	1,838	Established	Gastrointestinal
C. jejuni doylei 269.97	—	NC_009707	1.85	30.6	1,731	Established	Gastrointestinal
Campylobacter upsaliensis RM3195	—	AAFJ01	1.77	34.2	1,934	Established	Gastrointestinal
Campylobacter lari RM2100	—	NC_001239	1.57	29.6	1,544	Established	Gastrointestinal
Campylobacter fetus fetus 82-40	—	NC_008599	1.77	33.3	1,719	Established	Genital
C. fetus venerealis NCTC10354	—	AFGH01	1.87	33.2	1,718	Established	Genital
Campylobacter sputorum INTA08/209	—	JMTI0	1.78	29	1,869	Putative	Genital
Campylobacter showae RM3277	—	ACVQ01	2.07	45.7	2,361	Putative	Oral
Campylobacter gracilis RM3268	—	ACYG01	2.26	46.6	2,847	Putative	Oral
Campylobacter hominis ATCC BAA-381	—	NC_009714	1.71	31.7	1,687	Putative	Gastrointestinal
Campylobacter rectus RM3267	—	ACFU01	2.51	44.8	2,971	Putative	Oral
Campylobacter concisus 13826	—	NC_009802	2.1	39.2	1,989	Putative	Oral
Campylobacter curvus 525.92	—	NC_009715	1.97	44.5	1,934	Putative	Oral

average GC content of 29%. The chromosome contained 1.869 predicted protein-coding genes, 3 rRNA operons, and 42 tRNA genes. The genomic sequences were deposited in the GenBank database under accession number JMTI00000000.

The genome of *C. sputorum* bv. sputorum had the lowest GC content for a *Campylobacter* species reported so far. This feature is probably reflecting a host-associated lifestyle which tends to gradually lower GC content driven by adaptive evolution (Rocha and Danchin 2002). *Campylobacter sputorum* bv. sputorum also showed a reduced chromosome size in comparison with other members of the genus and presented the lowest synteny conservation among campylobacters, evidencing that sequence rearrangements have been shaping its genomic architecture, as expected for bacteria which have suffered host-restriction processes (Moran 2002, Moran and Plague 2004; supplementary fig. S1, Supplementary Material online). The species *C. sputorum* bv. sputorum had 181 unique protein-coding genes in comparison with their *Campylobacter* relatives, indicating that the pan-genome of this genus is still open and will probably increase as new complete genomes become available. Moreover, 83 out of these 181 genes exclusive for *C. bubulus* were not found in protein databases (ORFans). Previous analyses have reported that 20–30% of genes present in a novel genome may be ORFans (Medini et al. 2005), for *C. sputorum* bv. sputorum this percentage is quite lower (4%) probably reflecting a reductional process in the dispensable genes set for this species. The SAP (surface array protein) genes, which are the main antigenic determinants for the genital species *C. fetus*, were not present in *C. sputorum*. The genome of *C. sputorum* bv. sputorum presented no evidence for extrachromosomal replicons. Nevertheless, signatures for horizontal gene transfers were identified as some genes were found on other bacterial genomes whereas absent in campylobacters, for example, one gene coding for a putative hemolysin was probably acquired from *Wolinella succinogenes* and suffered posterior pseudogenization by nonsense mutations, showing that this kind of evolutionary process has also been shaping the genomic landscape of *C. sputorum* bv. sputorum. Additionally, two contiguous genes coding for an AAA+ ATPase and a restriction endonuclease were probably acquired from the gram-positive coccus *Eremococcus coleocola*, a relatively new species originally isolated from vaginal tissues from horses (Collins et al. 1999). This result provides strong evidence for the recent acquirement of new genes from bacteria that share the niche with *C. sputorum* bv. sputorum, even being phylogenetically distant. With the incorporation of this newly sequenced species we recalculated the core genome for *Campylobacter* genus, estimated in 669 genes.

Evolution of Pathogenicity

To study how the pathogenic character has evolved along the *Campylobacter* taxonomy we used 23 available genomes predefined in two groups, considering their documented clinical incidence: 1) *C. coli* strains, *C. jejuni* subsp. *jejuni* strains, *C. jejuni* subsp. *doylei*, *C. lari*, *C. upsaliensis*, *C. fetus* subsp.

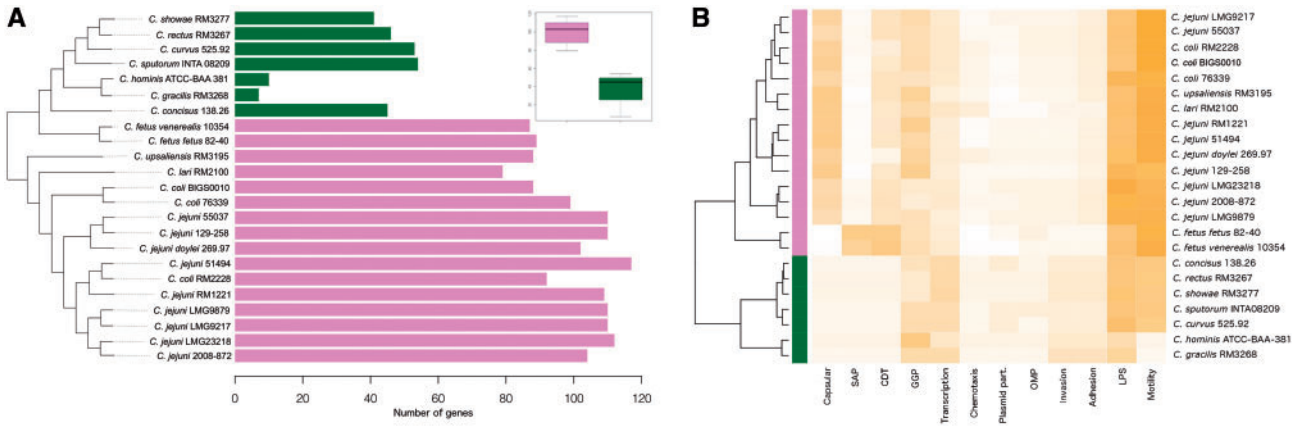

Fig. 1.—Barplot and heatmap of virulence genes identified per genome. Established pathogens are displayed in violet, and putative pathogens are displayed in dark green. (*A*) The total number of virulence genes is displayed as bar lengths. Genomes are clustered based on the inferred consensus phylogeny for *Campylobacter* genus. (*B*) Genomes are clustered in established (top) and putative (bottom) pathogens based on the presence/absence patterns for virulence genes belonging to 12 functional categories. Colors (white to orange) represent the number of genes.

Fetus, and *C. fetus* subsp. *venerealis* represent established pathogens in human and/or cattle, and 2) *C. sputorum*, *C. curvus*, *C. concisus*, *C. hominis*, *C. rectus*, *C. showae,* and *C. gracilis* represent putative pathogens. Based on this classification, we looked for the presence or absence of virulence genes on each genome and constructed a presence/absence matrix considering 255 different genes belonging to the following functional categories: Capsular, general glycosylation pathway, SAP, CDT (cytolethal distending toxin), transcription, chemotaxis, plasmid partitioning, outer membrane protein, invasion, adhesion, LPS (lipopolysaccharide), and motility. Figure 1*A* shows the number of virulence genes identified per genome displayed in accordance with a consensus phylogeny (see Materials and Methods for details). Established pathogens posed a significantly wider repertory of virulence genes ($P = 0.0002$, Fisher's exact test) which also correlated with their phylogenetic position. The unique exception was for *C. fetus* subspecies, that are phylogenetically closer to putative pathogens but have an expanded repertory of virulence genes as expected for established pathogens. Taking into account the number of genes belonging to each functional category, species clustered in two distinctive groups that match perfectly with the predefined established and putative pathogens (fig. 1*B*). Established pathogens were richer in genes coding for LPS, adhesion and, motility, although, the presence of capsular genes (SAP for *C. fetus* subspecies) and CDT were the most relevant features that distinguished established from putative pathogens ($P = 0.001$, Fisher's exact test). In particular, some authors have questioned the role of CDT as a virulence factor because some naturally occurring *C. jejuni* strains presented partial disruption or absence of CDT operon. As a complementary approach, we screened 85 additional *C. jejuni* genomes and found a prevalence of 90% for *cdtA*, 97% for *cdtB* (the main toxin component), and 98% for *cdtC*. The analysis of several publications (Asakura et al. 2007; Talukder et al. 2008; Ripabelli et al. 2010; Quetz et al. 2012) screening CDT genes in clinical cases or *C. jejuni* populations also showed a prevalence higher than 95%. Moreover, the insertional inactivation or complete deletion of *C. jejuni* CDT genes has been demonstrated to cause reduced invasiveness and adherence, attenuation and asymptomatic infections (Purdy et al. 2000; Jain et al. 2008). These results support the role of CDT as a virulence factor, which probably has a complementary activity with other virulence factors like motility proteins, adhesions, and invasins. However, the occurrence of CDT-negative strains isolated from clinical cases remains as an open question, in spite of being the vast minority. In summary, these results point that pathogenic potential of *Campylobacter* species may be correlated with the presence of certain virulence genes. However, considering the extreme complexity in defining bacterial pathogenicity these results are useful for suggesting general differences among established and putative pathogens, whereas deeper analyses involving experimental approaches should be conducted in the future to decipher the role of these genes during infection.

In order to elucidate the most probable evolutionary path that lead to the actual distribution of these genes among taxa, we implemented an ancestral character reconstruction approach using the presence or absence of virulence genes as current states. In first place, the MRCA (most recent common ancestor) for *Campylobacter* was probably a nonpathogenic or putative pathogen that lacked the vast majority of virulence genes (average probability for gene presence < 0.5) (fig. 2*A*). However, some virulence genes, such as the invasion antigen B (CiaB), were also found in the *Campylobacter* ancestor, which suggests that ancient campylobacters had the potential to invade host cells. The relevance of CiaB in pathogenic phenotypes will be further discussed afterwards.

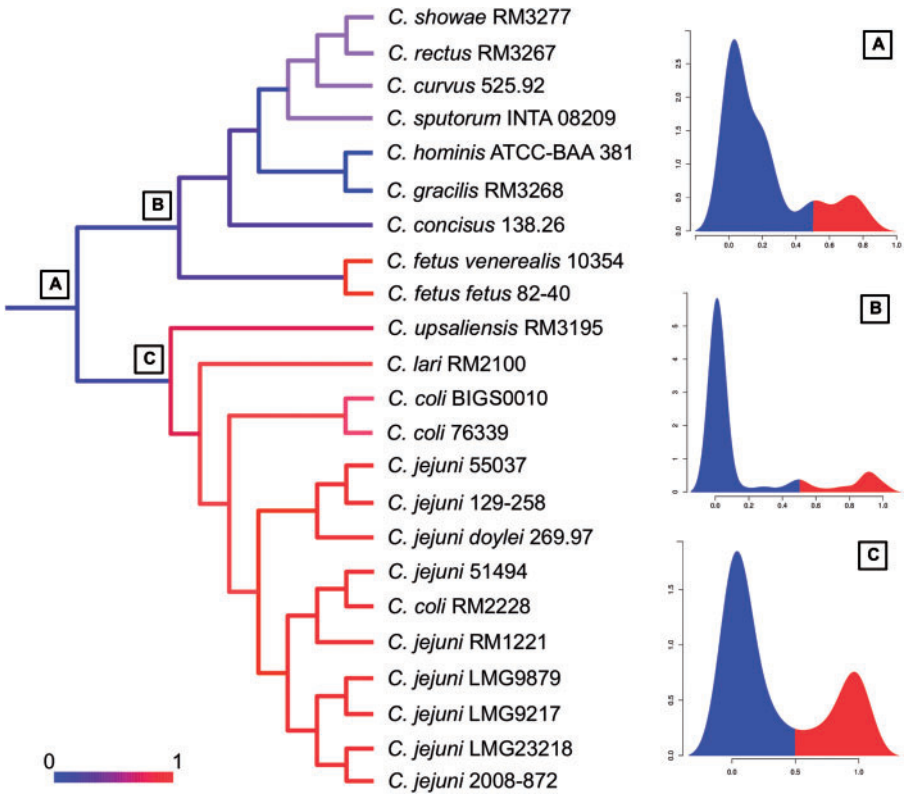

Fig. 2.—Ancestral character reconstruction. The consensus phylogeny for *Campylobacter* is colored according to the average probability for the absence ($P = 0$, blue) or the presence ($P = 1$, red) of virulence genes. The probability densities using the inferred states for each genes are shown for the *Campylobacter* MRCA (*A*), the MRCA for putative pathogens (*B*), and the MCRA for established pathogens (*C*).

The MRCA for putative pathogens showed a probability distribution for genes presence that resembled the *Campylobacter* MRCA (fig. 2*B*), indicating that the reduced amount of virulence genes present in putative pathogens were present in the ancestor or have been probably acquired through recent horizontal gene transfer events. On the contrary, the MRCA for established pathogens already carried genes for CDT, capsule, motility, LPS, and adhesion ($P > 0.9$), suggesting that these organisms have evolved from an ancestor with a significant virulence armament, acquired from other bacteria (fig. 2*C*). It is worth mentioning that no significant differences were found among *C. coli* strains belonging to clades 1, 2, and 3. Recent works have shown that *C. coli* clade 1 (the most frequently isolated from clinical cases) has suffered a progressive genomic introgression with *C. jejuni*, whereas clades 2 and 3 are mainly constituted by nonintrogressed isolates from riparian environments (Sheppard, Didelot, Jolley, et al. 2013). However, strains belonging to clades 2 and 3 have been also found in clinical cases and have genes associated with infection (Skarp-de Haan et al. 2014). Based on these results, it is probable that *C. coli* strains have evolved from the same pathogenic ancestor and shaped their genomes for environmental diversification while conserving genes for CDT, capsule, motility, LPS, or adhesion, being currently underreported in clinical cases due to niche separation. The analysis of *C. jejuni* strains belonging to different clonal complexes also revealed no significant differences in their repertories of virulence genes, showing that intraspecific diversification may be linked to the evolution of different genomic components.

Notoriously, *C. fetus* subspecies should be classified as established pathogens based on their virulence genes repertories and clinical presentations; however, they are phylogenetically closer to putative pathogens. The MRCA for *C. fetus* subspecies showed the presence of almost all genes found in the MRCA for established pathogens, suggesting that *C. fetus* subspecies evolved from a nonpathogenic ancestor shared with putative pathogens, but acquired a set of virulence genes from species belonging to established pathogens. This kind of horizontal evolution has been documented through plasmid transfer from *C. jejuni* to *C. fetus*, moreover *C. fetus* genomes show extensive evidence of recent horizontal gene transfer events (Moolhuijzen et al. 2009; Kienesberger et al. 2014). The lack of a genome for a sister species to *C. fetus*, like *Campylobacter hyointestinalis*, prevent more accurate estimates for the acquirement of these genes.

In summary, from the analysis of this set of genomes we can conclude that the most probable scenario for the

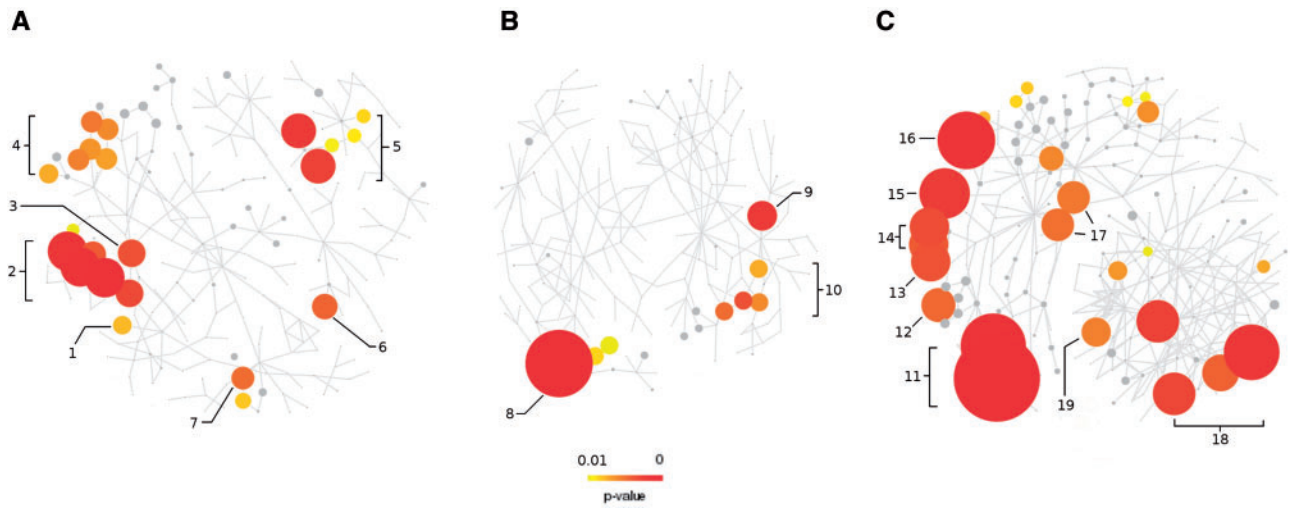

Fig. 3.—GO analysis. Three GO graphs for (*A*) oral versus nonoral, (*B*) genital versus nongenital, and (*C*) established versus putative. Significant GO terms ($P < 0.01$) for each graph are colored in a yellow to red gradient. Numbers encode the name of significant functional categories, (*A*) 1-pathogenesis, 2-response to virus, 3-proteolysis, 4-response to oxygen species, 5-lactate metabolism, 6-sulfate metabolism, 7-antibiotic resistance; (*B*) 8-protein methylation, 9-response to external stimulus, 10-nitrogen utilization; (*C*) 11-response to antibiotics, 12-locomotion, 13-oxydative stress, 14-starvation, 15-nitrogen transport, 16-chromosome partition, 17-adhesion, 18-vitamin biosynthesis, and 19-pathogenesis.

evolution of pathogenicity in *Campylobacter* is the accumulation of virulence factors that resulted in established pathogens, instead of an opposite scenario of pathogenicity attenuation by gene loss from a virulent ancestor. However, posterior gene loss events among putative pathogens should not be discarded, especially for *C. hominis*, which present the smallest virulence armament. Not surprisingly, it was originally isolated from healthy humans (Lawson et al. 1998) and has the lowest number of reported infections among sequenced species. Probably, a more complete representation of *Campylobacter* species could help to better understand the dynamics of horizontal gene transfers and their implications in pathogenicity.

Comparative Functional Analysis

Beyond the set of virulence genes analyzed so far are part of the best-known players in pathogenic phenotypes, the presence, absence or enrichment in other kind of virulence-associated or virulence lifestyle genes (typically coding for more general metabolic pathways) may be directly implied in pathogenicity (Wassenaar and Gaastra 2001). For this reason, we performed a comparative functional analysis of *Campylobacter* proteomes based on GO terms. This approach was also useful to have a first glace of the main metabolic functions associated to niche preferences (fig. 3).

Established pathogens were enriched in functions related to antibiotic resistance (GO:0046677). This feature is particularly interesting because of the central role of antibiotics in the treatment of bacterial infections. The most distinctive feature was the presence of a gene coding for the enzyme aminoglycoside n3-acetyltransferase among established pathogens, whereas absent in all putative pathogens. This enzyme is involved in the resistance to aminoglycosides, which has been extensively proved for *C. jejuni* and *C. coli* (Alfredson and Korolik 2007). The GO analysis also revealed that established pathogens were enriched in terms related to adhesion (GO:0007155) and motility (GO:0040011), in accordance with the results described in the previous section.

In contrast with putative pathogens, all species belonging to established pathogens are able to efficiently invade host cells (Konkel and Joens 1989; Mooney et al. 2003). When bacteria invade eukaryotic cells, they are immediately exposed to unfavorable conditions mainly associated to starvation (nutrients shortage) and multiple types of stresses, most notably oxidative stress (McDougald et al. 2002). Seven genes (coding for thiorredoxin reductase, thiol peroxidase, catalase, methionine sulfoxide reductase, superoxide dismutase, HMBPP reductase, and carbon starvation protein) were involved in the response to starvation and oxidative stress (GO:0006979 and GO:0042594, respectively) and were significantly ($P = 0.005$) more abundant in established pathogens (fig. 4) evidencing that these metabolic functions could be linked to the pathogenic potential.

A particularly relevant feature for species that can colonize genital tissues was the enrichment in genes for nitrogen metabolism (GO:0019740), which is an integrated mechanism that detects the depletion of the primary nitrogen source and activates genes for scavenging and transporting alternative nitrogen sources. There is scarce information about abundance of nutrients in the genital and urogenital tissues; however, it has been demonstrated that uropathogenic *Escherichia coli* strains need to activate nitrogen utilization

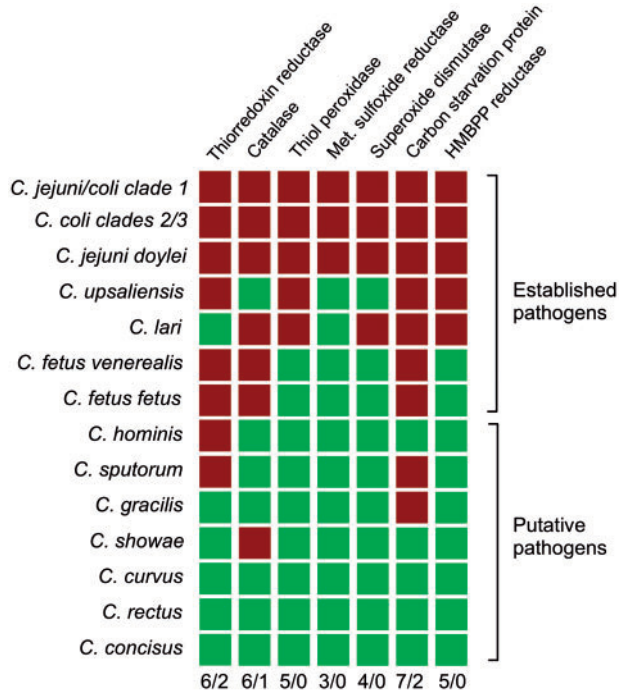

Fig. 4.—Distribution of genes belonging to oxidative stress and starvation. Red boxes show the presence of a gene in a certain genome whereas green boxes show its absence. Fractions at the bottom represent the counting of each gene in established and putative pathogens, respectively.

pathways during colonization of mice urinary tract (Hagan and Mobley 2007). Other genital bacteria, like *Gardnerella vaginalis*, also encode genes important for the utilization of various nitrogen sources (Yeoman et al. 2010) and the pathogen *Candida albicans* (despite nonbacterial) up regulates genes involved in nitrogen utilization when infecting genital tissues (Kumamoto 2008). These results indicate a possible role of nitrogen metabolism on the establishment of microorganisms in the apparently hostile genital environment. Genital campylobacters were also enriched in genes involved in protein methylation (GO:0006479). In general, methylation is involved in cell-environment interactions; however, this characteristic needs to be further investigated in order to establish its relation to genital tropism.

Oral campylobacters are suspected pathogens in periodontal diseases, often presenting a complex etiology mainly attributed to polymicrobial disruption of host homeostasis (Darveau 2010). Recently, the presence of sulfate-reducing bacteria in the complex oral flora has been proposed as implicated in the development of periodontal diseases (Vianna et al. 2008). Accordingly, oral campylobacters resulted to be enriched in functions related to sulfate metabolism (GO:0006790). These species were also enriched in genes for lactate metabolism (GO:0006089), which plays an important role in the development and maintenance of acidic conditions in vivo. Microbial flora present in cariogenic plaques produce lactate as the predominant glucose-derived product, which is considered to be the main acid involved in caries formation (Kara et al. 2006). Because dynamics of periodontal infections are complex, and beyond their direct incidence on oral diseases, the capacity of these *Campylobacter* species to produce lactate may be contributing to the development and establishment of infections, as other microorganisms directly associated to periodontal diseases (like *Streptococcus* or *Veillonella*) are benefited by this lactate-rich environment (McLean et al. 2012).

Secretomes, Compositional Differences, and Selection

To gain more insights on the mechanisms involved in niche preferences among *Campylobacter* species, we centered our attention on their predicted secretomes and the differential amino acids usage within the whole proteomes and the secretomes. Proteins with secretory signals are the main tools that bacteria use to interact with their environments (Desvaux et al. 2009), so secretome evolution may be driven by host-microorganism interactions which are determined by different types of tissue-specific molecules and environmental conditions. Extensive bioinformatics comparative studies of bacterial secretomes have suggested that secretome size is not correlated with pathogenic potential nor niche preferences at highest taxonomic levels (Song et al. 2009). Among *Campylobacter* genomes we found great differences in predicted secretome sizes, ranging from 80 proteins in *C. sputorum* to 210 in *C. gracilis*. Furthermore, when exploring the number of secreted proteins normalized by the species proteome size, we found that oral species secrete around 10% of their proteins, whereas nonoral species secrete around 6% (supplementary fig. S2, Supplementary Material online). These results indicate that particular, oral-exclusive secreted proteins should be playing an important role in niche preference for oral cavity and do not support the findings reported by Song et al. (2009); however, this discrepancy could be explained because the observed signal can be stronger for particular organisms at lower taxonomic levels. Among these oral-exclusive secreted proteins it is worth noting the presence of a divergent kind of metal scavenging TonB-dependent siderophore transporter (TBDT). Metal ions are essential cofactors needed for the correct functioning of most enzymes and bacteria have evolved special macromolecular mechanisms to sequester them from the environment when lacking (Schalk et al. 2011). TonB systems are formed by an energy transduction complex (TonB-ExbB-ExbD) anchored to the inner plasma membrane and a pore-forming TBDT anchored to the outer membrane (Noinaj et al. 2010). These systems have been well-characterized in *C. jejuni* and *C. coli* to a lesser extent, showing redundancy. Here, we found that genes coding for TonB-ExbB-ExbD complex were conserved in all genomes and showed great synteny conservation too. When exploring

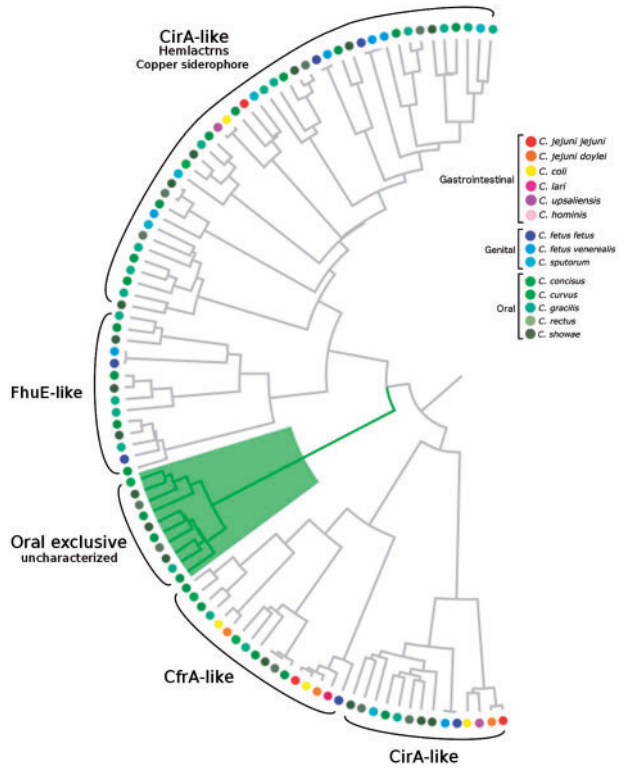

Fig. 5.—Phylogeny of TBDTs. The phylogenetic tree was constructed using 98 TBDT orthologs recovered from *Campylobacter* genomes. The oral-exclusive cluster is highlighted in dark green.

TBDT genes we found great diversity in sequence identity and copy number (up to 21 in *C. curvus*). Figure 5 shows the phylogenetic relationships among 98 recovered orthologs, highlighting the presence of different TBDT types like CirA, FuhE, CfrA, and an uncharacterized oral-exclusive cluster. In terms of genomic context, these oral-exclusive TBDTs were in proximity with genes coding for ATP-binding, permease and periplasmic proteins belonging to iron ABC transporters; a methlytransferase domain protein was habitually found next to the TBDT gene too. In *C. showae* and *C. rectus* we found two adjacent TBDT copies probably generated by paralogy, indeed, in *C. showae*, one of them was pseudogenized by nonsense mutations. Additionally, in all cases the genomic surroundings were rich in small predicted hypothetical proteins, suggesting that these TBDTs are placed in plastic regions suffering rearrangements and horizontal transfer events, as members of this divergent cluster were not found in other campylobacters and presented less than 25% of identity with the rest of TBDTs recovered from *Campylobacter* genomes. No significant differences in secretome size and composition were observed for genital and gastrointestinal campylobacters, suggesting that niche preferences do not depend on the evolution of the same set of genes for adaptation to different environments.

When considering the amino acids usage for secreted proteins, we identified significant differences between *Campylobacter* species belonging to different niches. The correspondence analysis displayed in figure 6*A* demonstrated how amino acids usage clearly discriminates oral, genital, and gastrointestinal species in distinctive clusters. The unique exception was for *C. hominis*, which clustered closer to genital species despite being gastrointestinal; this species carried an extremely reduced virulence genes repertory and posed the lowest number of documented infections. It is probable that these particular features are also being reflected in this discrepancy and further investigation is needed for elucidating why the observed amino acids usage was not correlated with the phenotype of this neglected species. Despite the variability in genome and proteome sizes, the behavior observed in figure 6*A* is maintained when using the whole proteome for amino acids usage calculations, suggesting global reach patterns that link nonsynonymous evolution with niche preferences. The comparison of amino acids usage from different bacteria has showed that adaptive pressures over amino acids are highly variable along taxonomy (Sharp et al. 2005); however, correlations between amino acids usage and niche preference or tissue tropism have been proposed only for viruses (Bahir et al. 2009). We suggest that differences found in amino acids usage among *Campylobacter* species may be attributed to adaptive evolution driven by niche-specific environmental conditions. This is also evident when analyzing the global GC content and genome sizes with respect to different niche preferences, especially for oral species which were distinguished by bigger genomes and higher values of GC content (fig. 6*B*).

In order to further investigate the correlation between secretome evolution and niche preferences we analyzed the genetic variability among DSB proteins. The DSB system (essentially conformed by *dsbA* and *dsbB* genes) is involved in imparting structural stability to proteins by catalyzing the oxidation of cysteine residues to form DSBs and is particularly important for the correct folding of secreted proteins. The high sequence variability found among bacterial DSBs indicates that they probably have different substrate specificities (Heras et al. 2009); hence, the presence of divergent sets of DSB systems may be linked to the observed differences in *Campylobacter* secretomes. DSB orthologs present in the 23 *Campylobacter* genomes and related genera (*Sulfurospirillum* and *Arcobacter*) were recovered using BLAST searches against annotated DSB genes and analyzed using phylogenies. The main component of DSB system (*dsbA*) was found in all genomes and copy number varied from 1 to 3 (supplementary table S1, Supplementary Material online). The phylogenetic reconstruction using *dsbA* orthologs showed the presence of different groups that correlated with niche preferences and evidenced a great deal of gene duplication. Figure 7 shows that groups 1 and 2 share a recent common ancestor and are formed by the same oral *Campylobacter* species,

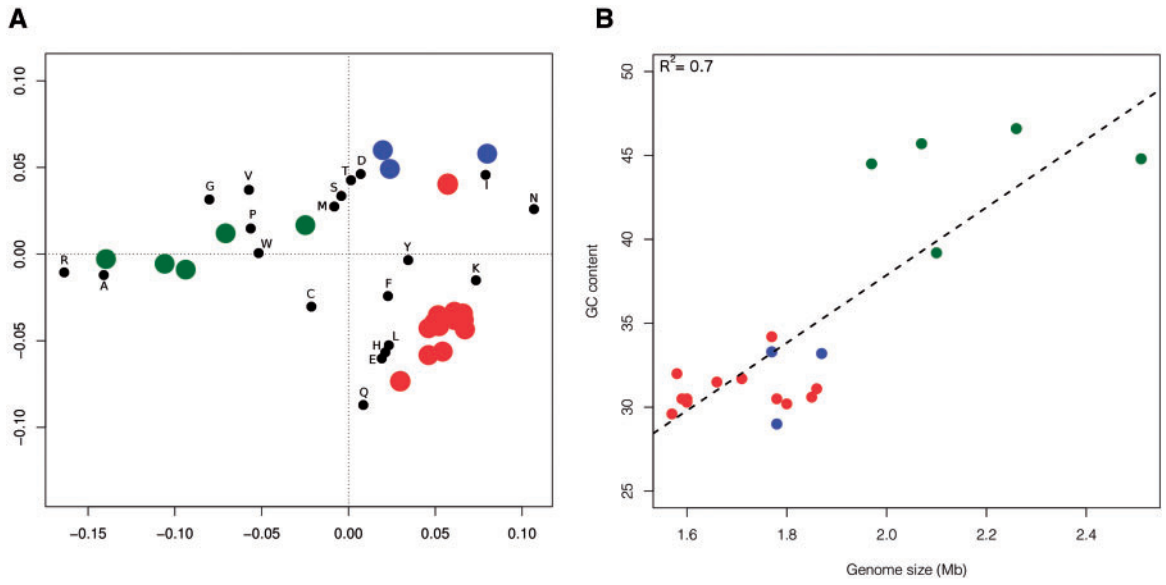

Fig. 6.—Correspondence analysis and whole-genome compositional features. Correspondence analysis using amino acids usage form secreted proteins (*A*) and linear correlation for genome size versus GC content (*B*). Small black circles represent each amino acid using the one-letter code. Big circles represent each genome colored according to niche preferences: gastrointestinal (red), genital (blue), and oral (green).

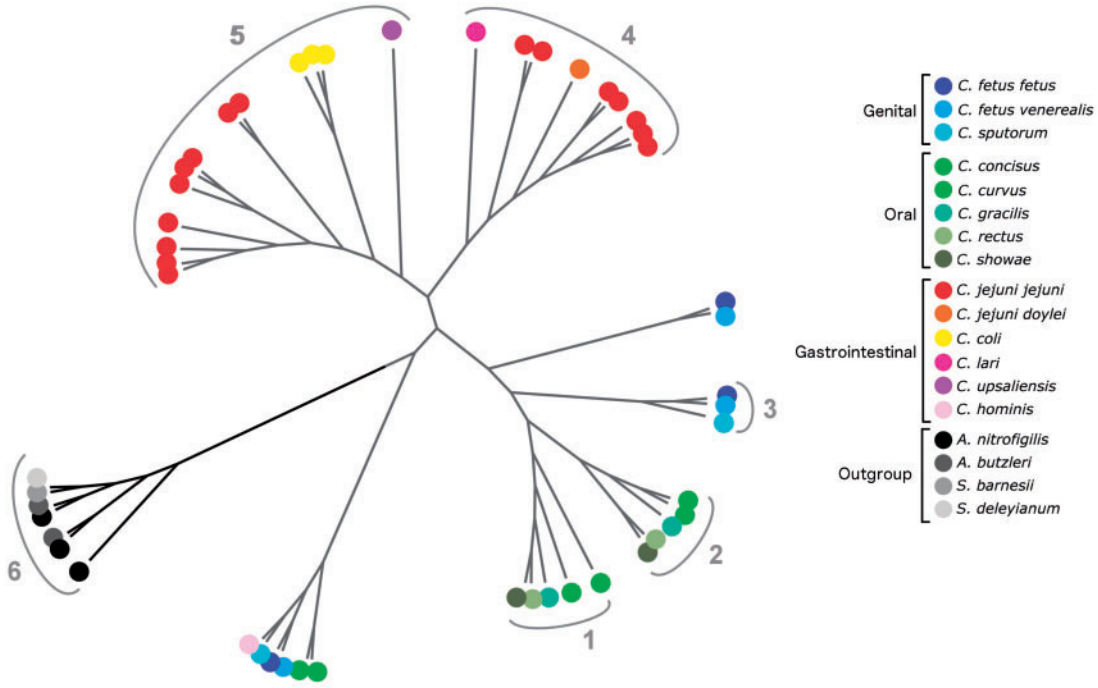

Fig. 7.—Phylogeny of DsbA. The phylogenetic tree using DsbA protein clusters *Campylobacter* and related genomes according to their niche preferences.

indicating recent paralogy for this divergent set of DSBs probably associated to niche preference for the oral cavity. Group 3 is just composed by the unique organisms capable of colonizing genital tissues (*C. fetus* subspecies and *C. sputorum*), reinforcing the hypothesis of niche-driven evolution of DSB proteins. This theory is additionally supported by the configuration of groups 4 and 5, exclusively formed by established pathogens (which are gastrointestinal).The ancestral genera *Sulfurospirillum* and *Arcobacter* clustered together (group 6), denoting an ancestral vertical evolution of DSB systems

among these taxa. In addition, these species showed the lowest gene diversification level, suggesting that DSB systems have experienced a duplication boost since *Campylobacter* diverged, with posterior specialization. Finally, this scenario is similar for DsbB protein, denoting the coevolution of this pair of functionally related genes (data not shown). We also found orthologs for *dsbD* and *dsbE* in the ancestral genera and in some *Campylobacter* species, suggesting gene loss events during the evolution of this genus, probably due to the nonessential functions of these genes.

Host Cells Invasion and Adhesion

The ability of *Campylobacter* species to invade host cells is a well-recognized virulence mechanism in all the established pathogens (Konkel and Joens 1989; Graham 2002; Mooney et al. 2003) and in some putative pathogens like *C. concisus* (Man et al. 2010). This phenotype is strongly correlated with the presence of the invasion antigen B (encoded by *ciaB* gene) which is the main genetic determinant for invasiveness in *Campylobacter*. In this study, we found single copy orthologs for this gene in all the species (average amino acid identity of 70%), even in *C. showae* and *C. hominis,* whose invasive capacity is apparently null (Man et al. 2010). These results open two alternative hypotheses: 1) just the presence of *ciaB* is not enough to warrant a successful invasion, considering that both *C. showae* and *C. hominis* presented a reduced repertory of virulence genes; or 2) sequence variation at amino acid level was responsible for function switching and/or specialization of this gene. In this sense, signals for diversifying selection were found on 34 over 607 (~5%) codon positions in the *ciaB* alignment (supplementary fig. S3, Supplementary Material online). Furthermore, we looked for shared positions within species belonging to the same niche in order to explain diversifying evolution as function of niche pressures. For *C. jejuni*, *C. coli*, *C. lari,* and *C. upsaliensis* (gastrointestinal established pathogens) we found 24 conserved positions that carried any different amino acid in the rest of the species, suggesting a strong diversifying pressure acting over *ciaB* gene for these phylogenetically related organisms. The scenario for genital species is slightly different because *C. fetus* subspecies are closely related but *C. sputorum* is phylogenetically distant, even though they shared 11 conserved positions that were different from the rest (three of them also showed positive selection signal), being the maximum number found for any possible species trio and significantly departing from the null distribution (see Materials and Methods for details). These results suggest that coevolution has been acting over these sites and reflect the probable specialization of CiaB to invade genital tissues. No associations were found for oral species.

Surface attachment to host cells is the previous step required for invasion. We previously described that genes coding for adhesins were overrepresented in established pathogens, evidencing that some adhesins are exclusive for these organisms. Although, we also found one gene coding for a fibronectin/fibrinogen-binding protein with ubiquitous distribution among campylobacters, suggesting that all species pose a basal attachment potential. The analysis of this gene showed the presence of 25 over 444 (~6%) sites under diversifying selection and 21 sites conserved among gastrointestinal pathogens whereas different in the rest. For this gene, no significant differences were found for genital or oral species. The role of diversifying selection has been previously highlighted for some *Campylobacter* genomes (Lefébure and Stanhope 2009), here we show how this evolutionary force is acting in some relevant genes and is probably driven by the particular environmental conditions found in different niches.

The Evolutionary Mechanism of *Campylobacter* Pathogenicity

The whole set of results obtained in this work allow us to accommodate an integrative hypothesis about *Campylobacter* evolution in terms of pathogenicity and niche preferences. Figure 8 shows a summary of the main forces shaping the evolutionary landscape of this genus. On one hand, horizontal gene transfers were probably the main evolutionary force involved in the emergence of some *Campylobacter* species as established pathogens. Probably the group of species conformed by *C. jejuni* (both subspecies) *C. coli*, *C. lari,* and *C. upsaliensis* gradually acquired a set of virulence genes from other bacteria and then transferred most of them to *C. fetus* subspecies. Based on this, pathogenic potential (established or putative pathogen) can be correlated with the presence/absence of certain genes with previous association to *Campylobacter* virulence (like CDT, capsule, or flagellum). On the other hand, gene diversification seemed to be playing a central role in the adaptation of species to different environmental conditions. This was suggested from the diversifying evolution of DSB orthologs, CiaB, and the whole secretome. However, the role of horizontal gene transfer events in niche preferences should not be discarded, because we found evidences for the acquisition of genes coding for secreted proteins in oral campylobacters and foreign genes in *C. sputorum* probably acquired from other non-*Campylobacter* genital species.

Despite the evolution of pathogenic potential and niche preferences should be somehow related, the presence of both established and putative pathogens colonizing the genital and gastrointestinal tracts indicated that they are not completely linked. This opens new questions regarding the relationship between niche and virulence, suggesting that genetic features that determine these phenotypes have different patterns of evolution. The following example involving genital species clearly describes this situation: The genome of *C. sputorum* did not present genes coding for SAP, which are the main antigenic determinants in *C. fetus* and have been associated to virulence in the context of genital

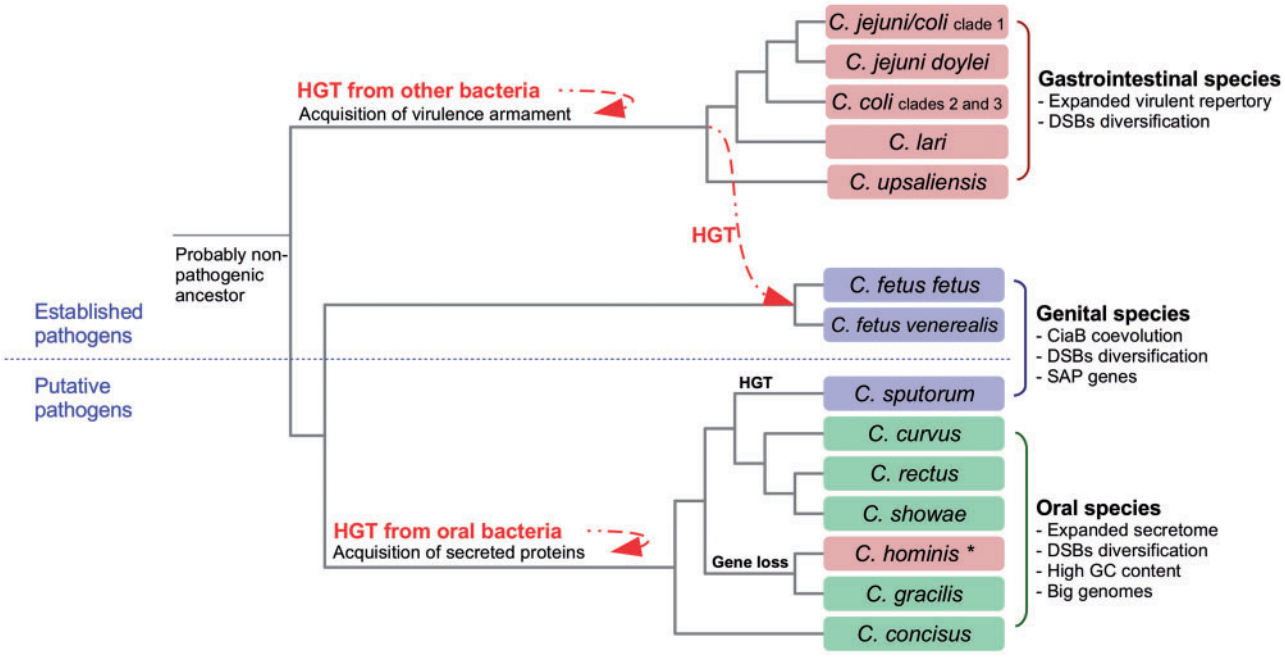

Fig. 8.—Main evolutionary processes in *Campylobacter*. This figure provides a phylogeny-based integrative view of the main evolutionary processes that have been shaping *Campylobacter* genomes in terms of pathogenicity and niche preferences. Species are highlighted in blue (genital), red (gastrointestinal), and green (oral). The species *C. hominis* is ticked off for not sharing the same genomic features than oral species, despite of belonging to the same phylogenetic group.

infections (Blaser and Pei 1993; Grogono-Thomas et al. 2000). On one hand, *C. sputorum* is capable of colonizing genital tissues without causing disease and did not code for SAP, suggesting that these proteins would not be essential for genital tropism. On the other hand, *C. fetus* is capable of causing infection in genital tissues and codes for SAP, so we propose that these proteins should have a role in virulence once the bacterium has been established in the tissue, more than in determining niche preference. How SAP genes emerged in *C. fetus* and why they were not transferred to other species, as well as other virulence genes, is an open question whose answer will involve a detailed study of horizontal gene transfer mechanisms in the context of *Campylobacter* infections.

Finally, the possibility of developing integrative comparative genomics analyses oriented to associate particular genomic features and evolutionary processes to phenotypes, not only depends on the availability of significant species for human health, but also in obtaining genomic information from neglected or less glamorous organisms. Additionally, the best scenario for performing these kinds of analyses should include many representative genomes form each species. The results presented here are product of comparing single representatives for certain campylobacters, evidently not considering the possible intraspecific variability of these species. However, this limitation could be improved in the near future as new genomes become available for different strains of the same species. This could provide the genetic information needed for refining our results and for gathering further genomic evidences for the evolution of pathogenicity and niche preferences among *Campylobacter* species.

Supplementary Material

Supplementary table S1 and figures S1–S3 are available at *Genome Biology and Evolution* online (http://gbe.oxfordjournals.org/).

Acknowledgments

The authors thank Trevor Lawley and David Harris from The Wellcome Trust Sanger Institute for invaluable assistance during genome sequencing process. The authors also thank Nicolás Sarute and Rodrigo Martino for reading the manuscript and providing constructive comments.

Literature Cited

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

Asakura M, et al. 2007. Comparative analysis of cytolethal distending toxin (cdt) genes among *Campylobacter jejuni*, *C. coli* and *C. fetus* strains. Microb Pathog. 42:174–183.

Aziz RK, et al. 2008. The RAST server: rapid annotations using subsystems technology. BMC Genomics 9:75.

Bahir I, Fromer M, Prat Y, Linial M. 2009. Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences. Mol Syst Biol. 5:1–14.

Bendtsen JD, Kiemer L, Fausboll A, Brunak S. 2005. Non-classical protein secretion in bacteria. BMC Microbiol. 5:58.
Blaser MJ, Pei Z. 1993. Pathogenesis of *Campylobacter fetus* infections: critical role of high-molecular-weight S-layer proteins in virulence. J Infect Dis. 167:372–377.
Butler J, et al. 2008. ALLPATHS-LG: de novo assembly of whole-genome shotgun microreads. Genome Res. 18:810–820.
Charif D, Lobry JR. 2007. Structural approaches to sequence evolution. Berlin/Heidelberg: Springer.
Chen L, et al. 2005. VFDB: a reference database for bacterial virulence factors. Nucleic Acids Res. 33: d325–d328.
Clark SC, Egan R, Frazier PI, Wang Z. 2013. ALE: a generic assembly likelihood evaluation framework for assessing the accuracy of genome and metagenome assemblies. Bioinformatics 29:435.
Conesa A, et al. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21: 3674–3676.
Collins MD, Rodriguez Jovita M, Lawson PA, Falsen E, Foster G. 1999. Characterization of a novel Gram-positive, catalase-negative coccus from horses: description of *Eremococcus coleocola* gen. nov., sp. nov. Int J Syst Bacteriol. 49:1381–1385.
Darveau RP. 2010. Periodontitis: a polymicrobial disruption of host homeostasis. Nat Rev Microbiol. 8:481–490.
Desvaux M, Hébraud M, Talon R, Henderson IR. 2009. Secretion and subcellular localizations of bacterial proteins: a semantic awareness issue. Trends Microbiol. 17:139–145.
Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.
Graham LL. 2002. *Campylobacter fetus* adheres to and enters INT 407 cells. Can J Microbiol. 48:995–1007.
Grogono-Thomas R, Dworkin J, Blaser MJ, Newewll DG. 2000. Roles of the surface layer proteins of *Campylobacter fetus* subsp. *fetus* in ovine abortion. Infect Immun. 68:1687–1691.
Gurtler M, Alter T, Kasimir S, Fehlhaber K. 2005. The importance of *Campylobacter coli* in human campylobacteriosis: prevalence and genetic characterization. Epidemiol Infect. 133:1081–1087.
Hagan EC, Mobley HL. 2007. Uropathogenic *Escherichia coli* outer membrane antigens expressed during urinary tract infection. Infect Immun. 75:3941–3949.
Hayashi K, et al. 2012. Molecular analysis of the 16S-23S rDNA internal spacer region (ISR) and truncated tRNA(Ala) gene segments in *Campylobacter lari*. World J Microbiol Biotechnol. 28: 2403–2410.
Heras B, et al. 2009. DSB proteins and bacterial pathogenicity. Nat Rev Microbiol. 7:215–225.
Jain D, Prasad KN, Sinha S, Husain N. 2008. Differences in virulence attributes between cytolethal distending toxin positive and negative *Campylobacter jejuni* strains. J Med Microbiol. 57:267–272.
Jun SR, Sims GE, Wu GA, Kim SH. 2010. Whole-proteome phylogeny of prokaryotes by feature frequency profiles: an alignment-free method with optimal feature resolution. Proc Natl Acad Sci U S A. 107: 133–138.
Kara D, Luppens SB, ten Cate JM. 2006. Differences between single- and dual-species biofilms of *Streptococcus mutans* and *Veillonella parvula* in growth, acidogenicity and susceptibility to chlorhexidine. Eur J Oral Sci. 114:58–63.
Kienesberger S, et al. 2014. Comparative genome analysis of *Campylobacter fetus* subspecies revealed horizontally acquired genetic elements important for virulence and niche specificity. PLoS One 9: e85491.
Konkel ME, Joens L. 1989. Adhesion to and invasion of Hep-2 cells by *Campylobacter* spp. Infect Immun. 57:2984–2990.
Kumamoto CA. 2008. Niche-specific gene expression during *C. albicans* infection. Curr Opin Microbiol. 11:325–330.
McDougald D, et al. 2002. Defences against oxidative stress during starvation in bacteria. Antonie van Leeuwenhoek 81:3–13.
Man SM. 2011. The clinical importance of emerging *Campylobacter* species. Nat Rev Gastroenterol Hepatol. 8:669–685.
Man SM, et al. 2010. Host attachment, invasion, and stimulation of proinflamatory cytokines by *Campylobacter concisus* and other non-*Campylobacter jejuni Campylobacter* species. J Infect Dis. 202. 1855:1865.
McLean JS, et al. 2012. Identifying low pH active and lactate-utilizing taxa within oral microbiome communities from healthy children using stable isotope probing techniques. PLoS One 7:e32219.
Medini D, Donati C, Tettelin H, Masignani V, Rappuoli R. 2005. The microbial pan-genome. Curr Opin Genet Dev. 15:589–594.
Mooney A, et al. 2003. Invasion of human epithelial cells by *Campylobacter upsaliensis*. Cell Microbiol. 5:835–847.
Moolhuijzen PM, et al. 2009. Genomic analysis of *Campylobacter fetus* subspecies: identification of candidate virulence determinants and diagnostic assays targets. BMC Microbiol. 9:86.
Moran NA. 2002. Microbial minimalism: genome reduction in bacterial pathogens. Cell 108:583–586.
Moran NA, Plague GR. 2004. Genomic changes following host restriction in bacteria. Curr Opin Genet Dev. 14:627–633.
Mshelia GD, Amin JD, Woldehiwet Z, Murray RD, Egwu GO. 2010. Epidemiology of bovine venereal Campylobacteriosis: geographic distribution and recent advances in molecular diagnostic techniques. Reprod Domest Anim. 45:e221–e230.
Murrell B, et al. 2012. Detecting individual sites subject to episodic diversifying selection. PLoS Genetics. 7:e1002764.
Nachamkin I, Szymanski CM, Blaser MJ. 2008. *Campylobacter*. Washington (DC): ASM Press.
Nenadic O, Greenacre M. 2007. Correspondence analysis in R, with two- and three-dimensional graphics: The ca package. J Stat Softw. 20: 1–13.
Noinaj N, Guillier M, Barnard TJ, Buchanan SK. 2010. TonB-dependet transporters: regulation, structure and function. Annu Rev Microbiol. 64:43–60.
Labarca JA, et al. 2002. *Campylobacter upsaliensis*: another pathogen for consideration in the United States. Clin Infect Dis. 34:59–60.
Lawson AJ, Linton D, Stanley J. 1998. 16S rRNA gene sequences of '*Candidatus Campylobacter hominis*', a novel uncultivated species, are found in the gastrointestinal tract of healthy humans. Microbiology 144:2063–2071.
Lefébure TL, Stanhope MJ. 2009. Pervasive, genome-wide positive selection leading to functional divergence in the bacterial genus. Campylobacter, 19:1224-1232.
Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13:2178–2189.
Linton D, Owen RJ, Stanley J. 1996. Rapid identification by PCR of the genus *Campylobacter* and of five *Campylobacter* species enteropathogenic for man and animals. Res Microbiol. 147: 707–718.
On SL, Atabay HI, Corry JE, Harrington CS, Vandamme P. 1998. Emended description of *Campylobacter sputorum* and revision of its infrasubspecific (biovar) divisions, including *C. sputorum* biovar paraureolyticus, a urease-producing variant from cattle and humans. Int J Syst Bacteriol. 48:195–206.
On SL, Ridgwell F, Cryan B, Azadian BS. 1992. Isolation of *Campylobacter sputorum* biovar sputorum from an axillary abscess. J Infect. 24: 175–179.
Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–290.
Parker CT, Miller GM, Horn ST, Lastovica AJ. 2007. Common genomic features of *Campylobacter jejuni* subsp. *doylei* distinguish them from *C. jejuni* subsp. jejuni. BMC Microbiol. 7:50.

Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 8:785–786.
Pruesse E, et al. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. 35:7188–7196.
Purdy D, et al. 2000. Characterization of cytolethal distending toxin (CDT) mutants of *Campylobacter jejuni*. J Med Microbiol. 49:473–479.
Ripabelli G, Tamburro M, Minelli F, Leone A, Sammarco ML. 2010. Prevalence of virulence-associated genes and cytolethal distending toxin production in *Campylobacter* spp. isolated in Italy. Comp Immunol Microbiol Infect Dis. 33:355–364.
Quetz JS, et al. 2012. *Campylobacter jejuni* infection and virulence-associated genes in children with moderate to severe diarrhoea admitted to emergency rooms in northeastern Brazil. J Med Microbiol. 61: 507–513.
Rocha EP, Danchin A. 2002. Base composition bias might result from competition for metabolic resources. Trends Genet. 18:291–294.
Schalk IJ, Hannauer M, Braud A. 2011. New roles for bacterial siderophores in metal transport and tolerance. Environ Microbiol. 13: 2844–2854.
Sebald M, Véron M. 1963. Teneur en bases de l'ADN et classification des vibrions. Ann Inst Pasteur. 105:897–910.
Sharp PM, Bailes E, Grocock RJ, Peden JF, Sockett RE. 2005. Variation in the strength of selected codon usage bias among bacteria. Nucl Acids Res. 33:1141–1153.
Sheppard SK, Didelot X, Jolley KA, et al. 2013. Progressive genome-wide introgression in agricultural *Campylobacter coli*. Mol Ecol. 22: 1051–1064.
Sheppard SK, Didelot X, Meric G, et al. 2013. Genome-wide association study identifies vitamin B_5 biosynthesis as a host specificity factor in *Campylobacter*. Proc Natl Acad Sci U S A. 29:11923–11927.
Siqueira JF, Rocas IN. 2002. *Campylobacter gracilis* and *Campylobacter rectus* in primary endodontic infections. Int Endod J. 36:174–190.
Skarp-de Haan CPA, et al. 2014. Comparative genomics of unintrogressed *Campylobacter coli* clades 2 and 3. BMC Genomics 15:129.
Skirrow MB. 1994. Diseases due to *Campylobacter*, *Helicobacter* and related bacteria. J Comp Pathol. 111:113–149.
Song C, Kumar A, Saleh M. 2009. Bioinformatic comparison of bacterial secretomes. Genomics Proteomics Bioinformatics 7:37–46.
Swain MT, et al. 2012. A post-assembly genome-improvement toolkit (PAGIT) to obtain annotated genomes from contigs. Nat Protoc. 7: 1260–1284.
Talukder KA, et al. 2008. Prevalence of virulence genes and cytolethal distending toxin production in *Campylobacter jejuni* isolates from diarrheal patients in Bangladesh. J Clin Microbiol. 46: 1485–1488.
Tee W, Luppino M, Rambaldo S. 1998. Bacteremia due to *Campylobacter sputorum* Biovar sputorum. Clin Infect Dis. 27:1544–1545.
van Bergen MAP, Linnane S, van Putten JP, Wagenaar JA. 2005. Global detection and identification of *Campylobacter fetus* subsp. *venerealis*. Rev Sci Tech. 24. 1017:1026.
Vianna ME, Holtgraewe S, Seyfarth I, Conrads G, Horz HP. 2008. Quantitative analysis of three hydrogenotrophic microbial groups, methanogenic archaea, sulfate-reducing bacteria, and acetogenic bacteria, within plaque biofilms associated with human periodontal disease. J Bacteriol. 190:3779–3789.
Wagenaar JA, et al. 2014. *Campylobacter fetus* infections in humans: exposure and disease. Clin Infect Dis. 58:1579–1586.
Wassenaar TM, Gaastra W. 2001. Bacterial virulence: can we draw the line? FEMS Microbiol Lett. 201:1–7.
Yeoman CJ, et al. 2010. Comparative genomics of *Gardnerella vaginalis* strains revals substantial differences in metabolic and virulence potential. PLoS ONE., doi:/10.1371/journal.pone.0012411.
Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18:821–829.

Associate editor: Tal Dagan