



OPEN

# Comparative pangenomic analysis of *Campylobacter fetus* isolated from Spanish bulls and other mammalian species

Nerea Pena-Fernández<sup>1,2,6</sup>, Medelin Ocejo<sup>1,2,6</sup>, Linda van der Graaf-van Bloois<sup>3</sup>, Jose Luís Lavín<sup>4</sup>, Nekane Kortabarria<sup>2</sup>, Esther Collantes-Fernández<sup>5</sup>, Ana Hurtado<sup>1,2</sup> & Gorka Aduriz<sup>2</sup>✉

*Campylobacter fetus* comprises two closely related mammal-associated subspecies: *Campylobacter fetus* subsp. *fetus* (*Cff*) and *Campylobacter fetus* subsp. *venerealis* (*Cfv*). The latter causes bovine genital campylobacteriosis, a sexually-transmitted disease endemic in Spain that results in significant economic losses in the cattle industry. Here, 33 *C. fetus* Spanish isolates were whole-genome sequenced and compared with 62 publicly available *C. fetus* genomes from other countries. Genome-based taxonomic identification revealed high concordance with in silico PCR, confirming Spanish isolates as *Cff* ( $n=4$ ), *Cfv* ( $n=9$ ) and *Cfv* biovar *intermedius* (*Cfvi*,  $n=20$ ). MLST analysis assigned the Spanish isolates to 6 STs, including three novel: ST-76 and ST-77 for *Cfv* and ST-78 for *Cff*. Core genome SNP phylogenetic analysis of the 95 genomes identified multiple clusters, revealing associations at subspecies and biovar level between genomes with the same ST and separating the *Cfvi* genomes from Spain and other countries. A genome-wide association study identified *pqqL* as a *Cfv*-specific gene and a potential candidate for more accurate identification methods. Functionality analysis revealed variations in the accessory genome of *C. fetus* subspecies and biovars that deserve further studies. These results provide valuable information about the regional variants of *C. fetus* present in Spain and the genetic diversity and predicted functionality of the different subspecies.

**Keywords** *Campylobacter fetus*, Whole-genome sequencing (WGS), Pangenome, Phylogeny, Genome-wide association study (GWAS), Gene function

*Campylobacter fetus* comprises three subspecies. Two of them, *Campylobacter fetus* subsp. *fetus* (*Cff*) and *Campylobacter fetus* subsp. *venerealis* (*Cfv*), are highly relevant veterinary pathogens commonly associated with mammals. The third one, *Campylobacter fetus* subsp. *testudium* (*Cft*), is mainly associated with reptiles and shows a clear genetic divergence from ruminant-associated *C. fetus* subspecies<sup>1–3</sup>. *Cff* can be found in the gastrointestinal tract of healthy ruminants, and by translocation across the intestinal mucosa, it can reach the bloodstream and colonise the gravid uterus causing abortions in sheep and sporadic abortions in cattle. *Cff*-caused sepsis and/or gastrointestinal disease have also been reported in immunocompromised humans<sup>4–6</sup>. *Cfv* is the causative agent of bovine genital campylobacteriosis (BGC), a bovine sexually transmitted disease distributed worldwide, and endemic in countries where natural breeding is used and cattle are extensively managed. BGC is included in the list of notifiable transmissible diseases of the World Organisation for Animal Health (WOAH, formerly known as Office International des Epizooties—OIE) and causes infertility, early embryonic death and abortions, resulting in significant economic losses for the beef cattle sector<sup>7,8</sup>.

<sup>1</sup>SERIDA, Servicio Regional de Investigación y Desarrollo Agroalimentario, Carretera de Oviedo, s/n, 33300 Villaviciosa, Spain. <sup>2</sup>Animal Health Department, NEIKER – Basque Institute for Agricultural Research and Development, Basque Research and Technology Alliance (BRTA), Bizkaia Science and Technology Park 812L, 48160 Derio, Spain. <sup>3</sup>Department of Biomolecular Health Sciences, Faculty of Veterinary Medicine, Utrecht University, Utrecht, The Netherlands. <sup>4</sup>Department of Applied Mathematics, NEIKER - Basque Institute for Agricultural Research and Development, Basque Research and Technology Alliance (BRTA), Bizkaia Science and Technology Park 812L, 48160 Derio, Spain. <sup>5</sup>Animal Health Department, Faculty of Veterinary Sciences, SALUVET, Complutense University of Madrid, Ciudad Universitaria s/n, 28040 Madrid, Spain. <sup>6</sup>These authors contributed equally: Nerea Pena-Fernández and Medelin Ocejo. ✉email: ahurtado@neiker.eus; gaduriz@neiker.eus

*Cff* and *Cfv* are closely related phenotypically and genotypically thus hampering their identification in the laboratory. The diagnostic techniques recommended by the WOAHA are culture followed by biochemical tests, the most relevant ones being the tolerance test to 1% glycine and the hydrogen sulfide (H<sub>2</sub>S) production test, both with positive result for *Cff* and negative for *Cfv*<sup>8</sup>. However, the *Cfv* biovar *intermedius* (*Cfvi*) shows intermediate results between *Cff* and *Cfv* (intolerance to 1% glycine as *Cfv* and ability to produce H<sub>2</sub>S like *Cff*), which can lead to misidentifications<sup>9,10</sup>. Nonetheless, recent research<sup>11</sup> has reported poor reproducibility of phenotypic tests and discrepancies with genotypic tests, which challenges the use of current biochemical methods for subspecies identification. Although several PCR methods have been described for the diagnosis of *C. fetus* subspecies, the most accurate are those targeting *C. fetus*-specific *nahE* and the *Cfv*-specific *ISCfe1* insertion element<sup>12,13</sup>. Differentiation between *Cfv* and *Cfvi* is also possible by PCR targeting the L-cysteine transporter (L-Cyst) related to H<sub>2</sub>S production<sup>14</sup>. Other molecular methods such as matrix-assisted laser desorption ionisation-time-of-flight (MALDI-TOF), pulsed-field gel electrophoresis (PFGE) or amplified fragment length polymorphism (AFLP) have also been described, but none of these molecular tests can reliably identify *C. fetus* subspecies<sup>15–17</sup>. Analyses based on multilocus sequence typing (MLST) showed that mammal-associated *C. fetus* subspecies were generally separated into two large genetically homogeneous clusters (cluster A, that includes bovine derived *Cfv/Cfvi* strains belonging to sequence type (ST) 4, and cluster B, that encompasses strains of different STs, mostly human derived *Cff*) with an apparently clonal population structure<sup>18,19</sup>. However, recent studies showed that correlating MLST STs with subspecies and biovar identity was not possible due to the presence of human *Cff* strains belonging to the bovine-associated ST-4 cluster<sup>11,20,21</sup>. Whole genome-based studies revealed highly syntenic genomes in *Cff* and *Cfv* with 92.9% sequence identity<sup>11,22</sup> alongside the presence of some gene regions exclusive in each subspecies resulting from evolutionary processes of adaptations to different hosts<sup>22</sup>.

A recent study revealed that BGC is endemic in Spain and widely distributed in Spanish cattle herds (especially in dehesa system herds) with a prevalence of 12.2% and 7.7% in herds and bulls, respectively<sup>23</sup>. In addition, this study also identified some specific mountain areas in Spain with a higher risk of BGC, where the use of communal pastures and bulls are frequent<sup>23</sup>. Considering these observations, it is important to investigate which subspecies of *C. fetus* are present in Spanish cattle and describe their genetic characteristics. For this purpose, complete genomes of *C. fetus* field isolates from Spanish cattle were sequenced and analysed with the following objectives: i) to taxonomically classify the isolates at subspecies level and, in the case of *Cfv*, at biovar level, ii) to study their genetic diversity by MLST typing, and iii) to perform comparative genomics of *C. fetus* field isolates with existing mammalian *C. fetus* genomes from different countries available in GenBank.

## Results

### Taxonomic identification and assignation of MLST types (ST)

The taxonomic analysis with Kraken identified four of the Spanish isolates as *Cff* and 29 as *Cfv*; this assignation was in complete agreement (100%) with the in silico PCR1. Differentiation of *Cfv* isolates at the biovar level based on L-Cyst PCR analysis (in silico PCR2) showed that 20 of the 29 *Cfv* isolates were *Cfvi* and 9 were *Cfv* (Table S1). Similarly, the 62 *C. fetus* genomes extracted from GenBank were taxonomically classified as follows: 27 *Cff*, 13 *Cfv*, and 22 *Cfvi*.

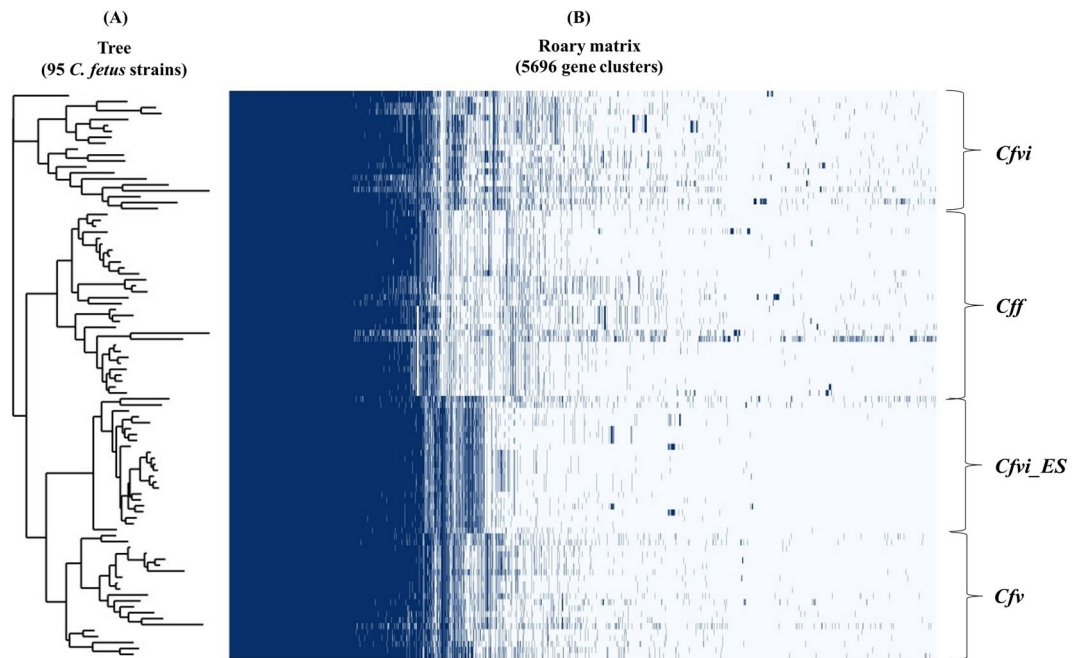
The average size of the *C. fetus* genomes analysed was 1.87 Mb (ranging from 1.73 to 2.19 Mb), with a mean G + C content of 33.1% (ranging from 32.9 to 33.4%). The mean number of genes per genome was 1,904, with *Cfvi* isolates having the highest average number of genes (1,962), followed by *Cfv* (1,946) and *Cff* (1,795) (Table S1).

In silico MLST analysis assigned the 33 Spanish genomes to 6 STs, including three novel types (Table S1). Thus, 20 *Cfvi* and 1 *Cfv* were assigned to ST-4, 2 *Cff* to ST-6 and 1 *Cff* to ST-3. The remaining 9 genomes showed novel STs: 5 *Cfv* were assigned to ST-76, 3 *Cfv* to ST-77, and 1 *Cff* to ST-78. ST-76 (strains *Cfv002*, *Cfv007*, *Cfv008*, *Cfv009* and *Cfv022*) was the result of a novel *aspA* allele resulting from two nucleotide transitions at positions 421 and 423 (G-A and A-G, respectively); ST-77 (strains *Cfv012*, *Cfv015* and *Cfv020*) carried a G-A nucleotide change at position 421 of the *aspA* gene; and, ST-78 (strain *Cff004*) had a nucleotide substitution of a thymine for a cytosine at position 126 of the *glnA* gene (Table S2). Most of the *Cfv* and *Cfvi* genomes downloaded from GenBank (31/35) belonged to ST-4, while the *Cff* genomes belonged to 8 different STs.

### Pangenome analysis of mammal-associated *C. fetus* strains

The analysis of the 95 ruminant-associated *C. fetus* genomes showed a total of 5696 genes including 1365 core genes (999 core genes and 366 soft-core genes) and 4331 accessory genes (948 shell genes and 3383 cloud genes). The genes presence/absence matrix revealed distinct patterns among the different taxonomic groups (Fig. 1B), which were supported by the PCA analysis (Figure S1) that showed a clear clustering of the genomes by subspecies (PERMANOVA,  $p_{adj} = 0.001$ ). Differences were also observed in the genetic profile of the Spanish *Cfvi* genomes compared to the rest of the genomes, while the *Cfvi* genomes from other countries showed a considerable heterogeneity, showing a more dispersed clustering (Figs. 1 and S1).

The phylogenetic analysis based on the core genome SNPs (1365 genes) showed multiple clusters (Fig. 2). The genomes taxonomically classified as *Cff*, originating from different host species (cattle, sheep, and human) and countries, displayed notable diversity as evidenced by the wide range of different STs, exceeding the diversity observed in *Cfv* and *Cfvi*. Nevertheless, despite this diversity, the *Cff* genomes consistently clustered together due to their shared STs. The *Cfv* (including *Cfvi*) genomes were all isolated from cattle and showed more similar core genomes. However, two subclusters were observed. The first one included all the Spanish *Cfvi* genomes (all of ST-4), and three ST-4 *Cfvi* genomes from Australia and South Africa. The remaining *Cfv* and *Cfvi* genomes formed a separate subcluster in which the genomes were separated at the biovar level into clearly different branches. The non-Spanish *Cfvi* genomes and *Cfvi9825* (taxonomically classified as *Cfv*) clustered together regardless of their ST (Fig. 2). Although genomes taxonomically classified as *Cfv* clustered together,



**Figure 1.** Pangenome analysis of 95 mammal-associated *C. fetus* genomes. **(A)** Dendrogram representing the grouping of the 95 *C. fetus* genomes according to the distribution of their accessory genes. **(B)** Roary matrix representing the complete genetic profile of each of the genomes based on presence/absence of core and accessory genes. *Cfvi*, *C. fetus* subsp. *venerealis* biovar *intermedius* isolates from other countries; *Cfvi\_ES*, *C. fetus* subsp. *venerealis* biovar *intermedius* from Spain; *Cff*, *C. fetus* subsp. *fetus*; *Cfv*, *C. fetus* subsp. *venerealis*.

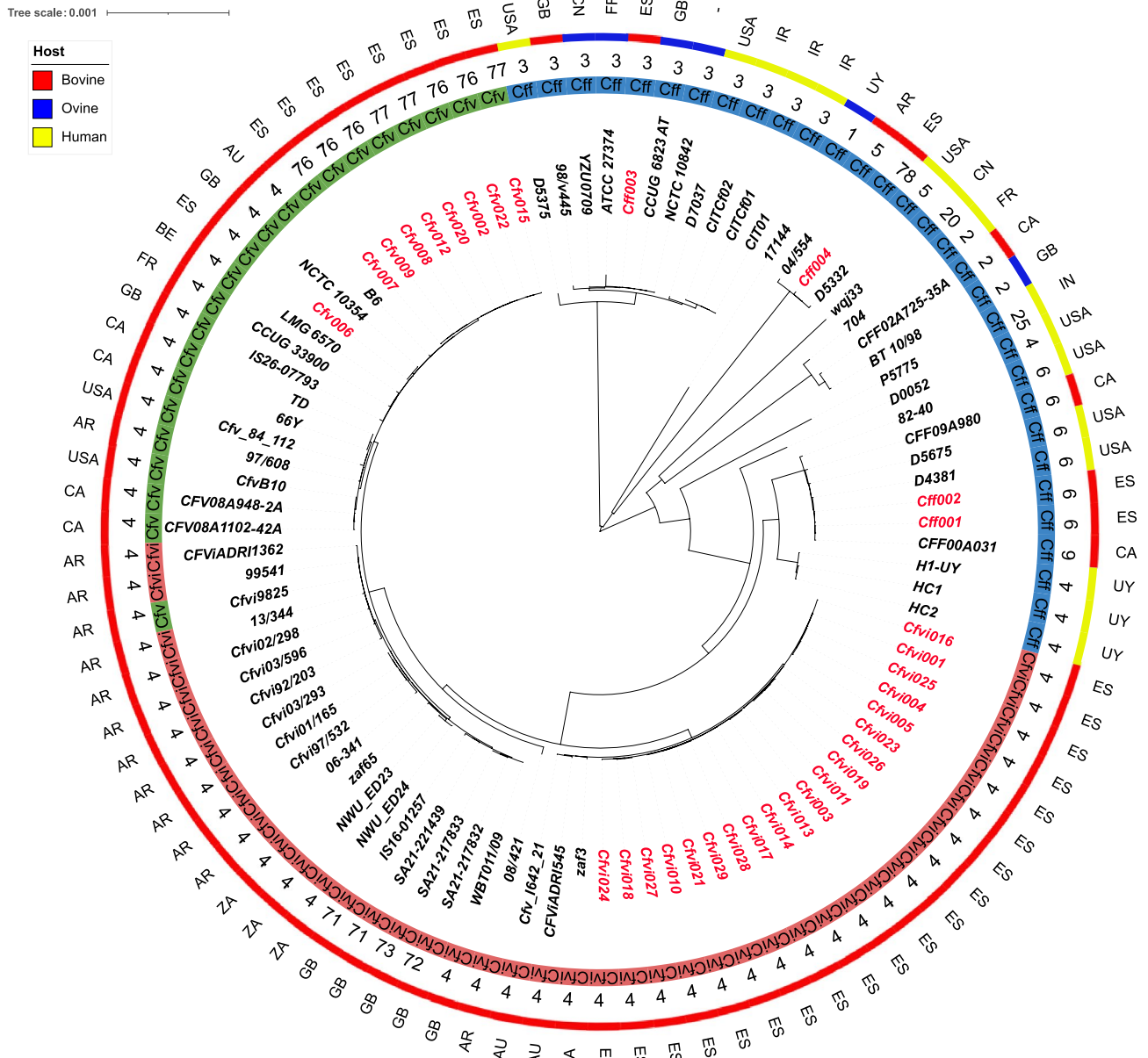
those originating from countries other than Spain (all assigned to ST-4) showed greater variability compared to the Spanish genomes, which were assigned to new STs (ST-76 and ST-77) and formed a separate cohesive subgroup. Phylogenetic analysis also revealed certain geographical association for *Cfv* and *Cfvi* genomes. Thus, *Cfvi* genomes from the same countries of origin clustered on the same branch, and for *Cfv* genomes, those from America were more similar to each other, while the European genomes were grouped separately. No clear geographical relations were observed for the *Cff* genomes.

Analysis of the accessory genome, like the core genome analysis, showed clear clustering of the genomes at the subspecies level. In this case, when presence/absence of accessory genes was compared, four different groups were identified (Fig. 1A), as the *Cff* genomes clustered on a single subcluster of the dendrogram. On the other hand, in contrast to the analysis of the core genomes, the *Cfvi* Spanish genomes shared a higher number of accessory genes with the *Cfv* genomes, while *Cfv* genomes from other countries clustered in a separate subcluster.

GWAS analysis identified a set of genes specific for each subspecies and the biovar *intermedius* (Tables S3–S5). A total of 180 genes strongly associated with *Cff* were identified, 66 of which were exclusively found in *Cff* genomes albeit not in all of them ( $n = 31$ ). Of the 233 genes strongly associated with *Cfvi*, 68 were found only in biovar *intermedius* genomes but were not present in all *Cfvi* genomes ( $n = 42$ ). On the other hand, 157 genes strongly associated with *Cfv* were detected, and the *pqqL* gene, which codes for a zinc-dependent peptidase of the M16 family, was present in all genomes of this subspecies ( $n = 22$ ). When these analyses were performed separating the Spanish genomes from the genomes of other countries, no specific genes associated to the Spanish genomes were found for any of the *C. fetus* subspecies or biovar.

### Functional characterization of genes comprising the different pangenome categories

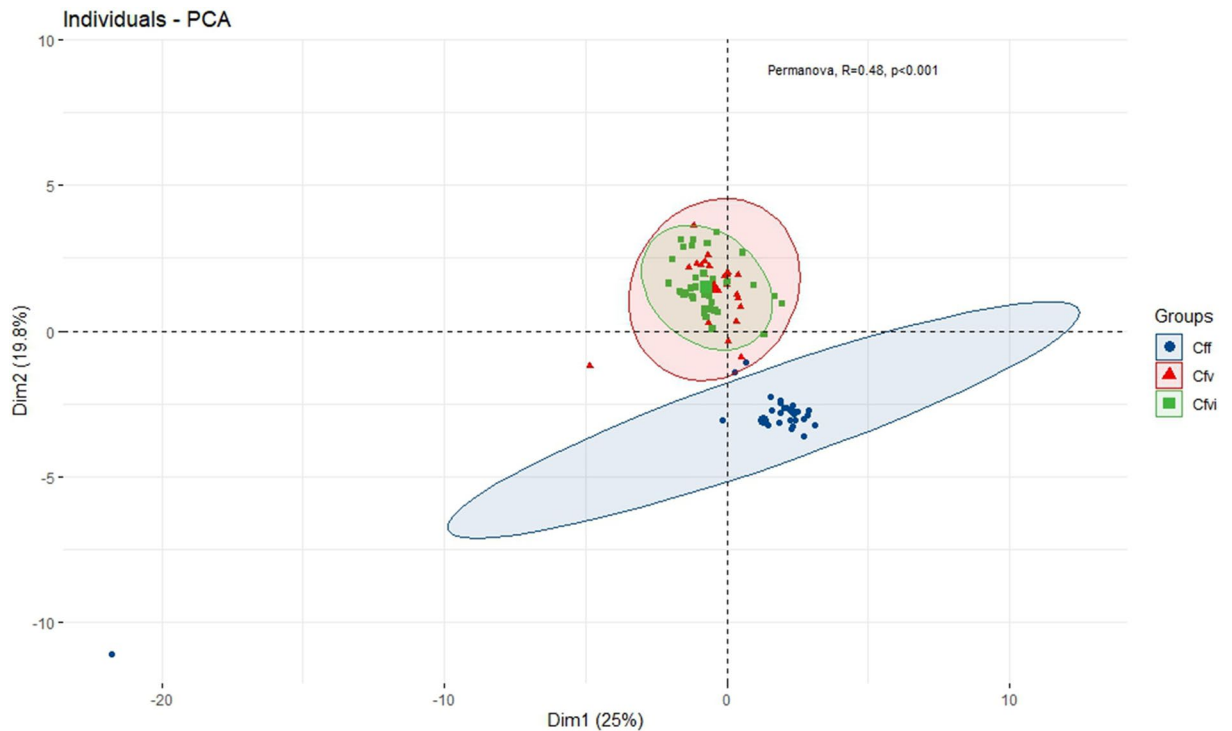
To determine the potential functionality associated with the genes belonging to the different pangenome categories (core and accessory), we performed a functional classification based on the KEGG database. Of the 5696 genes that comprised the pangenome, a total of 1746 genes could be annotated, 907 from the core genome (679 core, 228 soft-core) and 839 from the accessory genome (314 shell and 525 cloud genes). Most of the annotated core genes were related to genetic information processing (10.9%), signalling and cellular processes (10.4%) and carbohydrate metabolism (8.7%). The accessory genes were mostly related to signalling and cellular processes (16.8%), genetic information processing (11.7%) and membrane transport (11.7%). Functionality analysis also showed that most of the gene functions observed were mainly encoded by core genes and some of them, such as biodegradation and metabolism of xenobiotics, immune system, environmental adaptation, and aging, were encoded exclusively by core genes. On the other hand, processes linked to membrane transport (bacterial secretion system), replication and repair, cell communication such as quorum sensing (cell community in prokaryotes), infectious diseases and signalling and cellular processes related with cell defence, were mostly encoded by accessory genes (Figure S2).



**Figure 2.** Global phylogeny of *Campylobacter fetus* based on core genome SNP analysis of 95 *C. fetus* genomes. First (outer) ring represents country of origin of the strains. Second ring represents the isolation source of strains (red, bovine; yellow, humans; blue, ovine). Numbers in the third ring represent the MLST sequence type (ST). The fourth ring (coloured lines) represents the taxonomic assignment at the subspecies level (*Cfv*—*C. fetus* subsp. *venerealis*, green; *Cfvi*—*C. fetus* subsp. *venerealis* biovar *intermedius*, red; *Cff*—*C. fetus* subsp. *fetus*, blue). Isolates sequenced in this study are labelled in red.

PCA analysis based on gene functionality showed that there were significant differences between the *C. fetus* subspecies, including the *Cfvi* biovar (PERMANOVA  $p_{adj} = 0.001$ ) (Fig. 3). When the same analysis was performed considering the Spanish *Cfvi* genomes as a distinct group, significant differences in the functionality of the accessory genome were observed between the Spanish *Cfvi* genomes and the rest of the *Cfvi* genomes ( $p_{adj} = 0.001$ ) and the *Cff* genomes ( $p_{adj} = 0.001$ ). Conversely, no significant differences in functionality were observed between *Cfv* genomes and Spanish *Cfvi* genomes ( $p_{adj} = 0.079$ ). These results are consistent with those obtained in the study of the accessory genome composition where these two groups of genomes grouped in the same subcluster (Fig. 1A).

Overall, the proportion of genes encoding the different functions observed was significantly higher in the *Cff* genomes group ( $p_{adj} < 0.001$ ). In contrast, *Cfv* and biovar *intermedius* had a significantly higher proportion of genes related to membrane transport ( $p_{adj} < 0.004$ ), genetic information processing ( $p_{adj} < 0.05$ ), and communication, such as quorum sensing functions (cell community in prokaryotes) ( $p_{adj} < 0.05$ ) which represented a significantly higher proportion in the *Cfv* and biovar *intermedius* groups. The functions related to processes involved in bacterial infectious diseases were proportionally more represented in the Spanish *Cfvi* genomes ( $p_{adj} < 0.001$ ).



**Figure 3.** Principal component analysis based on the functional analysis of the accessory gene content of the genomes of *C. fetus* subspecies and biovar. *Cfv*—*C. fetus* subsp. *venerealis*; *Cfvi*—*C. fetus* subsp. *venerealis* biovar *intermedius*; *Cff*—*C. fetus* subsp. *fetus*.

and included three alleles of the cytolethal distending toxin B subunit gene (*cdtB*), a homologue of the *hp0169* gene of *Helicobacter pylori* which encodes a putative collagenase (group 1741, prokka classification), two alleles of the *phaC* gene of *Bordetella pertussis* (groups 289 and 292), and the *lepA* gene involved in the process of host cell invasion in *Legionella pneumophila*. Finally, functions related to transcription, xenobiotic degradation and metabolism, and the metabolism of other amino acids did not show significant differences in their distribution across subspecies (Figure S3).

## Discussion

Several studies have demonstrated that WGS is an efficient method to accurately characterize *C. fetus* subspecies and to investigate their genomic differences, valuable for the development of new and more accurate diagnostic tools<sup>5,24,25</sup>. Under this premise, in the current investigation, the genomes of 33 Spanish *C. fetus* isolates were sequenced and compared with other publicly available genomes of mammalian-associated *C. fetus* from other countries. Taxonomic identification by in silico PCR<sup>13</sup> and Kraken<sup>26</sup> showed perfect agreement for the identification of *Cff* and *Cfv*. These results were consistent with the PCA and phylogenetic analysis, which clustered the genomes at subspecies level. As previously reported<sup>27–29</sup> and confirmed in this study, the taxonomic results provided by Kraken are consistent with commonly used reference methods for the identification of bacterial genomes, making it a useful and precise classification tool.

Previous studies showed the close genetic relatedness and low genomic and ST variability among mammal-associated *C. fetus* subspecies and recommended the use of MLST for long-term epidemiological and phylogenetic analyses<sup>13,18</sup>. In this study, MLST separated the 95 *C. fetus* genomes into 14 different STs, most of them (9/14) associated to *Cff* isolated from different hosts. *Cfv*/*Cfvi* isolates (all from cattle) were mostly assigned to ST-4, a sequence type associated with the bovine lineage of *Cfv*/*Cfvi* isolates<sup>11,18</sup> and only sporadically detected in *Cff* isolates from humans<sup>19,20</sup>. Additionally, two new STs (ST-76 and ST-77) were described in Spanish *Cfv*. Until this study, STs other than ST-4 had only been described for *Cfvi*<sup>18,30,31</sup> and *Cfv* had been characterized as a genetically homogeneous species exclusively linked to ST-4. These results indicate a greater genetic diversity among *Cfv* isolates from Spain.

Core genome based phylogenetic analysis clustered the genomes at the subspecies level; the only exception was one genome (*Cfvi*9825) that grouped together with *Cfvi* genomes despite being taxonomically identified as *Cfv*. This genome has previously exhibited discordant biochemical classifications in various labs<sup>18,32</sup>. Consistent with another investigation<sup>11</sup>, a remarkable genetic similarity was found between strains sharing the same ST as they clustered together. Also, in agreement with the MLST analysis, phylogenetic analysis showed that the *Cff* genome group exhibited a higher genetic variability compared to the *Cfv* and *Cfvi* groups, which presented more homogeneous core genomes. This homogeneity may be attributed to the restrictive adaptation of *Cfv* and its biovar *intermedius* to a particular host<sup>18,21</sup>. Although no significant differences between *Cfv* and *Cfvi* genomes were found in other studies<sup>11,21</sup>, our phylogenomic analysis revealed a clear separation of *Cfv* genomes at the

biovar level. In addition, it was also observed that the *Cfvi* genomes from Spain showed a clonal structure and clustered separately from the rest of the *Cfv/Cfvi* genomes, suggesting the presence of a regional *Cfvi* variant circulating in Spain, similar to what has been reported in Argentina or Germany<sup>25,33</sup>. The situation was similar for the Spanish *Cfv* strains, which albeit being more closely related to *Cfv* from other countries, clustered separately. The fact that the Spanish *Cfvi* genomes shared a greater number of accessory genes with the *Cfv* genomes could indicate shared adaptive strategies or common functional traits.

Several PCR diagnostic methods are currently available to differentiate between mammal-associated *C. fetus* subspecies, but none of them are 100% accurate<sup>13,33</sup>. Therefore, the development of new and more reliable methods is crucial. GWAS has proven useful for the detection of new genetic targets associated with a trait of interest in bacterial populations<sup>34</sup>. In this study, a considerable number of genes strongly associated with each subspecies were identified, however in the case of *Cff* and *Cfvi* none of these genes were found to be present in all strains of each group. Only for *Cfv* a single gene (*pqqL*) was detected in all strains while being absent from *Cff* and *Cfvi*. The *pqqL* gene encodes a zinc-dependent peptidase, which in *E. coli* has been implicated in host adaptive functions in systemic infection processes in mammals<sup>35,36</sup>. Its presence in all *Cfv* genomes suggests that it might play an important role in the biology of this subspecies. Further investigations on the function of this gene could provide information on its contribution to *Cfv* survival, adaptation, or pathogenicity. In addition, this new marker could open a door to the development of more efficient *Cfv* typing methods at the biovar level, replacing current procedures (such as H<sub>2</sub>S production test and conventional PCR)<sup>8,32</sup>. However, a larger number of *C. fetus* genomes needs to be analysed to validate our observations.

The functional analysis of the *C. fetus* genomes showed that, similar to other *Campylobacter* species, the core genome of *C. fetus* is enriched in genes encoding fundamental processes for cell survival<sup>37,38</sup>. Compared to the core genome, the accessory genome usually encodes functions related to pathogenicity and adaptation to different environments or hosts and is not usually involved in functions essential for the survival of microorganisms<sup>38</sup>. In our study, we observed that the most abundant accessory genes encode mechanisms of cellular defence and adaptation to different environments. The subspecies of *C. fetus* are ecologically distinct, as *Cfv* is a pathogen restricted to the genital tract of cattle while *Cff* is considered an intestinal pathogen and can invade a wide variety of hosts<sup>4-7</sup>.

Our analysis of accessory genome functionality by groups showed significant differences between *C. fetus* subspecies and biovar with *Cfv* and *Cfvi* having a higher proportion of genes involved in quorum sensing and membrane transport as secretion systems compared to *Cff*. In other bacterial species such as *Bacillus*, quorum sensing has been shown to control gene expression synchronously in bacterial clusters during biofilm formation and virulence factor secretion<sup>39</sup>. Secretion systems facilitate the transfer of genetic material between bacteria; particularly in *Cfv*, the Type IV secretion system is involved in membrane transport functions and in the horizontal gene transfer<sup>40</sup>. It was also observed that the Spanish *Cfvi* strains presented a higher proportion of genes involved in bacterial infectious diseases processes that have been related to the ability to colonize hosts more efficiently<sup>24,41-43</sup>. Our study provides generic insight into the differences in accessory genome composition and functionality among mammalian-associated *C. fetus* subspecies. However, the significant proportion of unknown genes or genes with unidentified function, mainly in the accessory genome, underlines the strong need for further extensive analysis of the *C. fetus* genome. This will enable a more precise comprehension of the differences between subspecies and biovar, providing a more complete understanding of its genetic and functional diversity.

## Conclusions

Whole genome-based analysis allowed accurate identification of *C. fetus* subspecies and biovar and the assessment of their genome diversity. Taxonomic classification of Spanish *C. fetus* genomes revealed two novel MLST types (ST-76 and ST-77) for *Cfv* genomes and one novel ST (ST-78) for a *Cff* genome. Phylogenomic analyses clustered the *C. fetus* genomes at the subspecies and biovar level and separated the *Cfvi* Spanish genomes from *Cfvi* genomes from other countries, suggesting a regional variant circulating in Spain. Using GWAS, the *pqqL* gene was identified as uniquely present in all *Cfv* genomes thus being a potential candidate for the development of new and more accurate methods for the identification of *Cfv* biovars. Although significant differences in the functionality of the accessory genome of *C. fetus* subspecies were observed, further explorations are needed to determine the precise genetic and functional differences between *C. fetus* subspecies and biovar.

## Methods

### Strain selection

Thirty-three Spanish *C. fetus* isolates from bull preputial scraping (n = 31) and faeces (n = 1), and from cow vaginal mucus (n = 1) were selected for whole-genome sequencing (WGS) (Table S6); all isolates had been previously characterized in our laboratory using both phenotypic (biochemistry) and molecular methods (PCR)<sup>44</sup>. For genome comparative analyses, complete *C. fetus* genomes were retrieved from GenBank using the following search workflow: GenBank > Genome > *Campylobacter fetus* > Genome Assembly and Annotation Report (accessed, 15 May 2023). Only genomes from mammal-associated *C. fetus* subspecies and with scaffolds counts below 400 were considered. The quality of the GenBank genomes was then assessed with CheckM v1.1.6<sup>45</sup>, and those with less than 95% completeness and more than 5% contamination were discarded. Thus, a total of 62 *C. fetus* genomes from countries other than Spain met the inclusion criteria, i.e., 40 from cattle, six from sheep and 16 from humans (Table S6).

### Bacterial culture, genomic DNA extraction, whole-genome sequencing and assembly

The 33 Spanish *C. fetus* isolates were obtained from the isolate collections of SALUVET, NEIKER and the Spanish Reference Laboratory for Animal Campylobacteriosis (Laboratorio Central de Veterinaria: LCV, Madrid, Spain)

(Table S6). These isolates had been stored at  $-80^{\circ}\text{C}$  and were grown on 5% sheep blood-enriched Columbia agar (COS, Biomerieux, Marcy-l'Étoile, France) at  $37^{\circ}\text{C}$  under microaerophilic conditions (5%  $\text{O}_2$ , 10%  $\text{CO}_2$  and 85%  $\text{N}_2$ , GENbox Microaer, Biomerieux, Marcy-l'Étoile, France) for 48 h. DNA extraction was performed from single colony cultures following the protocol for gram-negative bacteria of the DNeasy Blood and tissues kit (QIAGEN, Hilden, Germany). Genomic DNA was sent to an external company for sequencing where library preparation was performed using the NEBNext Ultra™ II FS DNA Library Prep Kit (Illumina) and Illumina NovaSeq 6000 system was used to generate  $2 \times 150$  bp paired-end reads. The quality of the raw reads was assessed using FastQC v.0.11.9<sup>46</sup>. Trimmomatic v.0.38<sup>47</sup> was used for Illumina adapter removal and low-quality reads (reads with a quality score  $< 25$  over a sliding window size of 15 bp, and reads with a sequence length  $< 125$  bp) were filtered out using PRINSEQ v.0.20.4<sup>48</sup>. The resulting good quality reads were de novo assembled with SPAdes v.3.15.3<sup>49,50</sup>. The quality of the assemblies was assessed with QUAST v.5.0.2<sup>51</sup> and contigs of less than 500 bp were discarded with PRINSEQ v.0.20.4<sup>48</sup>.

### Genome taxonomic identification and MLST profile definition

Taxonomic classification of the genomes at the subspecies level was carried out by exact alignment of k-mers with Kraken2<sup>26</sup>. To confirm Kraken2 identification of *Cff* and *Cfv*, the draft genomes were subjected to in silico PCR targeting the *C. fetus*-specific *nahE* gene and the *Cfv*-specific *ISCfe1* insertion element (PCR1)<sup>13</sup> with the online tool Sequence Manipulation Suite (SMS) ([http://www.bioinformatics.org/sms2/pcr\\_products.html](http://www.bioinformatics.org/sms2/pcr_products.html)). Further taxonomic classification of the *Cfv* genomes to biovar level (*Cfv* or *Cfvi*) was achieved by in silico PCR targeting the L-Cyst transporter region (PCR2)<sup>32</sup>.

To determine the MLST STs, the draft genomes were queried against the *C. fetus* PubMLST<sup>52</sup> scheme using mlst v.2.19.0 (<https://github.com/tseemann/mlst>). New allelic variants and ST profiles derived from this study were submitted to PubMLST (<http://pub-mlst.org>) for new ST definition.

### Pangenome analysis

Gene prediction and annotation of the assembled genomes were performed with Prokka v1.11<sup>53</sup> using a curated database of *C. fetus* genomes owned by the Department of Biomolecular Health Sciences, Faculty of Veterinary Medicine, University of Utrecht. The annotated assemblies in Gff3 format were used as input for pangenome calculation using Roary v3.13.0<sup>54</sup> with a minimum percentage of identity for BlastP of 90. Pangenome gene categories were defined as follows: core genes (shared by 99–100% of the genomes studied); soft-core genes (95–99%); shell genes (15–95%); and cloud genes (0–15%). Parsnp v1.2<sup>55</sup> was used to construct a phylogenetic tree based on the core genome SNPs (Single Nucleotide Polymorphisms) among the 95 *C. fetus* genomes, using genome *Cff* 82–40 as reference. The resulting tree was rooted at the midpoint with Figtree v1.4.2 and visualised with iTOL<sup>56</sup>. FastTree v2.1.10<sup>57</sup> was used to construct a dendrogram based on the presence/absence of accessory genes. The Python roary\_plots.py script included in Roary was used to illustrate the pangenome gene presence/absence matrix. A genome-wide association study (GWAS) was performed to identify the genes associated with the *C. fetus* subspecies and biovar *intermedius*. This study was performed with Scoary v1.6.16<sup>58</sup> and the criteria for selecting genes included an Odds Ratio  $\geq 1$ , a Native *p* value of  $\leq 0.0075$ , a Benjamini-Hochberg (B-H) adjusted *p* value of  $\leq 0.05$  and a best pairwise score of  $\leq 0.5$ .

The genes comprising the different pangenome categories were transcribed into proteins using Geneious Prime v2023.0.3 software (Biomatters) for functional categories analysis. The obtained transcripts were used as input for the online bioinformatics tool BlastKOALA (<http://www.kegg.jp/blastkoala/>)<sup>59</sup>, an automated KEGG annotation and mapping server for functional categorisation of the gene set. The graphical representation of the different gene clusters and their functionalities was generated via ggplot2 v.3.4.0 R-package. To assess differences in proportion of gene functions among the different *C. fetus* subspecies (four groups in total, with Spanish *Cfvi* treated as a separate group), we conducted a Generalized Linear Model (GLM) analysis in R. Subspecies was used as the predictor variable and the relative abundance of gene functions served as the response variable. Subsequently, Bonferroni correction was applied on *p*-values to account for multiple testing, and results were considered statistically significant when  $p_{\text{adj}} < 0.05$ .

Principal component analysis (PCA) based on the presence/absence of pangenome genes and accessory gene function was calculated with R. Firstly, prcomp function from stats package v.4.2.2 was used to calculate the principal components, and then factoextra v.1.0.7 was implemented to account for the different PCA features calculation and access. Finally, the 3D graphical representation was possible using rgl v.1.0.1. Differences in accessory gene functionality between *C. fetus* subspecies were calculated by PERMANOVA using Bonferroni correction, considering  $p_{\text{adj}} < 0.05$  as statistically significant.

### Ethical approval and consent to participate

Sample collection was carried out by veterinary practitioners strictly following Spanish ethical guidelines and animal welfare regulations (Real Decreto 53/2013). The collection of this material, being considered a routine veterinary practice, did not require the approval of the Ethics Committee for Animal Experimentation. Informed oral consent was obtained from the farm owners at the time of sample collection. All methods were performed in accordance with the relevant guidelines and regulations and complied with ARRIVE guidelines<sup>60</sup>.

### Data availability

The Whole-Genome Sequencing project of *Campylobacter fetus* is deposited at GenBank under the BioProject accession number PRJNA1019261 and includes the 33 Spanish *Campylobacter fetus* genomes sequenced in this study, listed under the BioSample codes described in Table S6. Other datasets used and/or analysed during the current study are available from the corresponding author upon request.

Received: 27 November 2023; Accepted: 15 February 2024

Published online: 22 February 2024

## References

- van Bergen, M., van Putten, J. P. M., Dingle, K. E., Blaser, M. J. & Wagenaar, J. A. Isolation, identification, subspecies differentiation, and typing of *Campylobacter fetus*. In *Campylobacter* (eds Nachamkin, I. et al.) 213–225 (ASM Press, 2008).
- Fitzgerald, C. et al. *Campylobacter fetus* subsp. *testudinum* subsp. nov., isolated from humans and reptiles. *Int. J. Syst. Evol. Microbiol.* **64**, 2944–2948 (2014).
- Gilbert, M. J. et al. Comparative genomics of *Campylobacter fetus* from reptiles and mammals reveals divergent evolution in host-associated lineages. *Genome Biol. Evol.* **8**, 2006–2019 (2016).
- Sprenger, H., Zechner, E. L. & Gorkiewicz, G. So close and yet so far—Molecular microbiology of *Campylobacter fetus* subspecies. *Eur. J. Microbiol. Immunol.* **2**, 66–75 (2012).
- Nadin-Davis, S. A. et al. A comparison of fourteen fully characterized mammalian-associated *Campylobacter fetus* isolates suggests that loss of defense mechanisms contribute to high genomic plasticity and subspecies evolution. *PeerJ* **9**, e10586 (2021).
- Sahin, O., Yaeger, M., Wu, Z. & Zhang, Q. *Campylobacter* associated diseases in animals. *Annu. Rev. Anim. Biosci.* **5**, 21–42 (2017).
- Depner, K. Assessment of listing and categorisation of animal diseases within the framework of the Animal Health Law (Regulation (EU) No 2016/429): Trichomonosis. *EFSA J.* **15**, 4490 (2017).
- WOAH. Bovine genital campylobacteriosis. In *Manual of diagnostic tests and vaccines for terrestrial animals, chapter 3.4.4.* [https://www.oie.int/fileadmin/Home/eng/Health\\_standards/tahm/3.04.04\\_BGC.pdf](https://www.oie.int/fileadmin/Home/eng/Health_standards/tahm/3.04.04_BGC.pdf) (2021).
- Véron, M. & Chatelain, R. Taxonomic study of the genus *Campylobacter* Sebald and Veron and designation of the neotype strain for the type species, *Campylobacter fetus* (Smith and Taylor) Sebald and Veron. *Int. J. Syst. Bacteriol.* **23**, 122–134 (1973).
- van der Graaf-van Bloois, L. et al. First closed genome sequence of *Campylobacter fetus* subsp. *venerealis* bv. *intermedius*. *Genome Announc.* **2**, e01246 (2014).
- van der Graaf-Van Bloois, L. et al. Inconsistency of phenotypic and genomic characteristics of *Campylobacter fetus* subspecies requires reevaluation of current diagnostics. *J. Clin. Microbiol.* **52**, 4183–4188 (2014).
- Abril, C. et al. Discovery of insertion element ISCfe1: A new tool for *Campylobacter fetus* subspecies differentiation. *Clin. Microbiol. Infect.* **13**, 993–1000 (2007).
- van der Graaf-van Bloois, L. et al. Evaluation of molecular assays for identification *Campylobacter fetus* species and subspecies and development of a *C. fetus* specific real-time PCR assay. *J. Microbiol. Methods* **95**, 93–97 (2013).
- Farace, P. et al. *Campylobacter fetus* releases S-layered and immunoreactive outer membrane vesicles. *Rev. Argent. Microbiol.* <https://doi.org/10.1016/J.RAM.2021.06.001> (2021).
- On, S. L. W. & Harrington, C. S. Evaluation of numerical analysis of PFGE-DNA profiles for differentiating *Campylobacter fetus* subspecies by comparison with phenotypic, PCR and 16S rDNA sequencing methods. *J. Appl. Microbiol.* **90**, 285–293 (2001).
- Wagenaar, J. A., van Bergen, M. A. P., Newell, D. G., Grogono-Thomas, R. & Duim, B. Comparative study using amplified fragment length polymorphism fingerprinting, PCR genotyping, and phenotyping to differentiate *Campylobacter fetus* strains isolated from animals. *J. Clin. Microbiol.* **39**, 2283–2286 (2001).
- Bessède, E., Solecki, O., Sifré, E., Labadi, L. & Mé Graud, F. Identification of *Campylobacter* species and related organisms by matrix assisted laser desorption ionization-time of flight (MALDI-TOF) mass spectrometry. *Clin. Microbiol. Infect.* **17**, 1735–1739 (2011).
- van Bergen, M. et al. Clonal nature of *Campylobacter fetus* as defined by multilocus sequence typing. *J. Clin. Microbiol.* **43**, 5888–5898 (2005).
- Iraola, G. et al. Distinct *Campylobacter fetus* lineages adapted as livestock pathogens and human pathobionts in the intestinal microbiota. *Nat. Commun.* **8**, 1367 (2017).
- Iraola, G. et al. A rural worker infected with a bovine-prevalent genotype of *Campylobacter fetus* subsp. *fetus* supports zoonotic transmission and inconsistency of MLST and whole-genome typing. *Eur. J. Clin. Microbiol. Infect. Dis.* **34**, 1593–1596 (2015).
- van der Graaf-van Bloois, L. et al. Whole genome sequence analysis indicates recent diversification of mammal-associated *Campylobacter fetus* and implicates a genetic factor associated with H<sub>2</sub>S production. *BMC Genom.* **17**, 713 (2016).
- Kienesberger, S. et al. Comparative genome analysis of *Campylobacter fetus* subspecies revealed horizontally acquired genetic elements important for virulence and niche specificity. *PLoS ONE* **9**, e85491 (2014).
- Pena-Fernández, N. et al. Prevalence of bovine genital campylobacteriosis, associated risk factors and spatial distribution in spanish beef cattle based on veterinary laboratory database records. *Front. Vet. Sci.* **8**, 750183 (2021).
- Moolhuijzen, P. M. et al. Genomic analysis of *Campylobacter fetus* subspecies: Identification of candidate virulence determinants and diagnostic assay targets. *BMC Microbiol.* **9**, 86 (2009).
- Abdel-Glil, M. Y. et al. Genomic epidemiology of *Campylobacter fetus* subsp. *venerealis* from Germany. *Front. Vet. Sci.* **9**, 1069062 (2023).
- Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 1–13 (2019).
- Valenzuela-González, F., Martínez-Porchas, M., Villalpando-Canchola, E. & Vargas-Albore, F. Studying long 16S rDNA sequences with ultrafast-metagenomic sequence classification using exact alignments (Kraken). *J. Microbiol. Methods* **122**, 38–42 (2016).
- Imai, K. et al. Rapid and accurate species identification of mitis group *Streptococci* using the MinION Nanopore sequencer. *Front. Cell. Infect. Microbiol.* **1**, 11 (2020).
- Albert, M. J., Al-Hashem, G. & Rotimi, V. O. Multiplex gyrB PCR assay for identification of *Acinetobacter baumannii* is validated by whole genome sequence-based assays. *Med. Princ. Pract.* **31**(5), 493–496 (2022).
- Silva, M. F. et al. Genomic and phenotypic characterization of *Campylobacter fetus* subsp. *venerealis* strains. *Microorganisms* **9**, 1–14 (2021).
- Farace, P. D. et al. Phylogenomic analysis for *Campylobacter fetus* occurring in Argentina. *Vet. World* **14**(5), 1165–1179 (2021).
- Farace, P. D. et al. L-cysteine transporter-PCR to detect hydrogen sulfide-producing *Campylobacter fetus*. *PeerJ* **7**, e7820 (2019).
- Polo, C. et al. Evaluation of PCR assays for *Campylobacter fetus* detection and discrimination between *C. fetus* subspecies in bovine preputial wash samples. *Theriogenology* **172**, 300–306 (2021).
- Buchanan, C. J. et al. A genome-wide association study to identify diagnostic markers for human pathogenic *Campylobacter jejuni* strains. *Front. Microbiol.* **8**, 1224 (2017).
- Subashchandrabose, S., Smith, S. N., Spurbeck, R. R., Kole, M. M. & Mobley, H. L. T. Genome-wide detection of fitness genes in uropathogenic *Escherichia coli* during systemic infection. *PLoS Pathog.* **9**, 1–15 (2013).
- Jung, D. et al. Comparative genomic analysis of *Escherichia coli* isolates from cases of bovine clinical mastitis identifies nine specific pathotype marker genes. *Microb. Genom.* **7**, 000597 (2021).
- Zhong, C., Qu, B., Hu, G. & Ning, K. Pangenome analysis of *Campylobacter*: insights on the genomic diversity and virulence profile. *Microbiol. Spectr.* **10**, e0102922 (2022).
- Ali, A. et al. *Campylobacter fetus* subspecies: Comparative genomics and prediction of potential virulence targets. *Gene* **508**, 145–156 (2012).
- Rutherford, S. T. & Bassler, B. L. Bacterial quorum sensing: Its role in virulence and possibilities for its control. *Cold Spring Harb. Perspect. Med.* **2**, a012427 (2012).



40. Christie, P. J., Bhatti, M. & Laverde Gomez, J. A. The expanding bacterial type IV secretion lexicon. *Res. Microbiol.* **164**, 620–639 (2013).
41. Zhao, H. *et al.* Functional study of gene hp0169 in *Helicobacter pylori* pathogenesis. *Microb. Pathog.* **104**, 225–231 (2017).
42. Melvin, J. A., Scheller, E. V., Miller, J. F. & Cotter, P. A. *Bordetella pertussis* pathogenesis: Current and future challenges. *Nat. Rev. Microbiol.* **12**, 274 (2014).
43. Chen, J., Reyes, M., Clarke, M. & Shuman, H. A. Host cell-dependent secretion and translocation of the LepA and LepB effectors of *Legionella pneumophila*. *Cell. Microbiol.* **9**, 1660–1671 (2007).
44. Pena-Fernández, N. *et al.* Biochemical and molecular characterization of *Campylobacter fetus* isolates from bulls subjected to bovine genital campylobacteriosis diagnosis in Spain. [BMC Vet. Res., under review]. (2024)
45. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells and metagenomes. *Genome Res* **25**, 1043–55 (2015).
46. Andrews, S. Babraham Bioinformatics - FastQC a quality control tool for high throughput sequence data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (2010).
47. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
48. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**, 863–864 (2011).
49. Pribelski, A., Antipov, D., Meleshko, D., Lapidus, A. & Korobeynikov, A. Using SPAdes de novo assembler. *Curr. Protoc. Bioinforma.* **70**, e102 (2020).
50. Bankevich, A. *et al.* SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
51. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).
52. Jolley, K. A., Bray, J. E. & Maiden, M. C. J. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res.* **3**, (2018).
53. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
54. Page, A. J. *et al.* Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691–3693 (2015).
55. Treangen, T. J., Ondov, B. D., Koren, S. & Phillippy, A. M. The harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol.* **15**, 524 (2014).
56. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL): An online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127–128 (2007).
57. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).
58. Brynildsrud, O., Bohlin, J., Scheffer, L. & Eldholm, V. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biol.* **17**, 1–9 (2016).
59. Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **428**, 726–731 (2016).
60. Percie du Sert, N. *et al.* The ARRIVE guidelines 2.0: Updated guidelines for reporting animal research. *J. Cereb. Blood Flow Metab.* **40**, 1769–1777 (2020).

## Acknowledgements

The authors express their gratitude to Montserrat Agüero and Iratxe Pérez, from the Laboratorio Central de Veterinaria de Algete, for generously providing the *C. fetus* bacterial strains.

## Author contributions

G.A. and A.H. conceived the study and participated in its design. N.P.F. and N.K. performed laboratory analysis. N.P.F., M.O., J.L.L. and L.G. participated in the bioinformatics analysis. N.P.F. wrote the manuscript, with interpretation of results and discussion inputs from GA, AH, E.C., M.O., N.K., J.L.L. and L.G. All authors read and approved the final manuscript.

## Funding

This study was supported by the research projects RTA2017-00076-00-00 funded by MCIN/AEI/<https://doi.org/10.13039/501100011033> and FEDER “A way to do Europe”, as well as the Department of Economic Development, Sustainability, and Environment of the Basque Government (Project URAGAN 21-00012). N.P.F. was the recipient of the grant for doctoral formation Pre2018-086113 funded by MCIN/AEI/<https://doi.org/10.13039/501100011033> and by “ESF Investing in your future”.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-54750-1>.

**Correspondence** and requests for materials should be addressed to A.H. or G.A.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024