Investigating Major Recurring *Campylobacter jejuni* Lineages in Luxembourg Using Four Core or Whole Genome Sequencing Typing Schemes

Morgane Nennig[1,2]*, Ann-Katrin Llarena[3], Malte Herold[1], Joël Mossong[1], Christian Penny[4], Serge Losch[5], Odile Tresse[2] and Catherine Ragimbeau[1]

[1] Epidemiology and Microbial Genomics, Laboratoire National de Santé, Dudelange, Luxembourg, [2] INRAE, Oniris, SECALIM, Nantes, France, [3] Faculty of Veterinary Medicine, Norwegian University of Life Sciences, Oslo, Norway, [4] Luxembourg Institute of Science and Technology, Environmental Research and Innovation Department, Belvaux, Luxembourg, [5] Laboratoire de Médecine Vétérinaire de l'Etat, Veterinary Services Administration, Dudelange, Luxembourg

Campylobacter jejuni is the leading cause of bacterial gastroenteritis, which has motivated the monitoring of genetic profiles circulating in Luxembourg since 13 years. From our integrated surveillance using a genotyping strategy based on an extended MLST scheme including *gyrA* and *porA* markers, an unexpected endemic pattern was discovered in the temporal distribution of genotypes. We aimed to test the hypothesis of stable lineages occurrence by implementing whole genome sequencing (WGS) associated with comprehensive and internationally validated schemes. This pilot study assessed four WGS-based typing schemes to classify a panel of 108 strains previously identified as recurrent or sporadic profiles using this in-house typing system. The strain collection included four common lineages in human infection (N = 67) initially identified from recurrent combination of ST-*gyrA-porA* alleles also detected in non-human samples: veterinary (N = 19), food (N = 20), and environmental (N = 2) sources. An additional set of 19 strains belonging to sporadic profiles completed the tested panel. All the strains were processed by WGS by using Illumina technologies and by applying stringent criteria for filtering sequencing data; we ensure robustness in our genomic comparison. Four typing schemes were applied to classify the strains: (i) the cgMLST SeqSphere+ scheme of 637 loci, (ii) the cgMLST Oxford scheme of 1,343 loci, (iii) the cgMLST INNUENDO scheme of 678 loci, and (iv) the wgMLST INNUENDO scheme of 2,795 loci. A high concordance between the typing schemes was determined by comparing the calculated adjusted Wallace coefficients. After quality control and analyses with these four typing schemes, 60 strains were confirmed as members of the four recurrent lineages regardless of the method used (N = 32, 12, 7, and 9, respectively). Our results indicate that, regardless of the typing scheme used, epidemic or endemic signals were detected as reflected by lineage B (ST2254-*gyrA*9-*porA*1) in 2014 or lineage A (ST19-*gyrA*8-*porA*7), respectively. These findings support the clonal expansion of stable genomes in

Campylobacter population exhibiting a multi-host profile and accounting for the majority of clinical strains isolated over a decade. Such recurring genotypes suggest persistence in reservoirs, sources or environment, emphasizing the need to investigate their survival strategy in greater depth.

INTRODUCTION

Campylobacter spp. is the leading cause of bacterial foodborne diarrheal disease worldwide (WHO, 2013) and the main zoonotic agent in the European Union (EU) (EFSA and ECDC, 2019). In 2018, the reported EU-wide incidence of campylobacteriosis was 64.1 cases per 100,000 population and Luxembourg had one of the highest rates in Europe (103.8) (EFSA and ECDC, 2019). *Campylobacter* is responsible for a large health and economic burden world-wide with a cost-of-illness of $1.56 billion in the USA (Scharff, 2012; Devleesschauwer et al., 2017) and 8.28 disability adjusted life years (DALYs) per 100,000 population in Europe (Cassini et al., 2018).

More than 80% of cases of campylobacteriosis are caused by *Campylobacter jejuni* and poultry is considered the main reservoir of human infections (Mughini-Gras et al., 2012; Ragimbeau et al., 2014; Mossong et al., 2016; EFSA and ECDC, 2019). Transmission is commonly associated with cross-contamination during handling of raw meat, the consumption of undercooked meat or raw drinking milk (EFSA and ECDC, 2018). *C. jejuni* lives as a commensal bacterium in the gastrointestinal tract of wild and domestic birds and mammals, including cattle and sheep. Environmental transmission routes are less frequently reported, but risks include exposure during outdoor sports, swimming in natural waters or contact with garden soil (Stuart et al., 2010; Ellis-Iversen et al., 2012; Mughini-Gras et al., 2012; Bronowski et al., 2014; Mossong et al., 2016; Kuhn et al., 2018).

Unlike for other foodborne pathogens, molecular surveillance of *C. jejuni* has not been implemented in many European countries as the majority of human infections are thought to be sporadic with a low fatality rate (0.03% in EU in 2017) (EFSA and ECDC, 2019). Nevertheless, due to the high number of reported human cases in the EU, campylobacteriosis ranks third in cause of death behind listeriosis and salmonellosis. In addition, outbreaks caused by *Campylobacter* spp. are increasingly being identified and reported on a regular basis, often linked to consumption of untreated drinking water, raw milk or chicken liver paté (Jakopanec et al., 2008; Revez et al., 2014; Davis et al., 2016; Lahti et al., 2017; Kang et al., 2019; Hyllestad et al., 2020).

The generally high incidence recorded in Luxembourg over the last decade has motivated a national implementation of molecular monitoring of *Campylobacter* circulating in food, farm animals, and environmental waters, as part of an integrated surveillance (Ragimbeau et al., 2008; Berthe et al., 2013; Ragimbeau et al., 2014; Mossong et al., 2016). Monitoring the *C. jejuni* population circulating in a community can function as early warning signals for outbreaks and detect long-term changes in the bacterial population, such as emerging new virulence traits or antimicrobial resistance. Further, monitoring the types of *C. jejuni* in different reservoirs and environments can shed light on the epidemiology of campylobacteriosis in that region.

Initially, genotypes from the molecular monitoring were defined according to an in-house typing system originally developed for the Sanger sequencing method. This typing method consists of the seven housekeeping genes from the Multi Locus Sequence Typing (MLST) method (Maiden et al., 1998; Dingle et al., 2001) combined with allelic profiles from two additional loci: *porA* (Clark et al., 2007) and *gyrA* (Wang et al., 1993). Including *porA* and *gyrA* refines the resolution scale of MLST and creates a reliable extended MLST typing method. The *porA* locus encodes the major outer membrane protein and is highly polymorphic, but stable during human passage and within family outbreaks making it a suitable molecular marker for epidemiologic investigations (Cody et al., 2009). Jay-Russell et al. (2013) supported this finding by utilizing variations in *porA* sequences as a screening tool for discriminating genetically related strains in the situation of a large outbreak (Jay-Russell et al., 2013). Interestingly, specific point mutations within *porA* were identified as markers of hyper virulence for a *C. jejuni* clone causing abortion in ruminants and foodborne disease in humans (Sahin et al., 2012; Wu et al., 2016). A sequence-based *gyrA* method was recently developed and it provides information of isolates in two respects: (i) to distinguish the major nucleotide mutation (C257T) conferring the quinolone resistance (i.e., the peptide shift Thr86Ile), and (ii) to source-track clinical isolates according to a host signature in *gyrA* alleles, potentially predictive of domestic birds as source (Jesse et al., 2006; Ragimbeau et al., 2014). The discriminative power resulting from this extended MLST method indexed on a 9-loci basis is sufficient to define different lineages and human clusters (Dingle et al., 2008; Ragimbeau et al., 2014). This has recently been superseded by whole genome sequencing (WGS).

The advent of Next Generation Sequencing (NGS) technologies has significantly increased the amount of genetic information available for the characterization of bacterial isolates. Comparisons at the genome level are more relevant for defining relationships between isolates at unprecedented resolution while simultaneously allowing the full characterization of the virulome, resistome, and metabolome of the isolate. Phylogenetic approaches based on WGS data rely on calculating genetic

distances based on either SNPs (single nucleotide polymorphism) or allele differences (ADs) [known as core or whole genome MLST (cg/wgMLST)] (ECDC, 2016). Unlike other common food and waterborne bacterial pathogens [*Listeria monocytogenes* (Ragon et al., 2008) or *Salmonella enterica serovar Typhi* (Lan et al., 2009)], *Campylobacter* populations display high genetic diversity likely driven by horizontal genetic exchange (de Boer et al., 2002; Sheppard et al., 2011) and to a lesser extent by chromosomal mutations. As a result, SNP analyses that compare strains at the nucleotide level tend to overestimate genetic exchange events and, consequently, decimate the signals of the *Campylobacter* population structure (Sheppard et al., 2012). After conducting comparative studies between the SNP and the cgMLST approaches for different pathogens, it appears that the gene-by-gene approach is more suitable for identifying lineages with this recombining species (Dangel et al., 2019; Jajou et al., 2019). This gene-by-gene method defines allelic profiles from a set of common loci, known as core genome common to a representative panel of isolates. Including accessory loci, present in only a subsection of genomes and often associated with specific phenotypic traits of interest, improves the discriminatory power of the gene-by-gene analysis (Sheppard et al., 2012). For WGS analysis of *C. jejuni* and *C. coli*, several typing schemes have been developed, including two cgMLST schemes; a commercial cgMLST schema containing 637 loci from the SeqSphere+ software (Ridom GmbH, Münster, Germany; www.cgMLST.org) and the Oxford cgMLST schema with 1,343 loci (Cody et al., 2017). Two wgMLST schemes were also defined for *C. jejuni/coli* within the SeqSphere+ software (including the cgMLST and 958 accessory loci) and by the Oxford University (1,643 loci) (Cody et al., 2013). Moreover, two typing schemes were developed specifically for *C. jejuni*: a cgMLST (678 loci) and a wgMLST (2,795 loci) from the INNUENDO platform (Llarena et al., 2018). The method-dependent definition of a WGS-based genotype underlines the need for an international nomenclature to improve communication in outbreak investigation and in surveillance.

Through vigilant surveillance and molecular subtyping with extended MLST, we discovered an unexpected endemic pattern in the temporal distribution of genotypes associated with human infection over several years. The aim of this study was to investigate if these strains were indeed clonal by applying a higher resolution typing method, namely the WGS gene-by-gene approach. We simultaneously assessed the concordance between the four different typing schemes developed for *Campylobacter* spp. and their ability to separate closely related strains.

MATERIALS AND METHODS

Strain Selection
Five thousand *C. jejuni* isolates, from human and non-human sources collected in Luxembourg between 2006 and 2018, were inspected. Years 2009 and 2010 were not included as no molecular surveillance data were available. Genotypic data associated with this collection included extended MLST profiles indexed on nine loci: 7 targets of MLST (Dingle et al., 2001), the partial sequence of *gyrA*

(Ragimbeau et al., 2014), and the Sequence Variable Region of *porA* (Campylobacter MOMP database; Dingle et al., 2008). The nomenclature for displaying the results of this extended MLST was defined as follows: sequence type (ST), *gyrA* (allele number), and *porA* (allele number). For example, the combination of alleles including ST19 associated with *gyrA* allele number 8 and *porA* allele number 7 is displayed as follows: ST 19-8-7.

From these, a panel of strains with identical ST-*gyrA*-*porA* profiles over four successive years was selected, including some strains with one allele variation in either the *gyrA* or *porA* loci. Care was taken to achieve a representative strain collection from all available sources (clinical, food, animal, and environmental) and years (between 2006 and 2018). Finally, we also selected a control panel of "sporadic" isolates from patients lacking a recent travel history, i.e., only domestic cases, and whose ST-*gyrA*-*porA* profile occurred only once between 2011 and 2018. This control panel was used as outgroup.

Culture, DNA Extraction, Library Preparation, and WGS
All isolates were stored in −80°C in FBP medium (a combination of ferrous sulfate, sodium metabisulfite, sodium pyruvate and glycerol) (Gorman and Adley, 2004). For each strain, a loopful of frozen culture was spread on chocolate PolyVitex plates (BioMerieux, Marcy-l'Etoile, France) and incubated under micro-aerobic conditions (5% O_2, 10% CO_2, 85% N_2) at 42°C for 48 h. Then, a subculture of one colony was made again on chocolate PolyVitex agar, and incubated 16 h in the above-mentioned conditions. DNA was extracted with the DNA QIAamp Mini Kit (Qiagen, The Netherlands) according to the manufacturer's instructions. DNA was quantified with the Qubit 2.0 Fluorometer (Invitrogen, Belgium) and the Qubit[®] dsDNA BR Assay Kit (Life Technologies, Belgium). The DNA concentration was adjusted to be within the range of 30 to 170 μg/ml for subsequent sequencing. Libraries were prepared using the Nextera[TM] DNA Flex Library Prep Kit or the Nextera[TM] XT DNA Library Preparation Kit and sequenced on the MiSeq or the MiniSeq platforms achieving either 150- or 250-bp paired-end reads. All chemistry and instrumentations are supplied by Illumina, San Diego, CA, USA. Sequenced raw reads have been uploaded to ENA and are available under the accession project number PRJEB40465.

Genome Assembly and Quality Control (QC) Criteria
For the cgMLSTs SeqSphere+ and Oxford, the paired-end raw read data were *de novo* assembled using Velvet Optimizer v.1.1.04 implemented in Ridom SeqSphere+ v6.1 (Ridom GmbH, Münster, Germany) (Jünemann et al., 2013). Velvet Optimizer was run with automatic determination of the coverage cut-off and minimum contig length and only assemblies with >30x coverage, 1.6Mb ± 10% bp in size and maximum number of 150 contigs were included in the downstream analysis (Zerbino and Birney, 2008; Cody et al., 2017). For the cgMLST INNUENDO and wgMLST

INNUENDO, the raw data were assembled into contigs using the INNUca pipeline v. 4.2.1 with default settings (Machado et al. 2017). Only profiles with no more than 2% of missing loci in either cgMLST were included in the comparative study.

WGS-Based Typing Schemes for Genome Comparison

cgMLST and Accessory Schemes in SeqSphere+

For SeqSphere+, an *ad hoc* cgMLST scheme (N = 637 loci) for *C. jejuni/C. coli* developed by the commercial firm Ridom SeqSphere+ and publicly available at www.cgMLST.org was used. Details of the material and methods used for defining this typing scheme were kindly provided by Prof. Dr. Harmsen (**Supplementary Data S1**). The cgMLST scheme consisted of 637 genes (https://www.cgmlst.org/ncs/schema/145039/locus/). Using genomic data from previously described local outbreaks, a Complex Type (CT) threshold of thirteen was defined to give guidance for delineation of possibly related from not-related genomes (Mellmann et al., 2004). In addition, cgMLST (v1.3) was merged with a screening of the alleles of the accessory genes (N = 958). Altogether, the combined typing wgMLST scheme targets 1,595 loci and the nomenclature remain the same as in the cgMLST analyses with the definition of CTs, solely based on core genome analyses, with a cluster alert of 13.

cgMLST Oxford Scheme

Cody et al. (2017) designed a cgMLST scheme composed of 1,343 loci, available as an open-access and web-accessible analyses online (PubMLST - Campylobacter Sequence Typing; Jolley et al., 2018). The system assigns a unique profile ID from each isolate sequences submitted. Clustering to identify groups can be performed by selecting a threshold empirically chosen (depending on the discrimination power needed). However for this study, the scheme was implemented in SeqSphere+ for comparing strains by using an in-house nomenclature.

cgMLST and wgMLST INNUENDO Schemes

The cgMLST and wgMLST schemes from INNUENDO include 678 and 2,795 loci, respectively, and are publicly available at Zenodo (https://zenodo.org/record/1322564#.X5l_4IhKg2y, Rossi et al., 2018). The cgMLST and wgMLST profiles of the INNUca assembled genomes produced in this study were called using chewBBACA suite (v 2.0.17.1) (Silva et al., 2018). The achieved cgMLST profiles were added to the cgMLST allelic profiles of the 6,526 *C. jejuni* genomes of the INNUENDO dataset, which is also available at Zenodo (Allele_Profiles/Cjejuni_cgMLST_alleleProfiles.tsv, https://zenodo.org/record/1322564#.X5l_4IhKg2y, Rossi et al., 2018). Minimum Spanning Trees (MST) and goeBURST distances were calculated using the goeBURST Full MST algorithm implemented in PHYLOViZ 2.0, and used to define L1:L2:L3 profiles for the cgMLST at 4, 59, and 292 loci variance (Feil et al., 2004; Francisco et al., 2009; Francisco et al., 2012; Nascimento et al., 2017; Llarena et al., 2018). This classification system is hierarchical: L1 is the level

representing the highest resolution with a threshold of 4 and it is applied for outbreak detection and investigation, L2 is the intermediate level and is used for long-term longitudinal monitoring. L3 is defined as the level with the highest concordance with the seven-gene MLST classification (Llarena et al., 2018). The wgMLST INNUENDO defines genotypes based on the combination of alleles from the 2,795 loci; no rules were initially developed for clustering isolates with similar profiles.

Comparison of the Targets Included in Each cgMLST Schemes

To crosslink loci with different naming conventions across the four typing schemes, we compared the allele sequences in a pairwise manner. Allele sequences for cgMLST SeqSphere+ were downloaded from https://www.cgmlst.org/ncs/schema/145039/. Allele sequences for cgMLST Oxford were downloaded *via* the pubMLST RESTful API (scheme 4) (Jolley et al., 2017). Allele sequences for cgMLST INNUENDO were downloaded from Zenodo (Rossi et al., 2018). We selected the first allele sequence for each loci of the four typing schemes and performed pairwise reciprocal best hit comparison for the three schemes with the rbh function of the MMseqs2 toolkit ver. 11.e1a1c (Mirdita et al., 2019) using nucleotide search including forward and reverse strand, as well as default parameters. Hits with bitscore above 100 were selected and connected across schemes with a custom script in R 3.4.4. (R Core Team, 2018) using the igraph package 1.2.5. (Csárdi and Nepusz, 2006). Sets of matching loci within the three schemes were visualized with the UpSetR 1.4.0. package (Conway et al., 2017).

Typing System Concordance

The adjusted Wallace coefficient (AWC) (Wallace, 1983; Severiano et al., 2011) was used to estimate the concordance between the different typing schemes in classifying strains (Pinto et al., 2008) by the online Comparing Partitions tool (http://www.comparingpartitions.info), using the strain panel (**Supplementary Data S2**). The degree of equivalence is reflected by AWC. It indicates the probability that two strains with the same type by one method are also categorized into the same type by another method.

Detection of wgMLST Targets Shared by Recurrent Lineages

To determine the overlap of detected wgMLST INNUENDO targets, the allelic profiles of all strains were compared. We extracted lists of targets that appeared at least once within each of the respective lineages in the collection of strains to determine and visualize overlapping and unique sets with a webtool (http://www.molbiotools.com/listcompare.html).

Cluster Analyses

In SeqSphere+, *Campylobacter* isolates are classified in CTs in which the first CT assigned chronologically is definitively fixed in

the database and referred to as the CT founder (Ridom SeqSphere+, 2013). In contrast, the goeBURST algorithm produces a hierarchical classification with the gene-by-gene approach and aims to predict the founder of a clonal complex based on the allele frequency in the dataset. It assumes that the ancestral genotype is the predominant one, which subsequently generates variants. To deduct and visualize, the possible evolutionary relationships between strains, the goeBURST algorithm and its expansion to generate a complete MST implemented in PHYLOViZ 2.0 was used for the cgMLSTs SeqSphere+, INNUENDO, Oxford, the cgMLST SeqSphere+ combined with the accessory targets and the wgMLST INNUENDO (https://online2.phyloviz.net/index) (Feil et al., 2004; Nascimento et al., 2017).

The dynamic shared-genome based approach was performed on the MST generated for the cgMLST Oxford and the wgMLST INNUENDO in order to determine a clustering threshold. Genomic clusters were determined according to the definition of goeBURST groups, based on allelic differences ranging from 0.5 to 1% (Llarena et al., 2018). The in-house nomenclature for displaying the results of cgMLST Oxford and wgMLST INNUENDO were defined as follows: Ox+number and wg +number. For example, the Ox profile number 10 and the wg profile number 8 are displayed as follows: Ox10 and wg8, respectively. The wgMLST profiles were used in the dynamic shared-genome based approach for the comparison and only to increase resolution for clustering strains.

RESULTS

Recurrent Extended MLST Profiles in Campylobacteriosis

By focusing solely on human clinical isolates from our historical collection (N = 3,000), we identified approximately one hundred distinct ST-*gyrA*-*porA* combinations. Two-thirds (N = 2,010) of the human strains in the collection belong to 108 main combinations (**Supplementary Data S3**). Four lineages (ST19-8-7, ST2254-9-1, ST464-8-1678, and ST6175-9-1625, hereafter referred to as lineage A, B, C, and D, respectively) were selected due to the high number of strains (N ≥ 45) and their frequency in human infection over time (**Figure 1** and **Supplementary Data S3**). Some minor variations were accepted in *gyrA* and *porA* alleles: three variants of *porA* and one of *gyrA* in lineage A, and one variant of *porA* in lineage B (**Table 1**). In lineage A, the variation of *gyrA* alleles (*gyrA*1 instead of *gyrA*8) leads to the loss of the quinolone resistance (Wang et al., 1993; Payot et al., 2006). Concerning the *porA* variations, two are linked to deletions in lineage A and two to a non-synonymous mutation (one in lineage A and one in lineage B, respectively). Lineage A has appeared regularly after 2005, with an average of five strains per year and up to 23 in 2012, while 68% of all the strains belonging to lineage B were gathered in a peak in 2014 (**Figure 1**). For lineage C, strains displayed the same combination of alleles and occurred once in August 2008 and then reemerged from July 2014 to January 2018 (**Figure 1**). For lineage D, strains were characterized by the same allele combination (**Table 1**) and occurred once in June
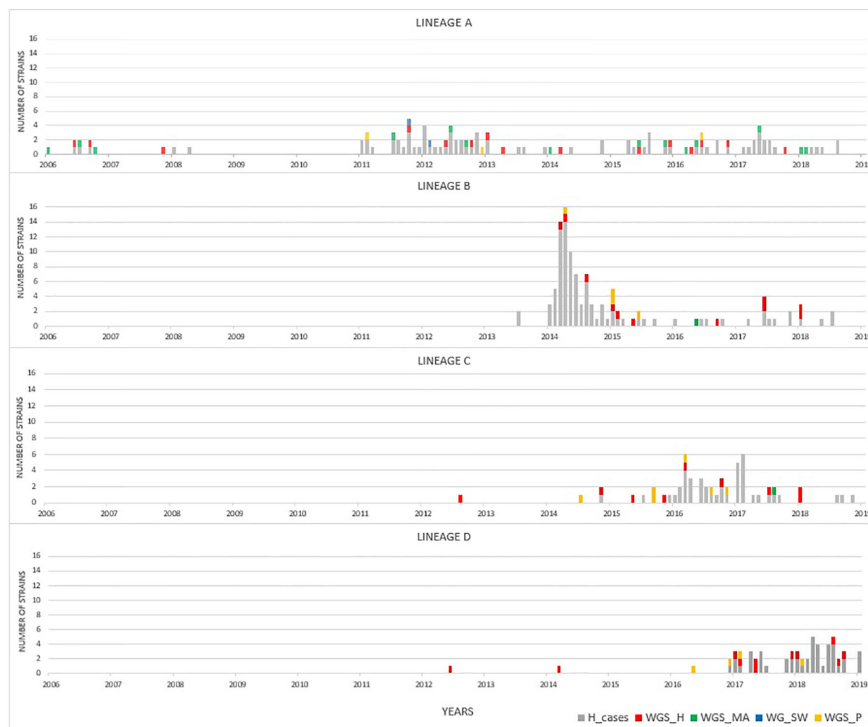
FIGURE 1 | Distribution of strains occurrence for lineages A to D over time. Clinical strains of the laboratory collection are displayed in gray (extended MLST typing). Colors represent to source of selected isolates that were analyzed by WGS: human (red), cattle and sheep (green), poultry (yellow), and surface water (blue) samples.

TABLE 1 | Distribution of main lineages, extended MLST, and variant of the strain collection.

Lineage	Main combination	Variants	Human	Poultry	Ruminants	Environmental	Total
A	ST 19 – *gyrA* 8 – *porA* 7	ST 19 – *gyrA* 8 – *porA* 582	13	3	14	2	37
		ST 19 – *gyrA* 8 – *porA* 2070	1	1	1		
		ST 19 – *gyrA* 8 – *porA* 2068	1				
		ST 19 – *gyrA* 1 – *porA* 7	1				
B	ST 2254 – *gyrA* 9 – *porA* 1	ST 2254 – *gyrA* 9 – *porA* 275	10	4	1	0	16
			1				
C	ST 464 – *gyrA* 8 – *porA* 1678		10	8	1	0	19
D	ST 6175 – *gyrA* 9 – *porA* 1628		12	4	1	0	17
Total			49	20	18	2	89

ST, Sequence Type.

2012, once in March 2014 and then regularly, from May 2016 to October 2018 (**Figure 1**).

Selection of a Strain Panel

Overall, the selected panel included strains from various sources as the four lineages occurring in human infections were also detected in non-human samples. Altogether, the collection included isolates from human (N = 67), poultry (N = 21), and ruminant (N = 18). To complete the panel, two strains from environmental sources (surface waters) assigned to lineage A were added (**Table 1**). A total of 108 strains was selected for the strain panel and subjected to WGS. To achieve equal distribution of strains over the study period, strains belonging to lineage A (N = 37 of 70), lineage B (N = 16 of 97), lineage C (N = 19 of 45), and lineage D (N = 17 of 58) were selected. In addition, 19 strains with a unique ST-*gyrA*-*porA* combination were included in the panel as an outgroup (**Supplementary Data S2**).

The acquired assemblies varied between 35× and 120× in depth of coverage and 1 to 150 contigs, associated with a percentage of good targets ranging from 98.6% to 99.8% (mean value = 99.3%) for cgMLST SeqSphere+ and from 98.0% to 99.3% (mean value = 98.4%) for the cgMLST Oxford. According to the quality criteria defined above (see Methods 2.3) as well as those of the INNUca pipeline, 15 genomes were discarded (14 with SeqSphere+ and 1 with INNUENDO criteria; 4, 1, 4 and 7 genomes were removed from lineages A, B, C and D, respectively). Consequently, genomes of 93 strains were included in the downstream analysis (**Supplementary Data S2**).

Comparison of the Loci Included in the Different Schemes

As the number of loci selected for the core genome varies between the schemes, we compared the respective sequences to assess the number of shared loci. We compared allele sequences by reciprocal best hits. All schemes shared 432 loci, constituting the majority of targets in cgMLST SeqSphere+ and cgMLST INNUENDO with 68% and 64% of targets respectively (**Figure 2**). The majority of targets that differed between cgMLST SeqSphere+ and cgMLST INNUENDO was present in cgMLST Oxford. The wgMLST INNUENDO had an additional 1,775 loci not present in any of the other three cgMLST schemas (**Supplementary Data S4**). The mean size of targets included in each cgMLST typing scheme ranges from 93 to 4,553 bp and

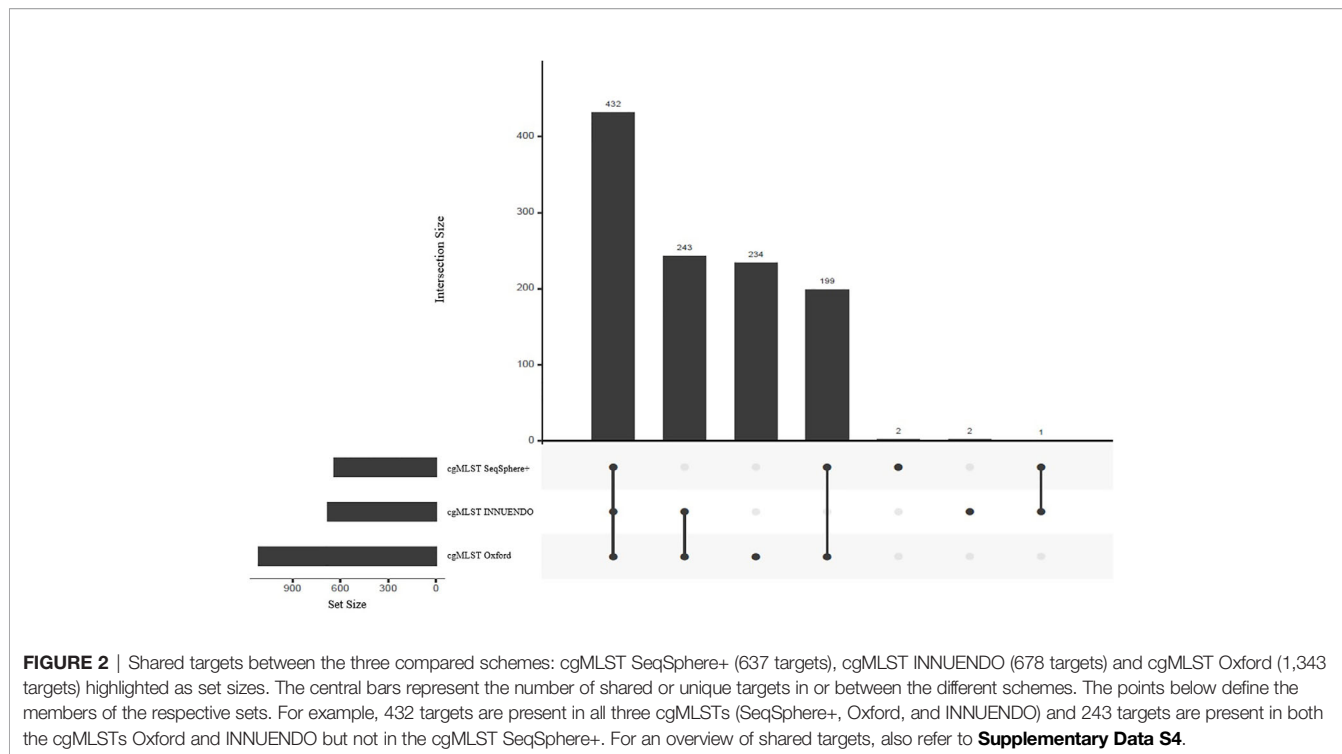the complete lists of targets are provided in **Supplementary Data S4**.

Gene-by-Gene WGS Analysis

With the dynamic shared-genome based approach using 1% allelic differences, thresholds of 11 and 9 AD were defined to classify the strains by the cgMLST Oxford and the wgMLST INNUENDO scheme respectively (**Table 2**). The number of partitions, or clusters, obtained with the different methods was very close: 28 for extended MLST, 22 for cgMLST SeqSphere+, 26 for cgMLST Oxford, and 24 for cgMLST INNUENDO. The largest number of partitions (N = 32) was obtained with the wgMLST INNUENDO analysis (**Supplementary Data S2**). From this pan-genome analysis including 2,795 targets, an average of 974 loci was detected in each lineage, with 870 loci shared between the four lineages (**Figure 3**).

For the analysis of unique combinations, all sporadic strains were classified distinctly by the typing schemes, with one exception regarding two strains that were classified in the same CT (CT 1639) with cgMLST SeqSphere+, in the same profile L1:L2:L3 (66:81:1) with cgMLST INNUENDO and in the same profile with wgMLST INNUENDO (wg30). The allelic profiles for the strains generated by all typing methods were clustered and visualized in PHYLOViZ online tool, in which all five typing schemes achieved very similar unrooted MSTs (**Figure 4**) (PHYLOViZ Online).

Lineage A (ST19-8-7, N = 34) had very limited genetic diversity according to our gene-by-gene WGS analyses. The cgMLST SeqSphere+ assigned all strains to the same CT (CT 82), as did the cgMLST Oxford scheme: Ox1 (**Table 3**). On the contrary, the cgMLST INNUENDO divided lineage A in two groups, of which the majority (33/34) were of the same L1:L2:L3 profile (1:9:1). The 34th strain had a different genotype at L1 level (2695:9:1). Thirty-two of 34 lineage A strains had an identical wgMLST INNUENDO profile: wg1, while two strains had a deviating wgMLST profile (wg5 and wg6). The strains belonging to profile wg1 were collected over a wide timespan (2006–2018) and a range of sources (human, veterinary, or environmental sources).

Lineage B (ST2254-9-1) had low genetic variability according to the cg/wgMLST analyses: altogether, 15 of 16 strains had a similar cgMLST SeqSphere+ (CT 51), cgMLST Oxford (Ox2), and cgMLST INNUENDO (19:49:4) profiles. The increased resolution offered by the wgMLST INNUENDO divided the

FIGURE 2 | Shared targets between the three compared schemes: cgMLST SeqSphere+ (637 targets), cgMLST INNUENDO (678 targets) and cgMLST Oxford (1,343 targets) highlighted as set sizes. The central bars represent the number of shared or unique targets in or between the different schemes. The points below define the members of the respective sets. For example, 432 targets are present in all three cgMLSTs (SeqSphere+, Oxford, and INNUENDO) and 243 targets are present in both the cgMLSTs Oxford and INNUENDO but not in the cgMLST SeqSphere+. For an overview of shared targets, also refer to **Supplementary Data S4**.

TABLE 2 | Characteristics of the different typing schemes to analyze WGS data from *C. jejuni*.

Typing scheme	Number of targets	Cluster Alert distance*
Extended MLST	9	1
cgMLST SeqSphere+	637	13
cgMLST Oxford	1,343	11
cgMLST INNUENDO	678	L1: 4, L2: 59, and L3: 292
wgMLST INNUENDO	2,795	9

The cluster alert distance is defined by a threshold value corresponding to the maximum number of different alleles between strains belonging to the same cluster.
MLST, Multi Locus Sequence Typing; cg, core genome; wg, whole genome.

strains in three types: 75% of the strains were of wg2 while the remaining quarter was divided between wg7 and wg8. The strains belonging to the genotype wg2 were isolated from 2014 to 2018 and from diverse sources (**Table 4**).

Lineage C (ST464-8-1678) was more variable than A and B: all 15 strains were of the CT 75 and the 29:70:7 according to the cgMLST SeqSphere+ and cgMLST INNUENDO, respectively. Contrary to this, the cgMLST Oxford split the panel into three: Ox3, Ox5, and Ox6 (**Table 5**). The wgMLST INNUENDO discriminated six different genotypes collected from diverse range of sources between 2014 and 2017 (**Table 5**).

For lineage D (ST6175-9-1625), all the 10 strains were gathered by the cgMLST SeqSphere+ in the same CT (CT 543), while with the cgMLSTs Oxford and INNUENDO and the wgMLST INNUENDO, one strain had a different profile from the others. The strains were isolated between 2017 and 2018 and from diverse sources (**Table 6**).
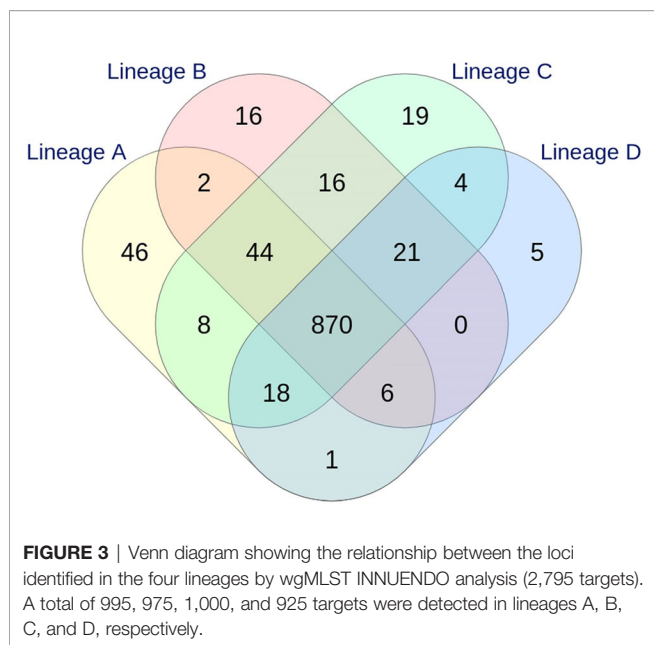
Concordance Between the Typing Methods

This analysis was performed on the 93 strains selected in the panel (Methods 3.2 and **Supplementary Data S2**). The cgMLST INNUENDO, the cgMLST Oxford, and the wgMLST INNUENDO had an AWC of 1.000 to the cgMLST SeqSphere+ schema, meaning that all strains clustering together using one of these three typing schemes are also classified together with the cgMLST SeqSphere+. The cgMLST Oxford had an AWC of 0.948 with the cgMLST INNUENDO and, conversely, the cgMLST INNUENDO had an AWC of 0.956 with cgMLST Oxford; 95% of the strains are clustered similarly using either cgMLSTs Oxford and INNUENDO. The majority (93.7% and 94.5%) of the strains that clustered with the cgMLST SeqSphere+ schema were also grouped by the cgMLST INNUENDO and the cgMLST Oxford, respectively. The wgMLST INNUENDO bundled 94.0% of the strains in a similar manner as the cgMLST INNUENDO and 99.8% as the cgMLST Oxford (**Table 7**).

DISCUSSION

From our long-term surveillance of campylobacteriosis at national scale, our data suggested the presence of recurring genotypes defined by an extended MLST method indexing 9-loci over a 13-year period. This study investigated the relationship of a collection of isolates classified in four commonly identified lineages in Luxembourg at genome level. The aim was to assess the potential occurrence of stable genomes through the concordance of different

FIGURE 3 | Venn diagram showing the relationship between the loci identified in the four lineages by wgMLST INNUENDO analysis (2,795 targets). A total of 995, 975, 1,000, and 925 targets were detected in lineages A, B, C, and D, respectively.

WGS-based typing schemes exploring and comparing isolates at the core genome level (cgMLSTs from SeqSphere+, Oxford, and INNUENDO) or at the pan genome scale (wgMLST INNUENDO).

Our findings suggested that the genetic population structure of *Campylobacter jejuni* is partly composed of clonal expansion of some genotypes that persist over a long period spanning up to 13 years. Contrary to the epidemic curve commonly detected in case of foodborne outbreaks, stable genetic lineages of this pathogen could emerge to observable frequency through an endemic pattern, causing human infections on a regular basis throughout the country. In order to delineate these lineages with more confidence, efforts were focused on (i) comparing data with a robust design, and (ii) defining cut-offs values aligned with previously published data from genetic variability of the species as well as pre-established threshold for the different typing schemes tested.

Our first concern was to avoid biases generated due to low quality in the raw data and/or assemblies by establishing defined criteria before applying the gene-by-gene approach (Clark et al., 2016; Cody et al., 2017; Llarena et al., 2018; Besser et al., 2019). Quality filtering is a key prerequisite for faithful comparison of genomic data and applied criteria should be clearly stated in all WGS related reports. In their studies, Cody et al. (2013) and Kovanen et al. (2014) implemented a quality threshold in filtering the length of the reads with fixed criteria before the assembly. In 2017, Cody et al. applied a maximum of 150 contigs covering at least 95% of cgMLST targets (Cody et al., 2017), whereas the INNUENDO pipeline included a QC step requiring an assembled depth of coverage of 30x associated with at least 98% of scheme targets found in the cgMLST analyses (Llarena et al., 2018). The Draft Standard of International Standardization Organization (ISO/DIS 23418) suggests a depth of coverage of at least 20x for Illumina short-read raw data and 95% of the read lengths should be over 120 bp (International Organization for Standardization) depending on the application. In our analyses,
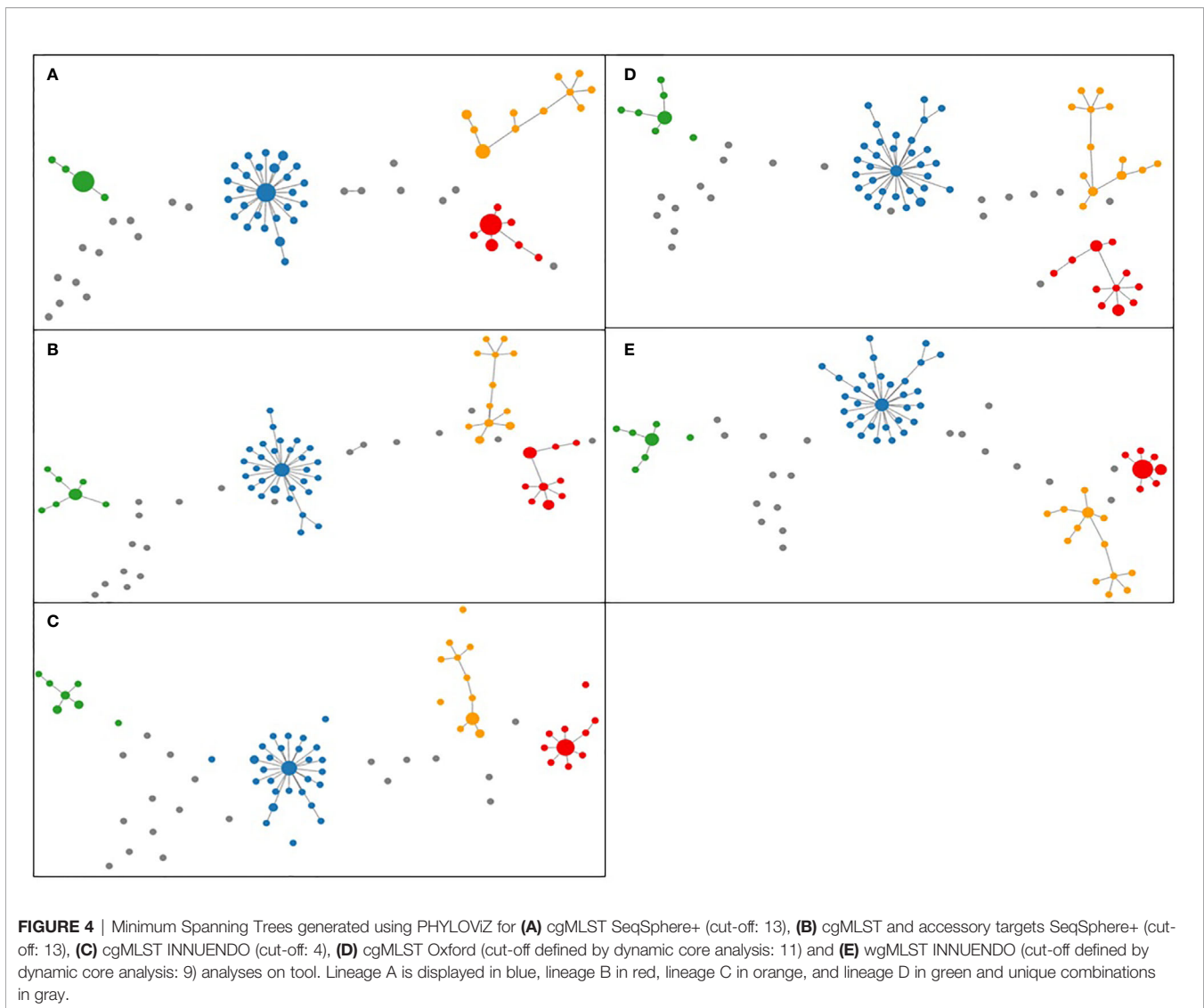
we implemented stringent criteria for quality filtering to ensure robustness and minimizing potential biases related to missing targets generated by poor quality sequencing data.

While a core genome of a bacterial species is expected to consist of a conserved panel of functional genes (also properly called housekeeping genes), mostly present in the genomes of interest and essential to the microorganism, the cgMLST of the three tested methods included a different number of loci. This discrepancy resulted from a more or less stringent definition of the core genome applied to a panel of reference genome varying in size and quality. It is also noteworthy that the cgMLST INNUENDO schema was specifically determined from the *C. jejuni* species while the two others have included some *C. coli* genomes to create their schemes. In summary, SeqSphere+ and INNUENDO selected targets present in at least 90% of the complete genomes (N = 12) or in 99.9% of draft genomes (N = 6,526), respectively (Llarena et al., 2018). The cgMLST Oxford was built from loci occurring in 95% of the *Campylobacter* sp. reference panel (N = 2,472) to take into consideration variation in sequence quality and applied algorithms (Cody et al., 2017). They proposed a more relaxed core genome definition as some isolates may contain mutations, leading to the reduction of the core genome size as more isolates are selected, and that analyses conducted on incomplete draft genomes might constitute a source of missing data (Cody et al., 2017). The finalized cgMLST Oxford scheme represents thus 82% of the reference genome NCTC11168, which places this typing scheme as an intermediate between a core genome and a whole genome MLST scheme with a total of 1,343 loci vs. 637/678 for the two others. A sample-set independent approach was recently proposed to select a conserved-sequence genome as a novel core genome methodology to address this issue (Van Aggelen et al., 2019).

Further, the locus definition is different in the various schemes as well as allele calling algorithms. In the so-called gene-by-gene approach, a locus does not necessarily correspond to the complete coding sequence of a gene but can constitute a specific region. Thus, each schema includes target sequences of varying length ranging from 100 bp to several kb. Surprisingly, the sizes distributions of the targets in the three cgMLSTs tested are very similar with approximately: 21% below 500 bp, 40% ranging from 500 to 1,000 bp, 26% ranging between 1,000 and 1,500 bp, and 13% above 1,500 bp. Except for the SeqSphere+ commercial platform, the design of allele-calling pipelines from the two others WGS-based schemes were published (Jolley et al., 2018; Silva et al., 2018). Both define alleles from sequence assemblies but perform a search by using nucleotide or translated sequences with BLASTN or BLASTP queries and by using "exemplar alleles" as reference or all alleles already recorded in the database. The procedure differs mainly when new sequences display no exact match with known alleles. However, both pipelines validate the nucleotide sequence after translation of DNA codons and include a threshold in percentage sequence identity and length.

A predefined allele distance threshold allows assignment of an unique identifier to genomes displaying a high level of similarities in their cg/wgMLST profiles. The cut-off distance

FIGURE 4 | Minimum Spanning Trees generated using PHYLOViZ for **(A)** cgMLST SeqSphere+ (cut-off: 13), **(B)** cgMLST and accessory targets SeqSphere+ (cut-off: 13), **(C)** cgMLST INNUENDO (cut-off: 4), **(D)** cgMLST Oxford (cut-off defined by dynamic core analysis: 11) and **(E)** wgMLST INNUENDO (cut-off defined by dynamic core analysis: 9) analyses on tool. Lineage A is displayed in blue, lineage B in red, lineage C in orange, and lineage D in green and unique combinations in gray.

value for distinguishing clusters is expressed as a number of ADs and is species or even lineage-specific. To calibrate this value, a test population commonly includes clonal outbreak strains as well as non-epidemiologically linked outgroups. Therefore, the established thresholds are based on strains collected over a relatively short period, and may thus not be appropriate for long-term surveillance. Genomic variations linked to insufficient sequencing quality and microevolutions generated during the gut passage are taken into account for classifying strains (Cody et al., 2013; Revez et al., 2013; Thomas et al., 2014; Barker et al., 2020). For instance, Cody et al. (2013) observed between 3 to 14 loci differences (of 1,643 loci in total) in *Campylobacter* sp., during human gut passage, mainly restricted to insertions and deletions in homopolymeric tracts in contingency loci regulating phase variations of surface structures (Jerome et al., 2011; Barker et al., 2020). To classify related-genomes, Cody et al. (2013) tested two methods: a hierarchical approach based on an increasing number of loci in order to detect closely related isolates and a pairwise comparison based on 1,026 loci shared by the 379 *C. jejuni*

genomes analyzed. Their results lead to the conclusion that the hierarchical approach is better suited to examine isolates epidemiologically related, while pairwise comparisons are preferable for the identification of outbreaks without initial suspicion (Cody et al., 2013). We assessed genomic clusters in our WGS data with goeBURST and we found that defined low cut-off values ranging from 6 to 11 AD and from 5 to 9 AD were appropriated to classify profiles generated with cgMLST Oxford and wgMLST INNUENDO schemes, respectively. By utilizing our newly established thresholds, the classification was consistent with the ones created by cgMLST methods that use a predefined threshold like SeqSphere+ (AD = 13 of 637 targets) and cgMLST INNUENDO (AD = 4 of 678 targets).

Overall, a high concordance in clustering strains was observed between the three cgMLST typing schemes, although congruence is higher between the cgMLSTs Oxford and INNUENDO schemes (predictive of each other in 95% of the cases) compared to the SeqSphere+ scheme. This was not expected, at first glance, as cgMLST schemes from SeqSphere+ and

TABLE 3 | Assignment of genetic profiles according to the different typing schemes for strains initially belonging to the lineage A.

Strain ID	Isolation Source	Year of isolation	ST – *gyrA* – *porA* – Extended MLST	CT – cgMLST SeqSphere+	Genotype – cgMLST Oxford	L1:L2:L3 – cgMLST INNUENDO	Genotype – wgMLST INNUENDO
Camp001	MA	2005	19 – 8 – 7	82	Ox1	1:9:1	wg1
Camp005	H	2006	19 – 8 – 7	82	Ox1	1:9:1	wg1
Camp003	H	2006	19 – 8 – 7	82	Ox1	1:9:1	wg1
Camp004	MA	2006	19 – 8 – 7	82	Ox1	1:9:1	wg1
Camp006	H	2007	19 – 8 – 7	82	Ox1	1:9:1	wg1
Camp002	MA	2006	19 – 8 – 7	82	Ox1	1:9:1	wg1
Camp009	H	2011	19 – 8 – 7	82	Ox1	1:9:1	wg1
Camp010	MA	2011	19 – 8 – 7	82	Ox1	1:9:1	wg1
Camp008	SW	2011	19 – 8 – 7	82	Ox1	1:9:1	wg1
Camp014	H	2012	19 – 8 – 7	82	Ox1	1:9:1	wg1
Camp012	MA	2012	19 – 8 – 7	82	Ox1	1:9:1	wg1
Camp011	SW	2012	19 – 8 – 7	82	Ox1	1:9:1	wg1
Camp015	MA	2012	19 – 8 – 7	82	Ox1	1:9:1	wg1
Camp016	V	2013	19 – 8 – 7	82	Ox1	1:9:1	wg1
Camp018	H	2013	19 – 8 – 7	82	Ox1	1:9:1	wg1
Camp017	H	2013	19 – 8 – 7	82	Ox1	1:9:1	wg1
Camp021	MA	2014	19 – 8 – 7	82	Ox1	1:9:1	wg1
Camp019	H	2014	19 – 8 – 7	82	Ox1	1:9:1	wg1
Camp020	MA	2014	19 – 8 – 7	82	Ox1	1:9:1	wg1
Camp023	MA	2015	19 – 8 – 7	82	Ox1	1:9:1	wg1
Camp029	MA	2016	19 – 8 – 7	82	Ox1	1:9:1	wg1
Camp028	H	2016	19 – 8 – 7	82	Ox1	1:9:1	wg1
Camp027	H	2016	19 – 8 – 7	82	Ox1	1:9:1	wg1
Camp032	V	2016	19 – 8 – 7	82	Ox1	1:9:1	wg1
Camp031	H	2016	19 – 8 – 7	82	Ox1	1:9:1	wg1
Camp035	H	2017	19 – 8 – 7	82	Ox1	1:9:1	wg1
Camp036	MA	2018	19 – 8 – 7	82	Ox1	1:9:1	wg1
Camp037	MA	2018	19 – 8 – 7	82	Ox1	1:9:1	wg1
Camp024	MA	2015	19 – 8 – 582	82	Ox1	1:9:1	wg1
Camp025	H	2015	19 – 8 – 2068	82	Ox1	1:9:1	wg1
Camp022	H	2015	19 – 8 – 2070	82	Ox1	1:9:1	wg1
Camp013	H	2012	19 – 8 – 7	82	Ox1	2695:9:1	wg1
Camp030	V	2016	19 – 8 – 7	82	Ox1	1:9:1	wg5
Camp034	H	2017	19 – 1 – 7	82	Ox11	1:9:1	wg6

In the column Isolation Source, H refers to clinical samples, MA to mammals (cattle and sheep), V to poultry, and SW to surface waters.
ST, Sequence Type; CT, Complex Type; MLST, Multi Locus Sequence Typing; cg, core genome; wg, whole genome.

INNUENDO have a close number of targets (637 vs. 678 targets, respectively) and share 68% of loci. The concordance between the cgMLST schemes Oxford and INNUENDO, both defined from a large collection of strains, suggests a more representative and stably defined core genome. It is noteworthy that in this study, the added value of the number of loci in the cgMLST Oxford cannot be truly attributed on its discriminative power as the datasets contain several clonal population. A largest test population, reflecting the genetic diversity within the *C. jejuni* species, would have been more appropriate for evaluating the resolution of the different typing schemes. As expected, cgMLST profiles could not be mapped with confidence to the wgMLST INNUENDO profiles including a significant larger number of targets. Differences in the accessory genome composition or in the allelic variations could explain these discrepancies. As all the lineages selected for this study originated from various hosts, it could be interesting to further investigate on a possible link between accessory genomes and niche adaption (Woodcock et al., 2017).

The clonality signal appearing through the concordance of the different typing schemes in classifying strains supports the idea of stability of these clones over time and sources. Two independent studies introduced the concept of monomorphic genotypes for *C. jejuni* within the generalist lineages Clonal Complex (CC) ST-21 (Wu et al., 2016) and ST-45 (Llarena et al., 2016). The first study investigated the genetic basis responsible for the hyper virulence of a known clone named "sheep abortion" (clone SA, ST-8), causing foodborne illnesses in human and ruminant abortion (Wu et al., 2016). The second study explored the population structure of the generalist ST-45-CC, overrepresented in human cases in Finland (Llarena et al., 2016). Considering another field, clonal expansion linked to the acquisition of antibiotic resistance has also already been highlighted in *Campylobacter* (Wimalarathna et al., 2013). Observing stable genotypes in *Campylobacter jejuni* over time are in accordance with these results, hypothesizing that predominant clonal evolution is a major adaptive evolutionary strategy in microbial pathogens (Tibayrenc and Ayala, 2017).

In our study, the best example for stable genome over time is lineage A (ST19-*gyrA*8-*porA*7) as its recurrence occurs over more than a decade, although at a low level, representing an average of 13.4% of human cases per year (data not shown).

TABLE 4 | Assignment of genetic profiles according to the different typing schemes for strains initially belonging to the lineage B.

Strain ID	Isolation Source	Year of isolation	ST – *gyrA* – *porA* – Extended MLST	CT – cgMLST SeqSphere+	Genotype – cgMLST Oxford	L1:L2:L3 – cgMLST INNUENDO	Genotype – wgMLST INNUENDO
Camp038	H	2014	2254 – 9 – 1	51	Ox2	19:49:4	wg2
Camp041	V	2014	2254 – 9 – 1	51	Ox2	19:49:4	wg2
Camp040	H	2014	2254 – 9 – 1	51	Ox2	19:49:4	wg2
Camp045	V	2015	2254 – 9 – 1	51	Ox2	19:49:4	wg2
Camp043	H	2015	2254 – 9 – 1	51	Ox2	19:49:4	wg2
Camp046	V	2015	2254 – 9 – 1	51	Ox2	19:49:4	wg2
Camp042	H	2015	2254 – 9 – 1	51	Ox2	19:49:4	wg2
Camp047	H	2015	2254 – 9 – 1	51	Ox2	19:49:4	wg2
Camp044	V	2015	2254 – 9 – 1	51	Ox2	19:49:4	wg2
Camp048	H	2016	2254 – 9 – 1	51	Ox2	19:49:4	wg2
Camp051	H	2017	2254 – 9 – 1	51	Ox2	19:49:4	wg2
Camp052	H	2018	2254 – 9 – 1	51	Ox2	19:49:4	wg2
Camp049	MA	2016	2254 – 9 – 1	51	Ox2	19:49:4	wg7
Camp053	H	2018	2254 – 9 – 1	51	Ox2	19:49:4	wg7
Camp050	H	2017	2254 – 9 – 275	51	Ox2	19:49:4	wg8

In the column Isolation Source, H refers to clinical samples, MA to mammals (cattle and sheep), V to poultry, and SW to surface waters.
ST, Sequence Type; CT, Complex Type; MLST, Multi Locus Sequence Typing; cg, core genome; wg, whole genome.

TABLE 5 | Assignment of genetic profiles according to the different typing schemes for strains initially belonging to the lineage C.

Strain ID	Isolation Source	Year of isolation	ST – *gyrA* – *porA* – Extended MLST	CT – cgMLST SeqSphere+	Genotype – cgMLST Oxford	L1:L2:L3 – cgMLST INNUENDO	Genotype – wgMLST INNUENDO
Camp059	V	2015	464 – 8 – 1678	75	Ox3	29:70:7	wg3
Camp058	H	2015	464 – 8 – 1678	75	Ox3	29:70:7	wg3
Camp067	V	2016	464 – 8 – 1678	75	Ox3	29:70:7	wg3
Camp065	H	2016	464 – 8 – 1678	75	Ox3	29:70:7	wg3
Camp064	V	2016	464 – 8 – 1678	75	Ox3	29:70:7	wg3
Camp069	MA	2017	464 – 8 – 1678	75	Ox3	29:70:7	wg3
Camp063	V	2016	464 – 8 – 1678	75	Ox3	29:70:7	wg3
Camp060	H	2015	464 – 8 – 1678	75	Ox3	29:70:7	wg9
Camp070	H	2017	464 – 8 – 1678	75	Ox3	29:70:7	wg10
Camp055	V	2014	464 – 8 – 1678	75	Ox3	29:70:7	wg11
Camp056	H	2014	464 – 8 – 1678	75	Ox3	29:70:7	wg11
Camp066	V	2016	464 – 8 – 1678	75	Ox3	29:70:7	wg11
Camp068	H	2017	464 – 8 – 1678	75	Ox3	29:70:7	wg11
Camp054	H	2012	464 – 8 – 1678	75	Ox5	29:70:7	wg12
Camp072	H	2018	464 – 8 – 1678	75	Ox6	29:70:7	wg13

In the column Isolation Source, H refers to clinical samples, MA to mammals (cattle and sheep), V to poultry, and SW to surface waters.
ST, Sequence Type; CT, Complex Type; MLST, Multi Locus Sequence Typing; cg, core genome; wg, whole genome.

TABLE 6 | Assignment of genetic profiles according to the different typing schemes for strains initially belonging to the lineage D.

Strain ID	Isolation Source	Year of isolation	ST – *gyrA* – *porA* – Extended MLST	CT – cgMLST SeqSphere+	Genotype – cgMLST Oxford	L1:L2:L3 – cgMLST INNUENDO	Genotype – wgMLST INNUENDO
Camp082	V	2017	6175 – 9 – 1625	543	Ox4	41:68:27	wg4
Camp083	H	2017	6175 – 9 – 1625	543	Ox4	41:68:27	wg4
Camp081	H	2017	6175 – 9 – 1625	543	Ox4	41:68:27	wg4
Camp084	MA	2017	6175 – 9 – 1625	543	Ox4	41:68:27	wg4
Camp080	H	2017	6175 – 9 – 1625	543	Ox4	41:68:27	wg4
Camp085	H	2018	6175 – 9 – 1625	543	Ox4	41:68:27	wg4
Camp086	V	2018	6175 – 9 – 1625	543	Ox4	41:68:27	wg4
Camp087	H	2018	6175 – 9 – 1625	543	Ox4	41:68:27	wg4
Camp089	H	2018	6175 – 9 – 1625	543	Ox4	41:68:27	wg4
Camp088	H	2018	6175 – 9 – 1625	543	Ox7	2724:68:27	wg14

In the column Isolation Source, H refers to clinical samples, MA to mammals (cattle and sheep), V to poultry and SW to surface waters.
ST, Sequence Type; CT, Complex Type; MLST, Multi Locus Sequence Typing; cg, core genome; wg, whole genome.

TABLE 7 | Adjusted Wallace coefficients values (CI 95%) for typing schemes comparison.

	Extended MLST	cgMLST SeqSphere+	cgMLST INNUENDO	cgMLST Oxford	wgMLST INNUENDO
Extended MLST		1.000	0.931	0.935	0.757
		(1.000–1.000)	(0.832–1.000)	(0.876–0.994)	(0.647–0.867)
cgMLST SeqSphere+	0.795		0.937	0.945	0.728
	(0.637–0.953)		(0.843–1.000)	(0.895–0.994)	(0.596–0.859)
cgMLST INNUENDO	0.790	1.000		0.956	0.729
	(0.630–0.951)	(1.000–1.000)		(0.909–1.000)	(0.597–0.862)
cgMLST Oxford	0.787	1.000	0.948		0.769
	(0.622–0.952)	(1.000–1.000)	(0.852–1.000)		(0.634–0.903)
wgMLST INNUENDO	0.829	1.000	0.940	0.998	
	(0.658–0.997)	(1.000–1.000)	(0.828–1.000)	(0.996–1.000)	

Thirty-two strains of 34 from diverse sources (human, cattle and sheep, poultry, and environmental samples) were gathered in the same genetic profile at the whole genome level. This result reflects that this lineage is likely derived from one common ancestor, which thereafter disseminated broadly to a variety of mammals and birds, clearly demonstrating an ability to disperse in the environment and adapt to different ecological niches. Thus, the question of the environmental transmission routes arises, particularly concerning animal reservoirs such as poultry and ruminants that could contribute to water contamination (Mughini-Gras et al., 2016). Persistent strains have already been identified, mainly in poultry farms and in milk, and it would be interesting to link lineage A with other contamination sources such as insects, rodents, drinking water, or the surrounding environment (Kudirkienė et al., 2010; Perez-Boto et al., 2012; Rauber-Würfel et al., 2019; Jaakkonen et al., 2020).

The lineage B (ST2254-*gyrA9-porA*1) arose unexpectedly from our national surveillance with an epidemic curve between March and April 2014 (>70 campylobacteriosis cases). Interestingly, after this episode, clinical isolates of this lineage were still collected but at a much lower frequency during the following four years. To put things into context, this particular ST was singular in 2014 and by querying the pubmlst.org database (Jolley et al., 2018); only a dozen strains had been recorded at that time including two from poultry origin. Interestingly, the same "clone" was finally isolated in the framework of the official controls conducted by the state veterinary laboratory in Luxembourg and supported chicken as a possible source of this outbreak. In molecular epidemiology, the expression "clone" generally refers to a set of independently isolated microbial organisms that have similar genotypic traits as a results of a shared common ancestor (Van Belkum et al., 2007). The analysis using different typing schemes gathered 80% of the tested strains from lineage B in the same CT, whereas only 50% of isolates from lineage C formed a cluster. These data support the occurrence of the most large-scale outbreak caused by *C. jejuni* ever identified in Luxembourg and linked to chicken imported from neighboring countries, as the local production is negligible. Two years after the epidemic episode, this clone was isolated from a bovine source for the first time, while the remainder of lineage B was mainly isolated from poultry. The extent of ecological niches suggests that strains from lineage B were able to cross ecological barriers and disseminate in the environment with a generalist profile (Sheppard et al., 2014).

Lineage C (ST464-*gyrA8-porA*1678) displayed two micro-epidemic peaks: one in March 2016 and a second in January and February 2017. Since then, its incidence has been low with less than 10 human cases per year since March 2017 and we observed a first sample of bovine source isolated in August 2017. An average of two human cases per month from December 2016 to May 2018 indicates the profile of an emerging clone tending to have an endemic profile. Notably, this lineage displays the *gyrA* allele 8, one of the nucleotide allele in *C. jejuni* containing the C257T mutation (i.e., the peptide shift Thr86Ile) which confers quinolone resistance (Ragimbeau et al., 2014). Indeed, dispersion of antimicrobial resistant lineages due to positive selection was previously described for bacterial pathogens, such as uropathogenic *Escherichia coli* (ST 131 for example) (Totsika et al., 2011; Yamaji et al., 2018) and *C. jejuni* (ST 464 for instance) (Cha et al., 2016). For lineage D (ST6175-*gyrA9-porA*1625), the first isolate was identified in 2012 from a human infection, then in 2014 and at the beginning of 2016. A link with poultry source was observed.

Whatever the typing scheme used, clear signals appeared in our molecular surveillance for identifying an outbreak (lineage B in 2014) or the phenomenon of recurrent clones, which cause of more than 50% of human infections in Luxembourg. This study provides new insights for the genomic surveillance of *Campylobacter* infections. Through the exploration of the large collection of data that we have initiated 15 years ago, we seek to demonstrate the strong interest in monitoring genotypes causing gastroenteritis in the sense that campylobacteriosis is not only of sporadic nature. A recent study based on collected WGS data in Denmark also supported these findings (Joensen et al., 2018).

Molecular surveillance of foodborne pathogens is currently implemented for *Salmonella* (Dangel et al., 2019), *Listeria* (Van Walle et al., 2018), and VTEC (Joensen et al., 2014) at the EU level (ECDC, 2019) and in the USA (Ribot et al., 2019). For *C. jejuni*, such monitoring in routine is hindered by the absence of a validated scheme at international level and the lack of evidence for the spread possibility of cross-border genotypes. The presence of recurring genotypes highlights the possible long-term existing of stable clones representing a risk factor of geographic spread that needs to be investigated further. Like for the acquisition of antibiotic resistance, persistent strains may have acquired specific phenotypic traits to adapt to other hosts or disperse in the environment. Habituation to ambient air (Rodrigues et al., 2015; Rodrigues et al., 2016), adhesion to inert surface (Sulaeman et al., 2010; Oh et al., 2016) and

biofilm formation (Reuter et al., 2010; Turonova et al., 2015) could contribute to the survival strategies of *C. jejuni* in the environment. In the future, studying the phenotypic traits of recurrent clones and their relationship to spatiotemporal persistence would broaden our understanding on *Campylobacter* adaptation and its transmission to humans.

DATA AVAILABILITY STATEMENT

Sequenced raw reads have been uploaded to ENA and are available under the accession project number PRJEB40465.

AUTHOR CONTRIBUTIONS

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcimb.2020.608020/full#supplementary-material

REFERENCES

Barker, C. R., Painset, A., Swift, C., Jenkins, C., Godbole, G., Maiden, M. C. J., et al. (2020). Microevolution of Campylobacter jejuni during long-term infection in an immunocompromised host. *Sci. Rep.* 10, 10109. doi: 10.1038/s41598-020-66771-7

Berthe, F., Hugas, M., and Makela, P. (2013). Integrating surveillance of animal health, food pathogens and foodborne disease in the European Union. *Rev. Off. Int. Epizoot.* 32, 521–528. doi: 10.20506/rst.32.2.2243

Besser, J. M., Carleton, H. A., Trees, E., Stroika, S. G., Hise, K., Wise, M., et al. (2019). Interpretation of whole-genome sequencing for enteric disease surveillance and outbreak investigation. *Foodborne Pathog. Dis.* 16, 504–512. doi: 10.1089/fpd.2019.2650

Bronowski, C., James, C. E., and Winstanley, C. (2014). Role of environmental survival in transmission of Campylobacter jejuni. *FEMS Microbiol. Lett.* 356, 8–19. doi: 10.1111/1574-6968.12488

Campylobacter MOMP database. Available at: https://pubmlst.org/campylobacter/info/porA_method.shtml (Accessed April 15, 2020).

Cassini, A., Colzani, E., Pini, A., Mangen, M.-J. J., Plass, D., McDonald, S. A., et al. (2018). Impact of infectious diseases on population health using incidence-based disability-adjusted life years (DALYs): results from the burden of communicable diseases in Europe study, European Union and European Economic Area countrieto 2013. *Euro. Surveill* 23, 17-00454. doi: 10.2807/1560-7917.ES.2018.23.16.17-00454

Cha, W., Mosci, R., Wengert, S. L., Singh, P., Newton, D. W., Salimnia, H., et al. (2016). Antimicrobial susceptibility profiles of human Campylobacter jejuni isolates and association with phylogenetic lineages. *Front. Microbiol.* 7, 589. doi: 10.3389/fmicb.2016.00589

Clark, C. G., Beeston, A., Bryden, L., Wang, G., Barton, C., Cuff, W., et al. (2007). Phylogenetic relationships of Campylobacter jejuni based on porA sequences. *Can. J. Microbiol.* 53, 27–38. doi: 10.1139/w06-099

Clark, C. G., Berry, C., Walker, M., Petkau, A., Barker, D. O. R., Guan, C., et al. (2016). Genomic insights from whole genome sequencing of four clonal outbreak Campylobacter jejuni assessed within the global C. jejuni population. *BMC Genomics* 17, 990. doi: 10.1186/s12864-016-3340-8

Cody, A. J., Maiden, M. J. C., and Dingle, K. E. (2009). Genetic diversity and stability of the porA allele as a genetic marker in human Campylobacter infection. *Microbiology* 155, 4145–4154. doi: 10.1099/mic.0.031047-0

Cody, A. J., McCarthy, N. D., Jansen van Rensburg, M., Isinkaye, T., Bentley, S. D., Parkhill, J., et al. (2013). Real-time genomic epidemiological evaluation of human Campylobacter isolates by use of whole-genome multilocus sequence typing. *J. Clin. Microbiol.* 51, 2526–2534. doi: 10.1128/JCM.00066-13

Cody, A. J., Bray, J. E., Jolley, K. A., McCarthy, N. D., and Maiden, M. C. J. (2017). Core genome multilocus sequence typing scheme for stable, comparative analyses of Campylobacter jejuni and C. coli human disease isolates. *J. Clin. Microbiol.* 55, 2086–2097. doi: 10.1128/JCM.00080-17

Conway, J. R., Lex, A., and Gehlenborg, N. (2017). UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* 33, 2938–2940. doi: 10.1093/bioinformatics/btx364

Csárdi, G., and Nepusz, T. (2006). The igraph software package for complex network research. *InterJ. Complex Syst* 1965, 1–9. Available at: https://www.semanticscholar.org/paper/The-igraph-software-package-for-complex-network-Cs%C3%A1rdi-Nepusz/1d2744b83519657f5f2610698a8ddd177ced4f5c [Accessed September 17, 2020].

Dangel, A., Berger, A., Messelhäußer, U., Konrad, R., Hörmansdorfer, S., Ackermann, N., et al. (2019). Genetic diversity and delineation of Salmonella Agona outbreak strains by next generation sequencing, Bavaria, Germano 2018. *Euro. Surveill.* 24, 1800303. doi: 10.2807/1560-7917.ES.2019.24.18.1800303

Davis, K. R., Dunn, A. C., Burnett, C., McCullough, L., Dimond, M., Wagner, J., et al. (2016). Campylobacter jejuni infections associated with raw milk consumption - Uta. *MMWR Morb. Mortal. Wkly. Rep.* 65, 301–305. doi: 10.15585/mmwr.mm6512a1

de Boer, P., Wagenaar, J. A., Achterberg, R. P., van Putten, J. P. M., Schouls, L. M., and Duim, B. (2002). Generation of Campylobacter jejuni genetic diversity in vivo. *Mol. Microbiol.* 44, 351–359. doi: 10.1046/j.1365-2958.2002.02930.x

Devleesschauwer, B., Bouwknegt, M., Mangen, M.-J. J., and Havelaar, A. H. (2017). "Chapter 2 - Health and economic burden of Campylobacter," in

Campylobacter. Ed. G. Klein (London, UK: Academic Press), 27–40. doi: 10.1016/B978-0-12-803623-5.00002-2

Dingle, K. E., Colles, F. M., Wareing, D. R. A., Ure, R., Fox, A. J., Bolton, F. E., et al. (2001). Multilocus sequence typing system for Campylobacter jejuni. *J. Clin. Microbiol.* 39, 14–23. doi: 10.1128/JCM.39.1.14-23.2001

Dingle, K. E., McCarthy, N. D., Cody, A. J., Peto, T. E. A., and Maiden, M. C. J. (2008). Extended sequence typing of Campylobacter spp., United Kingdom. *Emerg. Infect. Dis.* 14, 1620–1622. doi: 10.3201/eid1410.071109

ECDC (2016).Expert opinion on whole genome sequencing for public health surveillance. In: *European Centre for Disease Prevention and Control.* Available at: http://ecdc.europa.eu/en/publications-data/expert-opinion-whole-genome-sequencing-public-health-surveillance (Accessed August 9, 2018).

ECDC (2019).ECDC strategic framework for the integration of molecular and genomic typing into European surveillance and multi-country outbreak investigations. In: *European Centre for Disease Prevention and Control.* Available at: https://www.ecdc.europa.eu/en/publications-data/ecdc-strategic-framework-integration-molecular-and-genomic-typing-european (Accessed November 15, 2019).

EFSA and ECDC (2018). The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2017. *EFSA J.* 16, 5500. doi: 10.2903/j.efsa.2018.5500

EFSA and ECDC (2019). The European Union One Health 2018 zoonoses report. *EFSA J.* 17, e05926. doi: 10.2903/j.efsa.2019.5926

Ellis-Iversen, J., Ridley, A., Morris, V., Sowa, A., Harris, J., Atterbury, R., et al. (2012). Persistent environmental reservoirs on farms as risk factors for Campylobacter in commercial poultry. *Epidemiol. Infect.* 140, 916–924. doi: 10.1017/S095026881100118X

Feil, E. J., Li, B. C., Aanensen, D. M., Hanage, W. P., and Spratt, B. G. (2004). eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *J. Bacteriol.* 186, 1518–1530. doi: 10.1128/JB.186.5.1518-1530.2004

Francisco, A. P., Bugalho, M., Ramirez, M., and Carriço, J. A. (2009). Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. *BMC Bioinf.* 10, 152. doi: 10.1186/1471-2105-10-152

Francisco, A. P., Vaz, C., Monteiro, P. T., Melo-Cristino, J., Ramirez, M., and Carriço, J. A. (2012). PHYLOVIZ: phylogenetic inference and data visualization for sequence based typing methods. *BMC Bioinf.* 13, 87. doi: 10.1186/1471-2105-13-87

Gorman,, and Adley, (2004). An evaluation of five preservation techniques and conventional freezing temperatures of –20°C and –85°C for long-term preservation of Campylobacter jejuni. *Lett. Appl. Microbiol.* 38, 306–310. doi: 10.1111/j.1472-765X.2004.01490.x

Hyllestad, S., Iversen, A., MacDonald, E., Amato, E., Borge, B.Å.S., Bøe, A., et al. (2020). Large waterborne Campylobacter outbreak: use of multiple approaches to investigate contamination of the drinking water supply system, Norway, June 2019. *Eurosurveillance* 25, 2000011. doi: 10.2807/1560-7917.ES.2020.25.35.2000011

International Organization for Standardization Draft standard of International Standardization Organization. Available at: https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/07/55/75509.html (Accessed April 27, 2020). ISO/CD 23418. ISO.

Jaakkonen, A., Kivistö, R., Aarnio, M., Kalekivi, J., and Hakkinen, M. (2020). Persistent contamination of raw milk by Campylobacter jejuni ST-883. *PloS One* 15, e0231810. doi: 10.1371/journal.pone.0231810

Jajou, R., Kohl, T. A., Walker, T., Norman, A., Cirillo, D. M., Tagliani, E., et al. (2019). Towards standardisation: comparison of five whole genome sequencing (WGS) analysis pipelines for detection of epidemiologically linked tuberculosis cases. *Euro. Surveill.* 24, 1900130. doi: 10.2807/1560-7917.ES.2019.24.50.1900130

Jakopanec, I., Borgen, K., Vold, L., Lund, H., Forseth, T., Hannula, R., et al. (2008). A large waterborne outbreak of campylobacteriosis in Norway: the need to focus on distribution system safety. *BMC Infect. Dis.* 8, 128. doi: 10.1186/1471-2334-8-128

Jay-Russell, M. T., Mandrell, R. E., Yuan, J., Bates, A., Manalac, R., Mohle-Boetani, J., et al. (2013). Using major outer membrane protein typing as an epidemiological tool to investigate outbreaks caused by milk-borne Campylobacter jejuni isolates in California. *J. Clin. Microbiol.* 51, 195–201. doi: 10.1128/JCM.01845-12

Jerome, J. P., Bell, J. A., Plovanich-Jones, A. E., Barrick, J. E., Brown, C. T., and Mansfield, L. S. (2011). Standing genetic variation in contingency loci drives

the rapid adaptation of Campylobacter jejuni to a novel host. *PloS One* 6, e16399. doi: 10.1371/journal.pone.0016399

Jesse, T. W., Englen, M. D., Pittenger-Alley, L. G., and Fedorka-Cray, P. J. (2006). Two distinct mutations in gyrA lead to ciprofloxacin and nalidixic acid resistance in Campylobacter coli and Campylobacter jejuni isolated from chickens and beef cattle*. *J. Appl. Microbiol.* 100, 682–688. doi: 10.1111/j.1365-2672.2005.02796.x

Joensen, K. G., Scheutz, F., Lund, O., Hasman, H., Kaas, R. S., Nielsen, E. M., et al. (2014). Real-Time Whole-Genome Sequencing for Routine Typing, Surveillance, and Outbreak Detection of Verotoxigenic Escherichia coli. *J. Clin. Microbiol.* 52, 1501–1510. doi: 10.1128/JCM.03617-13

Joensen, K. G., Kuhn, K. G., Müller, L., Björkman, J. T., Torpdahl, M., Engberg, J., et al. (2018). Whole-genome sequencing of Campylobacter jejuni isolated from Danish routine human stool samples reveals surprising degree of clustering. *Clin. Microbiol. Infect.* 24, 201.e5–201.e8. doi: 10.1016/j.cmi.2017.07.026

Jolley, K. A., Bray, J. E., and Maiden, M. C. J. (2017). A RESTful application programming interface for the PubMLST molecular typing and genome databases. *Database (Oxford)* 2017, bax060. doi: 10.1093/database/bax060

Jolley, K. A., Bray, J. E., and Maiden, M. C. J. (2018). Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res.* 3, 124. doi: 10.12688/wellcomeopenres.14826.1

Jünemann, S., Sedlazeck, F. J., Prior, K., Albersmeier, A., John, U., Kalinowski, J., et al. (2013). Updating benchtop sequencing performance comparison. *Nat. Biotechnol.* 31, 294–296. doi: 10.1038/nbt.2522

Kang, C. R., Bang, J. H., and Cho, S.II (2019). Campylobacter jejuni foodborne infection associated with cross-contamination: outbreak in Seoul in 2017. *Infect. Chemother.* 51, 21–27. doi: 10.3947/ic.2019.51.1.21

Kovanen, S. M., Kivistö, R. L., Rossi, M., Schott, T., Kärkkäinen, U.-M., Tuuminen, T., et al. (2014). Multilocus Sequence Typing (MLST) and Whole-Genome MLST of Campylobacter jejuni isolates from human infections in three districts during a seasonal peak in Finland. *J. Clin. Microbiol.* 52, 4147–4154. doi: 10.1128/JCM.01959-14

Kudirkienė, E., Malakauskas, M., Malakauskas, A., Bojesen, A. M., and Olsen, J. E. (2010). Demonstration of persistent strains of Campylobacter jejuni within broiler farms over a 1-year period in Lithuania. *J. Appl. Microbiol.* 108, 868–877. doi: 10.1111/j.1365-2672.2009.04490.x

Kuhn, K. G., Nielsen, E. M., Mølbak, K., and Ethelberg, S. (2018). Determinants of sporadic Campylobacter infections in Denmark: a nationwide case-control study among children and young adults. *Clin. Epidemiol.* 10, 1695–1707. doi: 10.2147/CLEP.S177141

Lahti, E., Löfdahl, M., Ågren, J., Hansson, I., and Olsson Engvall, E. (2017). Confirmation of a campylobacteriosis outbreak associated with chicken liver pâté using PFGE and WGS. *Zoonoses Public Health* 64, 14–20. doi: 10.1111/zph.12272

Lan, R., Reeves, P. R., and Octavia, S. (2009). Population structure, origins and evolution of major Salmonella enterica clones. *Infect. Genet. Evol.* 9, 996–1005. doi: 10.1016/j.meegid.2009.04.011

Llarena, A.-K., Zhang, J., Vehkala, M., Välimäki, N., Hakkinen, M., Hänninen, M.-L., et al. (2016). Monomorphic genotypes within a generalist lineage of Campylobacter jejuni show signs of global dispersion. *Microb. Genom.* 2, e000088. doi: 10.1099/mgen.0.000088

Llarena, A.-K., Ribeiro-Gonçalves, B. F., Silva, D. N., Halkilahti, J., Machado, M. P., Silva, M. S. D., et al. (2018). INNUENDO: a cross-sectoral platform for the integration of genomics in the surveillance of food-borne pathogens. *EFSA Support. Public.* 15, 1498E. doi: 10.2903/sp.efsa.2018.EN-1498

Machado, M. P., Halkilahti, J., Jaakkonen, A., Silva, D. N., Mendes, I., Nalbantoglu, Y., et al. (2017) *INNUca GitHub.* Available at: https://github.com/B-UMMI/INNUca

Maiden, M. C., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., et al. (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. U. S. A.* 95, 3140–3145. doi: 10.1073/pnas.95.6.3140

Mellmann, A., Mosters, J., Bartelt, E., Roggentin, P., Ammon, A., Friedrich, A. W, et al. (2004). Sequence-based typing of flaB is a more stable screening tool than typing of flaA for monitoring of Campylobacter populations. *J. Clin. Microbiol.* 42, 4840–4842. doi: 10.1128/JCM.42.10.4840-4842.2004

Mirdita, M., Steinegger, M., and Söding, J. (2019). MMseqs2 desktop and local web server app for fast, interactive sequence searches. *Bioinformatics* 35, 2856–2858. doi: 10.1093/bioinformatics/bty1057

Mossong, J., Mughini-Gras, L., Penny, C., Devaux, A., Olinger, C., Losch, S., et al. (2016). Human campylobacteriosis in Luxembour–2013: a case-control study combined with multilocus sequence typing for source attribution and risk factor analysis. *Sci. Rep.* 6, 20939. doi: 10.1038/srep20939

Mughini-Gras, L., Smid, J. H., Wagenaar, J. A., de Boer, A. G., Havelaar, A. H., Friesema, I. H. M., et al. (2012). Risk factors for campylobacteriosis of chicken, ruminant, and environmental origin: a combined case-control and source attribution analysis. *PloS One* 7, e42599. doi: 10.1371/journal.pone.0042599

Mughini-Gras, L., Penny, C., Ragimbeau, C., Schets, F. M., Blaak, H., Duim, B., et al. (2016). Quantifying potential sources of surface water contamination with Campylobacter jejuni and Campylobacter coli. *Water Res.* 101, 36–45. doi: 10.1016/j.watres.2016.05.069

Nascimento, M., Sousa, A., Ramirez, M., Francisco, A. P., Carriço, J. A., and Vaz, C. (2017). PHYLOViZ 2.0: providing scalable data integration and visualization for multiple phylogenetic inference methods. *Bioinformatics* 33, 128–129. doi: 10.1093/bioinformatics/btw582

Oh, E., Kim, J.-C., and Jeon, B. (2016). Stimulation of biofilm formation by oxidative stress in Campylobacter jejuni under aerobic conditions. *Virulence* 7, 846–851. doi: 10.1080/21505594.2016.1197471

Payot, S., Bolla, J.-M., Corcoran, D., Fanning, S., Mégraud, F., and Zhang, Q. (2006). Mechanisms of fluoroquinolone and macrolide resistance in Campylobacter spp. *Microbes Infect.* 8, 1967–1971. doi: 10.1016/j.micinf.2005.12.032

Perez-Boto, D., Garcia-Peña, F. J., Abad-Moreno, J. C., and Echeita, M. A. (2012). Dynamics of populations of Campylobacter jejuni in two grandparent broiler breeder farms: persistent vs. transient strains. *Vet. Microbiol.* 159, 204–211. doi: 10.1016/j.vetmic.2012.03.042

PHYLOViZ. Available at: http://online2.phyloviz.net/index (Accessed September 14, 2020).

Pinto, F. R., Melo-Cristino, J., and Ramirez, M. (2008). A confidence interval for the Wallace coefficient of concordance and its application to microbial typing methods. *PloS One* 3, e3696. doi: 10.1371/journal.pone.0003696

PubMLST *Campylobacter Sequence Typing.* Available at: https://pubmlst.org/campylobacter/ (Accessed September 18, 2020).

R Core Team (2018). *R: A language and environment for statistical computing* (Vienna, Austria: R Foundation for Statistical Computing). Available at: https://www.r-project.org/ (Accessed June 30, 2020).

Ragimbeau, C., Schneider, F., Losch, S., Even, J., and Mossong, J. (2008). Multilocus sequence typing, pulsed-field gel electrophoresis, and fla short variable region typing of clonal complexes of Campylobacter jejuni strains of human, bovine, and poultry origins in Luxembourg. *Appl. Environ. Microbiol.* 74, 7715–7722. doi: 10.1128/AEM.00865-08

Ragimbeau, C., Colin, S., Devaux, A., Decruyenaere, F., Cauchie, H.-M., Losch, S., et al. (2014). Investigating the host specificity of Campylobacter jejuni and Campylobacter coli by sequencing gyrase subunit A. *BMC Microbiol.* 14, 205. doi: 10.1186/s12866-014-0205-7

Ragon, M., Wirth, T., Hollandt, F., Lavenir, R., Lecuit, M., Le Monnier, A., et al. (2008). A new perspective on Listeria monocytogenes evolution. *PloS Pathog.* 4, e1000146. doi: 10.1371/journal.ppat.1000146

Rauber-Würfel, S., de, F., Voss-Rech, D., dos Santos Pozza, J., Coldebella, A., Santiago Silva, V., et al. (2019). Population dynamics of thermotolerant Campylobacter in broilers reared on reused litter. *Foodborne Pathog. Dis.* 16, 738–743. doi: 10.1089/fpd.2019.2645

Reuter, M., Mallett, A., Pearson, B. M., and van Vliet, A. H. M. (2010). Biofilm formation by Campylobacter jejuni is increased under aerobic conditions. *Appl. Environ. Microbiol.* 76, 2122–2128. doi: 10.1128/AEM.01878-09

Revez, J., Schott, T., Llarena, A.-K., Rossi, M., and Hänninen, M.-L. (2013). Genetic heterogeneity of Campylobacter jejuni NCTC 11168 upon human infection. *Infect. Genet. Evol.* 16, 305–309. doi: 10.1016/j.meegid.2013.03.009

Revez, J., Llarena, A.-K., Schott, T., Kuusi, M., Hakkinen, M., Kivistö, R., et al. (2014). Genome analysis of Campylobacter jejuni strains isolated from a waterborne outbreak. *BMC Genomics* 15, 768. doi: 10.1186/1471-2164-15-768

Ribot, E. M., Freeman, M., Hise, K. B., and Gerner-Smidt, P. (2019). PulseNet: entering the age of next-generation sequencing. *Foodborne Pathog. Dis.* 16, 451–456. doi: 10.1089/fpd.2019.2634

Ridom SeqSphere+ (2013). *Ridom SeqSphere+ Documentation - Task Procedure.* Available at: https://www.ridom.de/seqsphere/ug/v60/Task_Procedure.html#Cluster-Alert_Settings (Accessed April 14, 2020).

Rodrigues, R. C., Pocheron, A.-L., Hernould, M., Haddad, N., Tresse, O., and Cappelier, J.-M. (2015). Description of Campylobacter jejuni Bf, an atypical aero-tolerant strain. *Gut Pathog.* 7, 30. doi: 10.1186/s13099-015-0077-x

Rodrigues, R. C., Haddad, N., Chevret, D., Cappelier, J.-M., and Tresse, O. (2016). Comparison of proteomics profiles of Campylobacter jejuni strain Bf under microaerobic and aerobic conditions. *Front. Microbiol.* 7, 1596. doi: 10.3389/fmicb.2016.01596

Rossi, M., Santos Da Silva, M., Ribeiro-Gonçalves, B. F., Silva, D. N., Machado, Oleastro, M., et al. (2018). INNUENDO whole genome and core genome MLST schemas and datasets for Campylobacter jejuni. (Version 1.0) (Data set). *Zenodo.* doi: 10.5281/zenodo.1322564

Sahin, O., Fitzgerald, C., Stroika, S., Zhao, S., Sippy, R. J., Kwan, P., et al. (2012). Molecular evidence for zoonotic transmission of an emergent, highly pathogenic Campylobacter jejuni clone in the United States. *J. Clin. Microbiol.* 50, 680–687. doi: 10.1128/JCM.06167-11

Scharff, R. L. (2012). Economic burden from health losses due to foodborne illness in the United States. *J. Food Prot.* 75, 123–131. doi: 10.4315/0362-028X.JFP-11-058

Severiano, A., Pinto, F. R., Ramirez, M., and Carriço, J. A. (2011). Adjusted Wallace coefficient as a measure of congruence between typing methods. *J. Clin. Microbiol.* 49, 3997–4000. doi: 10.1128/JCM.00624-11

Sheppard, S. K., McCarthy, N. D., Jolley, K. A., and Maiden, M. C. J. (2011). Introgression in the genus Campylobacter: generation and spread of mosaic alleles. *Microbiology* 157, 1066–1074. doi: 10.1099/mic.0.045153-0

Sheppard, S. K., Jolley, K. A., and Maiden, M. C. J. (2012). A gene-by-gene approach to bacterial population genomics: whole genome MLST of Campylobacter. *Genes (Basel)* 3, 261–277. doi: 10.3390/genes3020261

Sheppard, S. K., Cheng, L., Méric, G., de Haan, C. P. A., Llarena, A.-K., Marttinen, P., et al. (2014). Cryptic ecology among host generalist Campylobacter jejuni in domestic animals. *Mol. Ecol.* 23, 2442–2451. doi: 10.1111/mec.12742

Silva, M., Machado, M. P., Silva, D. N., Rossi, M., Moran-Gilad, J., Santos, S., et al. (2018). chewBBACA: a complete suite for gene-by-gene schema creation and strain identification. *Microb. Genom.* 4, e000166. doi: 10.1099/mgen.0.000166

Stuart, T. L., Sandhu, J., Stirling, R., Corder, J., Ellis, A., Misa, P., et al. (2010). Campylobacteriosis outbreak associated with ingestion of mud during a mountain bike race. *Epidemiol. Infect.* 138, 1695–1703. doi: 10.1017/S095026881000049X

Sulaeman, S., Bihan, G. L., Rossero, A., Federighi, M., Dé, E., and Tresse, O. (2010). Comparison between the biofilm initiation of Campylobacter jejuni and Campylobacter coli strains to an inert surface using BioFilm Ring Test®. *J. Appl. Microbiol.* 108, 1303–1312. doi: 10.1111/j.1365-2672.2009.04534.x

Thomas, D. K., Lone, A. G., Selinger, L. B., Taboada, E. N., Uwiera, R. R. E., Abbott, D. W., et al. (2014). Comparative variation within the genome of Campylobacter jejuni NCTC 11168 in human and murine hosts. *PloS One* 9, e88229. doi: 10.1371/journal.pone.0088229

Tibayrenc, M., and Ayala, F. J. (2017). "Chapter Six - Is predominant clonal evolution a common evolutionary adaptation to parasitism in pathogenic parasitic protozoa, fungi, bacteria, and viruses?," in *Advances in Parasitology.* Eds. D. Rollinson and J. R. Stothard (San Diego, United States: Academic Press), 243–325. doi: 10.1016/bs.apar.2016.08.007

Totsika, M., Beatson, S. A., Sarkar, S., Phan, M.-D., Petty, N. K., Bachmann, N., et al. (2011). Insights into a multidrug resistant Escherichia coli pathogen of the globally disseminated ST131 lineage: genome analysis and virulence mechanisms. *PloS One* 6, e26578. doi: 10.1371/journal.pone.0026578

Turonova, H., Briandet, R., Rodrigues, R., Hernould, M., Hayek, N., Stintzi, A., et al. (2015). Biofilm spatial organization by the emerging pathogen Campylobacter jejuni: comparison between NCTC 11168 and 81-176 strains under microaerobic and oxygen-enriched conditions. *Front. Microbiol.* 6, 709. doi: 10.3389/fmicb.2015.00709

Van Aggelen, H., Kolde, R., Chamarthi, H., Loving, J., Fan, Y., Fallon, J. T., et al. (2019). A core genome approach that enables prospective and dynamic monitoring of infectious outbreaks. *Sci. Rep.* 9, 7808. doi: 10.1038/s41598-019-44189-0

Van Belkum, A. V., Tassios, P. T., Dijkshoorn, L., Haeggman, S., Cookson, B., Fry, N. K., et al. (2007). Guidelines for the validation and application of typing methods for use in bacterial epidemiology. *Clin. Microbiol. Infect.* 13, 1–46. doi: 10.1111/j.1469-0691.2007.01786.x

Van Walle, I., Björkman, J. T., Cormican, M., Dallman, T., Mossong, J., Moura, A., et al. (2018). Retrospective validation of whole genome sequencing-enhanced

surveillance of listeriosis in Europto 2015. *Euro. Surveill* 23, 1700798. doi: 10.2807/1560-7917.ES.2018.23.33.1700798

Wallace, D. L. (1983). A method for comparing two hierarchical clusterings. *J. Am. Stat. Assoc.* 78, 569–576. doi: 10.2307/2288118

Wang, Y., Huang, W. M., and Taylor, D. E. (1993). Cloning and nucleotide sequence of the Campylobacter jejuni gyrA gene and characterization of quinolone resistance mutations. *Antimicrob. Agents Chemother.* 37, 457–463. doi: 10.1128/AAC.37.3.457

WHO (2013). The global view of campylobacteriosis. *WHO*. Available at: https://www.who.int/foodsafety/publications/campylobacteriosis/en/ (Accessed November 11, 2019).

Wimalarathna, H. M., Richardson, J. F., Lawson, A. J., Elson, R., Meldrum, R., Little, C. L., et al. (2013). Widespread acquisition of antimicrobial resistance among Campylobacter isolates from UK retail poultry and evidence for clonal expansion of resistant lineages. *BMC Microbiol.* 13, 160. doi: 10.1186/1471-2180-13-160

Woodcock, D. J., Krusche, P., Strachan, N. J. C., Forbes, K. J., Cohan, F. M., Méric, G., et al. (2017). Genomic plasticity and rapid host switching can promote the evolution of generalism: a case study in the zoonotic pathogen Campylobacter. *Sci. Rep.* 7, 9650. doi: 10.1038/s41598-017-09483-9

Wu, Z., Periaswamy, B., Sahin, O., Yaeger, M., Plummer, P., Zhai, W., et al. (2016). Point mutations in the major outer membrane protein drive hypervirulence of a rapidly expanding clone of Campylobacter jejuni. *Proc. Natl. Acad. Sci. U. S. A.* 113, 10690–10695. doi: 10.1073/pnas.1605869113

Yamaji, R., Rubin, J., Thys, E., Friedman, C. R., and Riley, L. W. (2018). Persistent pandemic lineages of uropathogenic Escherichia coli in a college community from 1999 to 2017. *J. Clin. Microbiol.* 56, e01834-17. doi: 10.1128/JCM.01834-17

Zerbino, D. R., and Birney, E. (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18, 821–829. doi: 10.1101/gr.074492.107