

Comparative genomics of quinoloneresistant and susceptible *Campylobacter jejuni* of poultry origin from major poultry producing European countries (GENCAMP)

Leekitcharoenphon, Pimlapas; GarciaGraells, Cristina; Botteldoorn, Nadine; Dierick, Katelijne; Kempf, Isabelle; Olkkola, Satu; Rossi, Mirko; Nykäsenoja, Suvi; Malorny, Burkhard; Stingl, Kerstin; Battisti, Antonio; Franco, Alessia; Mossong, Joël; Veldman, Kees; Mevius, Dik; Wasyl, Dariusz; Wieczorek, Kinga; Osek, Jacek; Clemente, Lurdes; Lacatus, Angela; Nicorescu, Isabela; García, Monserrat Agüero; Escobar, Cristina De Frutos; Ferrer, Manuel Duran; UgarteRuiz, María; Anjum, Muna F; Teale, Christopher; Agren, Joakim; Hendriksen, Rene S.; Aarestrup, Frank Møller

Published in:
EFSA Supporting Publications

Link to article, DOI:
[10.2903/sp.efsa.2018.EN-1398](https://doi.org/10.2903/sp.efsa.2018.EN-1398)

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Leekitcharoenphon, P., GarciaGraells, C., Botteldoorn, N., Dierick, K., Kempf, I., Olkkola, S., ... Aarestrup, F. M. (2018). Comparative genomics of quinoloneresistant and susceptible *Campylobacter jejuni* of poultry origin from major poultry producing European countries (GENCAMP). EFSA Supporting Publications, 15(5), [1398E]. DOI: 10.2903/sp.efsa.2018.EN-1398

DTU Library

Technical Information Center of Denmark

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

APPROVED: 12 March 2018

doi:10.2903/sp.efsa.2018.EN-1398

Comparative genomics of quinolone-resistant and susceptible *Campylobacter jejuni* of poultry origin from major poultry producing European countries (GENCAMP)

Pimlapas Leekitcharoenphon¹, Cristina Garcia-Graells², Nadine Botteldoorn², Katelijne Dierick², Isabelle Kempf³, Satu Olkkola⁴, Mirko Rossi⁴, Suvi Nykäsenoja⁴, Burkhard Malorny⁵, Kerstin Stingl⁵, Antonio Battisti⁶, Alessia Franco⁶, Joël Mossong⁷, Kees Veldman⁸, Dik Mevius⁸, Dariusz Wasyl⁹, Kinga Wieczorek¹⁰, Jacek Osek¹⁰, Lurdes Clemente¹¹, Angela Lacatus¹², Isabela Nicorescu¹², Monserrat Agüero García¹³, Cristina De Frutos Escobar¹³, Manuel Duran Ferrer¹³, María Ugarte-Ruiz¹⁴, Muna F Anjum¹⁵, Christopher Teale¹⁵, Joakim Ågren¹⁶, Rene S. Hendriksen¹, Frank M. Aarestrup¹

¹National Food Institute, Technical University of Denmark, WHO Collaborating Center for Antimicrobial Resistance in Food borne Pathogens and Genomics, European Union Reference Laboratory for Antimicrobial Resistance, Lyngby, Denmark. ²Institut Scientifique Santé Publique (WIV-ISP) Faculty of Veterinary, University of Zaragoza. ³ANSES, Ploufragan Laboratory, France. ⁴Microbiology Research Unit, Finnish Food Safety Authority Evira, Finland. ⁵Federal Institute for Risk Assessment (BfR), NRL for the Analysis and Testing of Zoonoses (*Salmonella*), Berlin. ⁶Istituto Zooprofilattico Sperimentale del Lazio e della Toscana, Rome, Italy. ⁷Laboratoire National de Santé, Dudelange, Luxembourg. ⁸Central Veterinary Institute (CVI) part of Wageningen UR, Lelystad, The Netherlands. ⁹Department of Microbiology, National Veterinary Research Institute, Pulawy, Poland. ¹⁰Department of Hygiene of Food of Animal Origin, National Veterinary Research Institute, Pulawy, Poland. ¹¹INIAV - National Institute for Agrarian and Veterinary Research, Microbiology and Mycology Laboratory, Lisbon, Portugal. ¹²Institute for Hygiene and Veterinary Public Health, Romania. ¹³Laboratorio Central de Veterinaria-Sanidad Animal, Algete, Madrid, Spain. ¹⁴VISAVET Health Surveillance Centre. Universidad Complutense Madrid, Spain. ¹⁵Department of Bacteriology, APHA Weybridge, Surrey, UK. ¹⁶Department of Microbiology, National Veterinary Institute, Uppsala, Sweden

Abstract

A total of 502 *Campylobacter jejuni* isolates from poultry in 12 different European countries (10 of them the largest poultry production countries in Europe) were whole genome sequenced to examine the genomic diversity of fluoroquinolone resistant (FQ-R) and susceptible (FQ-S) *C. jejuni* across the poultry producing European countries and to determine whether the emergence of fluoroquinolone resistance among *C. jejuni* is related to the transmission through countries or to the selection through fluoroquinolone use in the individual countries. A high genomic diversity was observed. The isolates clustered in four main clusters. All trees revealed that the isolates were clustered according to the presence/absence of the *gyrA* mutations causing fluoroquinolone resistance and ST-types. The cgMLST trees of only FQ-R and FQ-S isolates showed that isolates from the same country of origin were distributed into multiple clusters similarly to the trees combining FQ-R and FQ-S isolates. The different phylogenetic methods, ranging from single nucleotide polymorphisms analysis to gene-by-gene approaches such as rMLST, cgMLST, wgMLST and core genome tree, provided concordant results, but it is not known which is the most accurate method for identifying the country of origin of the isolates. Allele frequency analysis of isolates under this study and a selection of previously published *C. jejuni* genomes in ENA showed association of geographical origin of poultry *C. jejuni* populations between Romania-Poland, Italy-Germany-England, Portugal-The Netherlands and USA-Luxembourg. Allele frequency and phylogenetic analysis indicated that the isolates from Finland were genetically different from *C. jejuni* populations from other European countries included in this study.

Trade pattern and antimicrobial use in livestock were not significantly associated with allele frequency or populations of *C. jejuni*, but data available to investigate these associations were limited.

© Technical University of Denmark - National Food Institute, 2018

Key words: *Campylobacter*, fluoroquinolone, antimicrobial resistance, Europe, whole genome sequencing, comparative genomics, cgMLST

Question number: EFSA-Q-2016-00517

Correspondence: biocontam@efsa.europa.eu

Disclaimer: The present document has been produced and adopted by the bodies identified above as author(s). In accordance with Article 36 of Regulation (EC) No 178/2002, this task has been carried out exclusively by the author(s) in the context of a grant agreement between the European Food Safety Authority and the author(s). The present document is published complying with the transparency principle to which the Authority is subject. It cannot be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Suggested citation: Technical University of Denmark - National Food Institute, 2018. Comparative genomics of quinolone-resistant and susceptible *Campylobacter jejuni* of poultry origin from major poultry producing European countries (GENCAMP). EFSA supporting publication 2018:EN-1398. 35 pp. doi:10.2903/sp.efsa.2017.EN-1398

ISSN: 2397-8325

© Technical University of Denmark - National Food Institute, 2018

Summary

A total of 502 *Campylobacter jejuni* collected in 2014 from poultry in 12 different European Member States (MSs) were whole genome sequenced (WGS). Ten MSs were selected and included the study as being the largest poultry producing countries in Europe. Additionally two MSs were included the study due to available genomic data. All MSs provided minimal inhibitory concentration (MIC) data in 2013 and 2014 for ciprofloxacin resistance among *C. jejuni* from poultry. Of these 502 *C. jejuni*, 307 isolates contained the *gyrA* mutations (T86I, T86A and T86V) conferring fluoroquinolone resistance. The isolates were examined for the presence of antimicrobial resistance genes using ResFinder, and analysed by the conventional seven genes multilocus sequence typing (MLST, ST-types), core genome multilocus sequence typing (cgMLST, cgST-types), and phylogenetic analysis using core genome MLST (cgMLST), whole genome MLST (wgMLST), ribosomal MLST (rMLST), core genome, single nucleotide polymorphism (SNP)-analysis and the nucleotide difference tree (ND tree). In order to determine how the *C. jejuni* clades fit within the global *C. jejuni* populations, additional 536 previously published *C. jejuni* genomes were retrieved from ENA.

A total of 144 different ST-types were observed among the 502 *C. jejuni* isolates. Only 264 genomes could be assigned to 169 known cgST-types, whereas the remaining 238 genomes were of unknown cgST-types. The SNP tree was built from 49,224 SNPs identified from 1,020,450 valid positions out of 1,641,481 total bp of the reference genome (NCTC 11168). The ND tree was created using the same reference genome as for the SNP tree. The rMLST, cgMLST and wgMLST trees were constructed based on 52 loci, 1,343 loci and 1,643 loci schemes, respectively. The core genome tree was constructed based on 557 shared core genes. In general, all of the phylogenetic analyses produced four main clusters and a comparison showed that the trees were concordant at the internal nodes, with the cgMLST and wgMLST being the most concordant. All trees revealed that the isolates were clustered according to the presence/absence of the *gyrA* mutations for resistance to fluoroquinolones and ST-types. The cgMLST trees of only fluoroquinolone resistant (FQ-R) and susceptible (FQ-S) isolates showed that isolates from the same country of origin were distributed into multiple clusters similarly to the trees combining FQ-R and FQ-S isolates. This result suggested that the presence/absence of *gyrA* mutations did not influence the clustering of country of origin. There was a certain association of specific types within the different countries, especially allele frequency analysis showed association of geographical origin of poultry *C. jejuni* isolates from Romania-Poland, Italy-Germany-England, Portugal-the Netherlands and USA-Luxemburg. The allele frequency and phylogenetic analysis indicated that the isolates from Finland were different from *C. jejuni* populations from other European countries included in this study. The allele frequency indicated that the populations of *C. jejuni* isolates from poultry were more similar than the populations of *C. jejuni* isolates from other host species. Trade patterns and antimicrobial use in livestock were not significantly associated with allele frequency or populations of *C. jejuni* but available antimicrobial usage data refers to sales of veterinary antimicrobials for all livestock species, not specifically poultry. The lack of precise data on the antimicrobial use in poultry from the countries with large poultry production and the lack of metadata, for instance the information of poultry farms, trade and the amount of antimicrobials used in each of antimicrobial classes from the country of isolation, limit the ability to identify whether the emergence of antimicrobial resistance among *C. jejuni* is related to the selection through the antimicrobial use in these individual countries or to the transmission through countries.

In conclusion, this study showed a high diversity of *C. jejuni* across the 10 countries with the largest poultry production in Europe (Germany, Romania, Italy, Poland, the Netherlands, Spain, the United Kingdom, Belgium, France and Portugal). All phylogenetic approaches were concordant at internal nodes. The *C. jejuni* isolates were clustered based on *gyrA* mutations, ST-types and animal host species. None of the phylogenetic approaches used could identify the country of origin. Allele frequency could determine the association of *C. jejuni* population between countries of origin. Based on available data, no significant association was observed between trade patterns of poultry or antimicrobial use in livestock and population of *C. jejuni*.

Table of contents

Abstract.....	1
Summary	3
1. Introduction.....	5
1.1. Background	5
2. Objectives	5
3. Materials and methods	6
3.1. Materials	6
3.2. Methods	6
4. Results	11
5. Discussion and Conclusions	27
References.....	29
Abbreviations	32
Appendix A – Questionnaire to participants on selection and collection of isolates and related metadata	33
Annex A – Supplementary Table	35

1. Introduction

1.1. Background

This grant was awarded by EFSA to: Technical University of Denmark, National Food Institute (DTU Food)

Beneficiary: National Food Institute, Technical University of Denmark

Grant title: Genomic epidemiology of *Campylobacter jejuni* from poultry

Grant number: GP/EFSA/BIOCONTAM/2016/01

2. Objectives

Antimicrobial resistance is increasing globally and is threatening treatment options for humans and animals (WHO, 2014; Aarestrup, 2015). In addition to the use of antimicrobial agents for humans, livestock animals are considered an important source of antimicrobial resistant bacteria transmitted through the food chain, potentially compromising human health (Aarestrup, 2015).

Campylobacter jejuni is in many countries the most important foodborne pathogen. The emergence of fluoroquinolone resistance in *C. jejuni* is considered a major human health threat (Engberg et al., 2001) and one of the main reasons for categorizing the use of quinolones for humans as a major human health hazard (Collignon et al., 2009). In recent years, a major increase in fluoroquinolone resistance has been observed in *C. jejuni* from poultry in Europe (EFSA and ECDC, 2016).

The poultry production in Europe involves several Member States (MSs) with grand-parent flocks in one country, parent flocks in a second country, and production flocks potentially in a third country. It has previously been shown that antimicrobial resistance might be selected through use of antimicrobial agents in the grandparent flocks in one country and transmitted vertically through the production pyramid, without additional selective pressure in other countries (Aarestrup, 2015).

Genomic diversity of *C. jejuni* has in recent years mainly been studied using the conventional seven genes multilocus sequence typing or by whole genome sequencing (WGS) (Sheppard and Maiden, 2015). In poultry, many of the genomic lineages are globally distributed (Sheppard et al., 2010). There is also evidence for phylogeographic structuring and proliferation of specific lineages in different countries (McTavish et al., 2008; Asakura et al., 2012). In addition, studies have shown an association between specific clonal types and the emergence of fluoroquinolone resistance (Kinana et al., 2006; Habib et al., 2009; Kovač et al., 2014; Cha et al., 2017).

There are currently no studies that provide evidence on whether the observed emergence of fluoroquinolone resistance among *C. jejuni* might be related to vertical transmission or selected through quinolone use in the individual countries. Nor is there any evidence regarding whether it is related to a single or multiple clones or whether other epidemiological and/or microbiological factors might explain the increasing frequency of fluoroquinolone resistance in *C. jejuni*.

The main objectives of the study were:

- to determine the genomic diversity of FQ-R and FQ-S *C. jejuni* across the 12 European countries, including the 10 largest poultry production countries in the EU, and 2 additional countries;
- to investigate whether the diversity observed may be related with selected explanatory variables, such as country-specific use of quinolones and/or poultry trade connections.

Thus, this study examines the genomic diversity of fluoroquinolone-resistant (FQ-R) and fluoroquinolone-susceptible (FQ-S) *C. jejuni* in 12 European countries, including the 10 largest poultry producers in the EU, investigates whether trade patterns and antimicrobial use in livestock might be associated with allele frequency of *C. jejuni* or population of *C. jejuni* and thereby provides an initial basis for more detailed studies in the future to control and prevent the further spread of fluoroquinolone resistant *C. jejuni*.

3. Materials and methods

3.1. Materials

Selection and collection of isolates

At the kick-off meeting of the GENCAMP project held on the 18th of August 2016, it was agreed to enrol in the study the main poultry-producing MSs in Europe. The following 10 countries were targeted: Belgium, France, Italy, Germany, Poland, Portugal, Romania, Spain, the Netherlands and the United Kingdom (UK).

An invitation letter was sent on the 12th of September 2016, to the directors of the National Reference Laboratories (NRLs) for antimicrobial resistance in the above mentioned MSs describing the goals and the objectives of the study. The letter also included a description of what strains and which number of strains would be selected for the study, both FQ-S and FQ-R *C. jejuni* of poultry origin reported in the European Union Summary Report on Antimicrobial Resistance (EUSR-AR) in zoonotic and indicator bacteria from humans, animals and food in 2014. In addition, the letter also contained a template for a material transfer agreement (MTA) to ensure protection of any property rights and proper use of the material. A link to a questionnaire (Appendix A) was provided to all contact persons with the purpose to indicate the availability of metadata.

Finland and Luxemburg also showed interest in the study and provided additional *C. jejuni* genomes as a supplement to the collected isolates.

3.2. Methods

Collection and identification of isolates

A total of 511 FQ-R and FQ-S *C. jejuni* strains were collected from the ten target countries, largest European poultry producers, with Germany providing DNA of 70 additional strains instead of the isolates. Finland and Luxemburg showed interest to participate in the study and provided 86 additional genome sequences. The isolates from Finland represented *C. jejuni* present in Finland broiler production in 2014. The number of isolates requested from the participating countries was greater than the number of isolates agreed to be whole genome sequenced according to the grant agreement. The reason to request a larger number of isolates was to ensure reaching the agreed number considering the possibility to have contaminations or mis-identifications. The isolates were sub-cultured upon arrival on *Campylobacter* Charcoal Differential Agar (CCDA) and incubated for 24h at 42°C using microaerophilic conditions (6% O₂, 7% CO₂, 7% H₂ and 80% N₂). Subsequently, the strains were further cultured onto Columbia agar supplemented with 5% cattle blood and incubated as described above. The cultures were assessed by the naked eye for contaminations and stored at -80°C if assessed pure. Species identification was determined according to the polymerase chain reaction (PCR) protocol by Denis et al. (1999)¹ and further subjected to antimicrobial susceptibility by disk diffusion using ciprofloxacin disks of 5 µg and according to the European Committee on Antimicrobial Susceptibility Testing (EUCAST).²

Selection of isolates, DNA extraction and whole genome sequencing

Out of the 581 *C. jejuni* strains/DNA, a subset of 419 strains were selected by selecting approximately every second isolate per country after sorting the isolates by date. The selection did not take into account

¹ As reported in the protocol recommended by the EURL-AR at https://www.eurl-ar.eu/CustomData/Files/Folders/21-protocols/280_protocol-for-campylobacter-november-2013.pdf

² http://www.eucast.org/ast_of_bacteria/disk_diffusion_methodology/

the geographical region of origin or any other data provided. The pre-selection was conducted to select a number of strains compliant with the one originally planned in the Grant agreement for this study (i.e. about 400 strains). The DNA was extracted from 380 isolates randomly selected among those received (nine MSs) with a slight over-representation of strains resistant to FQ-R compared to FQ-S profiles as the aim of the study was to characterise resistant *C. jejuni* strains, using an Invitrogen Easy-DNATM Kit (Invitrogen, Carlsbad, CA, USA). DNA concentrations were determined using the Qubit dsDNA BR assay kit (Invitrogen). The subset of 380 strains and 39 DNA samples (Germany) from the ten MSs initially included in the study, were whole genome sequenced. The genomic DNA was prepared for Illumina pair-end sequencing using the Illumina (Illumina, Inc., San Diego, CA) NexteraXT® Guide 150319425031942 following the protocol revision C.³

A sample of the pooled NexteraXT Libraries of the isolates or DNA from all MSs except for Romania were loaded onto a Illumina HiSeq reagent cartridge using HiSeq Reagent Kit v2. The libraries were sequenced using an Illumina HiSeq platform. Due to the late arrival of the Romanian isolates, these were whole genome sequenced using an Illumina MiSeq platform and corresponding an Illumina MiSeq reagent cartridge using MiSeq Reagent Kit v2.

The additional 59 isolates from Luxemburg were fragmented by Nextera XT and sequenced by Illumina MiSeq. The additional 27 genomes from Finland were sequenced by Illumina HiSeq. The raw reads were submitted to the European Nucleotide Archive (ENA) under project number PRJEB23492. The ENA accession number for individual strains can be found in the Supplementary Table in Annex A ('List genomes' sheet). The raw reads were *de novo* assembled using SPAdes 3.7.0.

Additional published genome sequences

In order to determine how the *Campylobacter jejuni* clades of this study fit within the global *C. jejuni* genomes, a set of 536 previously published *Campylobacter jejuni* genomes was retrieved from ENA.⁴ The published *C. jejuni* genomes were retrieved from different studies that are available on ENA. The genomes were selected based on the availability of the following metadata: country of isolation, isolation date and year and isolation source. Among the genomes from Canada, France, UK and USA, there are highly redundant genomes from the same sources, years and study. The genomes from those countries were randomly chosen to cover as many sources, years as possible and the maximum numbers of selected genomes per country was 120 to reduce the bias of number of selected genomes in a selected country. The global set consisted of strains from Antarctica (n=3), Canada (n=120), Denmark (n=4), Estonia (n=4), Finland (n=20), France (n=100), Japan (n=1), Lithuania (n=2), Spain (n=59), Sweden (n=7), UK (n=98) and USA (n=118). The isolates originated mainly from poultry, human, cattle and sheep, and also included a few from cat, dog, environment, horse, raccoon, and pigs.

Species identification, conventional seven multilocus sequence typing, screening for antimicrobial resistance genes, and chromosomal point mutations

The assembled sequences were analyzed using *in silico* bioinformatics tools to confirm species, the MLST sequence type (ST-type) for *C. jejuni*, acquired antimicrobial resistance genes (beta-lactams, nitroimidazole, MLSs- Macrolides-Lincosamides-StreptograminsB, sulphonamides, tetracyclines, fosfomycin, oxazolidinones, aminoglycosides, glycopeptides, rifampicin, trimethoprim, phenicols, fusidic acid and colistin resistance) and chromosomal point mutations for *Campylobacter* using the following pipelines: KmerFinder (version 2.0) (Hasman et al., 2014), MLST (version 1.7) (Larsen et al., 2012), and ResFinder (version 3.0) (Zankari et al., 2012) available from the Center for Genomic Epidemiology (CGE).⁵

³ http://support.illumina.com/downloads/nextera_xt_sample_preparation_guide_15031942.html

⁴ <https://www.ebi.ac.uk/ena>

⁵ <http://cge.cbs.dtu.dk/services/>

Criteria for choosing bioinformatics approaches

Different phylogenetic approaches have been tested. We used all available phylogenetic methods in this study, as there is no evaluation of phylogenetic approaches on *C. jejuni*. The methods chosen range from single nucleotide polymorphisms level to gene-by-gene approaches such as rMLST, cgMLST, wgMLST and core genome tree. We further applied population genetics for instance allele frequency to elucidate the link between trade, antimicrobial use in livestock and population of *C. jejuni* (allele frequency).

Single nucleotide polymorphisms tree

Single nucleotide polymorphisms (SNPs) were identified using the pipeline CSI phylogeny (Kaas et al., 2014) available from CGE. The paired-end reads were mapped to the reference genome, *Campylobacter jejuni* subsp. *jejuni* NCTC 11168 (accession number; NC_002163.1), using Burrows-Wheeler Aligner (BWA) version 0.7.2 (Li and Durbin, 2009). The SNPs were called using 'mpileup' module in SAMTools version 0.1.18 (Li et al., 2009). Subsequently, the SNPs were selected when they met the following criteria: 1) a minimum distance of 15 bps between each SNP (pruning), 2) a minimum of 10% of the average depth, 3) the mapping quality was above 25, 4) the SNP quality was more than 30, and 5) all INDELs were excluded. The qualified core SNPs were concatenated to a single alignment corresponding to the position of the reference genome. The concatenated sequences were subjected to parsimony tree construction using PhyML (Guindon et al., 2010) with the HKY85 substitution model.

ND tree

The reference genome was split into *k*-mers and stored in a hash table. When all of the raw reads had been mapped, the significance of the base call at each position was evaluated by calculating the number of reads *X* having the most common nucleotide at that position, and the number of reads *Y* supporting other nucleotides. A Z-score was calculated as $Z = (X-Y)/\sqrt{X+Y}$. The value of 1.96 was used as a threshold for *Z* corresponding to a p-value of 0.001. It was further required that $X > 10*Y$.

Each pair of sequences was compared and the number of nucleotide differences in positions called in all sequences was counted. A matrix with these numbers was given as input to a UPGMA algorithm implemented in the neighbour program⁶ in order to construct the tree. The ND tree approach was implemented as a pipeline tool on the CGE website⁷ (Leekitcharoenphon et al., 2014).

rMLST, cgMLST and wgMLST trees

The rMLST (52 loci), cgMLST (1,343 loci) and wgMLST (1,643 loci) loci sequences were retrieved from <http://pubmlst.org/campylobacter> (Cody et al., 2017). The rMLST, cgMLST and wgMLST loci sequences were indexed for use with KMA (*k*-mer alignment) with a *k*-mer size of 16. Raw reads were mapped with KMA against the rMLST, cgMLST and wgMLST sequences to identify the best-matched allele numbers. The threshold for read mapping were match = 1, mismatch = -2, gap-opening = -3, gap-extension = -1. There is no threshold for read depth. Trying for best matching alleles is more of a rule than an exception when doing cgMLST, this means that for every read we get there is almost always several equally well matching alleles. In order to resolve this with high precision, the ConClave sorting scheme is used. While all equally well matching alleles are saved in the first iteration, in a second iteration the most likely allele is chosen as the one with the highest number of matches.

KMA is a hash-based mapping method, designed to map raw reads directly against highly redundant databases (Clausen et al., 2017). In order to do this, KMA uses the 'ConClave' scoring scheme, where all the best alignments for each read are initially saved and a global alignment score is calculated for each

⁶ <http://evolution.genetics.washington.edu/phylip.html>

⁷ <http://www.cge.cbs.dtu.dk/services/NDtree/>

template sequence. Where this global score is calculated as the sum of all alignment scores for each template, for which the template was among the best mapping templates. When all mappings between reads and templates have been performed, one best-matching template is chosen for each read as the one with the highest global alignment score, this way dealing with the redundant sequences in a strictly deterministic manner (Clausen et al., 2017).

The alignment itself is based on seed and extend, where matching *k*-mers are extended to longer stretches of an identical sequence between query and template. Seed-extends are then pieced together using the Needleman-Wunsch algorithm (Needleman et al., 1970). The ends of the sequences are aligned by a modified Needleman-Wunsch, where the alignment is cut when one of the sequences end (Clausen et al., 2017). KMA uses the Needleman-Wunsch algorithm between seed-extends to get a high resolution of indels and mismatches. The filtering of detected genes are done in the post-processing of the KMA mappings, where stop-codons and frame shifts are not taken into account. KMA predicts the most likely alleles based on the q-value.

KMA is freely available at: <https://bitbucket.org/genomicepidemiology/kma> and <https://cge.cbs.dtu.dk/services/kma>. The cgMLSTFinder tool can be found at: <https://cge.cbs.dtu.dk/services/cgMLSTFinder-1.0/>.

The pairwise dissimilarities (distances) between genomes were identified based on the allele profile using 'gower' distance method in R. The rMLST, cgMLST and wgMLST trees were constructed from the distance matrix using hierarchical clustering in R. R codes for distance calculation and hierarchical clustering are as follows:

```
allele_profile <- read.table("matrix_allele_profile.txt", sep = "|t", row.names=1, colClasses =
"factor")
library(cluster)
library(ape)
tree <- as.phylo(hclust(daisy(allele_profile, metric="gower")))
write.tree(phy=tree, file="tree.newick")
```

Core genome ST type (cgST) was assigned by comparing core-genome alleles numbers with the core-genome allele profiles from Cody et al. (2017) with the cut-off ≤ 50 missing loci (Cody et al., 2017).

Core genome tree

The shared core genome was determined using SeqSphere+ (Ridom GmbH, Münster, Germany) software⁸ via assembled genomes. First suitable MLST genes were determined using the *C. jejuni* genome RM1221 (NCBI accession NC_003912) with default settings with default parameters as described previously (Ruppitsch et al., 2015). The 1,271 genes were then queried to each assembly. Truncated and missing genes from each isolate were discarded from the final analysis. The shared core genome consisted of 557 genes for the 502 *C. jejuni* isolates in this study. The core genome tree was built from the shared core genome using neighbour joining method.

Tree comparison

The trees were compared branch to branch using 'cophyloplot' function in R. Example of R code is below:

```
association <- matrix(ncol=2, nrow=59)
association[,1] <- association[,2] <- tree$tip.label
cophyloplot(tree1, tree2, assoc=association, length.line=0, space=1000, gap=0)
```

⁸ <http://www.ridom.de/seqsphere/>

The mean relative similarity between trees for all nodes was determined using function 'all.equal.list' from R package 'ape'. Example of R code is below:

```
library(ape)
tree1 <- read.tree("tree1.newick")
tree2 <- read.tree("tree2.newick")
all.equal.list(tree1,tree2)
```

Principal coordinate analysis of allele frequency

The allele frequency was calculated at host and country level from allele numbers of the conventional seven genes multilocus scheme from the 502 *C. jejuni* isolates and the 536 previously published *C. jejuni* genomes (Supplementary Table in Annex A - 'Allele frequency by source' sheet). In order to perform the principal coordinate analysis (PCoA) for determining the association between host and country based on allele frequency, the allele frequency matrix was calculated for the Bray-Curtis (BC) dissimilarities using R package vegan (Legendre and Gallagher, 2001). The PCoA plot was carried out using 'cmdscale' function in R. An example of R code can be found online.⁹

Trade connectivity of poultry, antimicrobial use in livestock and Procrustes analysis

The trade connectivity of poultry was the quantity in tons of grandparent and parent female chickens that were imported to the EU MSs included in this study, based on Eurostat figures for year 2014. The data can be found in the Supplementary Table in Annex A - 'Trade' sheet.

Data for national livestock antimicrobial usage (AMU) was obtained from the European Medicines Agency's 2014 European Surveillance of Veterinary Antimicrobial Consumption (ESVAC) report and was stratified by major drug family (EMA, 2016). The mass of active compound sold for use in all food-producing species (including horses) in 2014 was normalized by dividing with the population correction unit (PCU) in tonnes – approximating the biomass. The PCU is the unit that allows interspecies integration by adjusting for import/export and differences in average weight between species when they are most likely to receive antimicrobial treatment. The country-specific livestock drug use can be found in Annex A - 'AMU' sheet.

In order to test the association between country-specific trade connectivity and the allele frequency and the association between country-specific AMU patterns and the allele frequency, a Procrustes analysis using the vegan R package was performed. PCoA was generated from Bray-Curtis (BC) dissimilarities using the R package vegan (Legendre and Gallagher, 2001) for the PCU-corrected AMU, the trade connectivity data and allele frequency in country level. The symmetric Procrustes correlation coefficients, p-value and plots between trade connectivity against the allele frequency and AMU against the allele frequency were obtained using the 'protest' and 'procrustes' functions in vegan (Oksanen et al., 2016).¹⁰

⁹ https://bitbucket.org/patrickmunk/mganalysis/src/4d6ff8823ec467658abd3b5a15650fa68a1d3b75/R/plotting_functions.R?at=master&fileviewer=file-view-default

¹⁰ R codes for PCoA plot and procrustes analysis can be found at the following link: https://bitbucket.org/patrickmunk/mganalysis/src/4d6ff8823ec467658abd3b5a15650fa68a1d3b75/R/plotting_functions.R?at=master&fileviewer=file-view-default

4. Results

Number of isolates

The first batch of *C. jejuni* isolates was received from Poland on the 5th of October 2016 and the last batch from Romania on the 27th of February 2017. According to the grant agreement, 419 of the 581 strains and DNA samples were whole genome sequenced. In total, including the genomes provided by Finland and Luxemburg, 505 *Campylobacter* spp. genomes were included in the study (Table 1). KmerFinder detected three genomes as *Campylobacter coli*. Therefore, these were excluded from the analysis and the final set of *C. jejuni* genomes for this study consisted of 502 genomes.

Table 1: Number of *Campylobacter* spp. isolates and genomes included in the study

Country	Number of isolates / DNA	Number of genomes	Number of genomes used in this study	Number of FQ-R genomes containing <i>gyrA</i> mutations
Romania	66	45	45	32
Italy	66	58	58	42
Poland	55	52	52	42
The Netherlands	70	38	38	21
Spain	54	34	34	32
United Kingdom	60	40	38	22
Belgium	57	39	38	18
France	70	64	64	35
Portugal	13	10	10	9
Germany	70 (DNA)	39	39	19
Finland	-	59	59	16
Luxembourg	-	27	27	19
Total	581	505	502	307

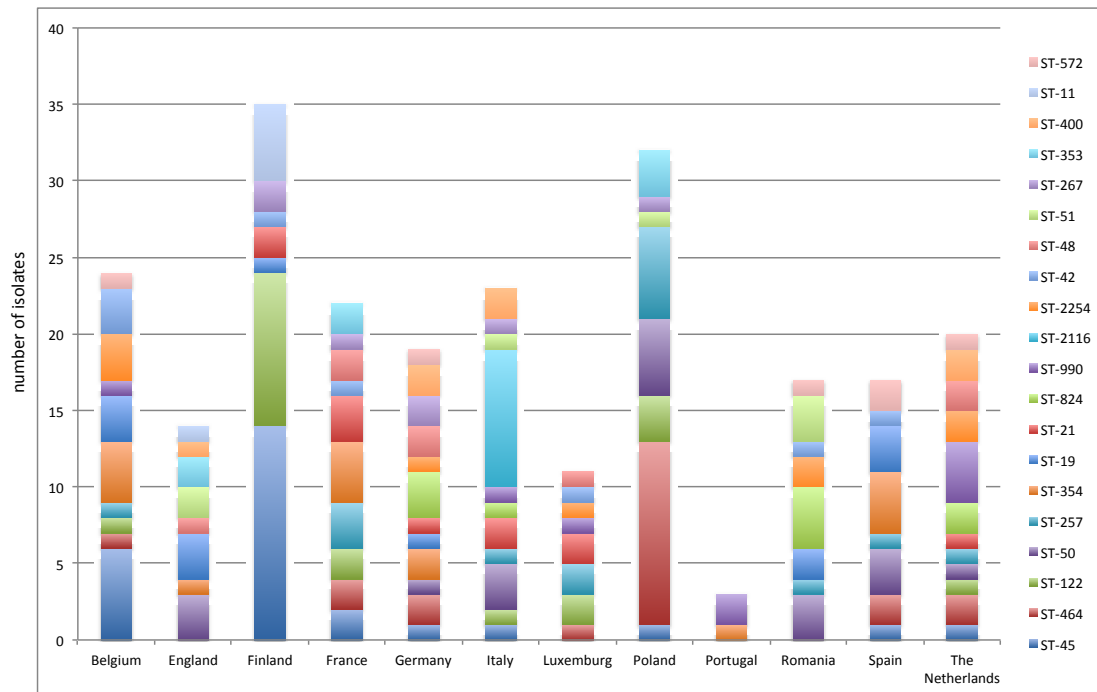
Chromosomal point mutations, acquired antimicrobial resistance genes, conventional seven genes multilocus sequence types (ST-types) and core genome multilocus sequence types (cgST-types)

The detection of chromosomal point mutations in *Campylobacter* spp. was performed using ResFinder version 3.0. The point mutations recognised by ResFinder for *Campylobacter* spp. included those in the *gyrA* (quinolones resistance), 23S rRNA (macrolides), *cmeR* (macrolides), L22 ribosomal protein (erythromycin) and *rpsL* (streptomycin) genes. The majority of *C. jejuni* (n=307 strains) contained *gyrA* mutations (T86I, T86A and T86V) conferring fluoroquinolone resistance and 195 strains did not have *gyrA* or any other mutations responsible for additional antimicrobial resistance. The number of FQ-R genomes per country is listed in Table 1. More detail on positions and amino acid changes of the point mutations can be found in the Supplementary Table in Annex A ('List genomes' and 'Point mutation in *Campylobacter*' sheets).

The following antimicrobial resistance genes were observed in the isolates using ResFinder: *bla*_{OXA-449}, *bla*_{OXA-465}, *bla*_{OXA-184}, *bla*_{OXA-448}, *bla*_{OXA-447}, *bla*_{OXA-466}, *bla*_{OXA-61}, *bla*_{OXA-185}, *bla*_{OXA-446}, *lnu*(C), *tet*(O), *aph*(3')-III, *aadE* (Supplementary Table in Annex A - 'List genomes' sheet).

The isolates were ascribed to diverse ST types. There were 144 different ST-types assigned by the seven house keeping gene scheme from *C. jejuni*. There are 21 genomes that remain unknown ST. Comparing core-genome alleles numbers with the core-genome allele profiles from Cody et al. (2017), 264 genomes

can be assigned to 169 different cgST types with ≤ 50 missing loci. The other 238 genomes remain unknown cgST types (Supplementary Table in Annex A ('cgST' sheet)). The distribution of top 20 ST-types was illustrated in Figure 1 and in Supplementary Table in Annex A ('ST top 20' sheet). The number of ST-types in each country was from 7 – 12 ST-types (excluding Portugal as there were only 10 genomes included in this study). The *C. jejuni* strains from Finland, Italy and Poland had 1 or 2 major ST-types, ST-45 (Finland), ST-122 (Finland), ST-2116 (Italy) and ST-464 (Poland).

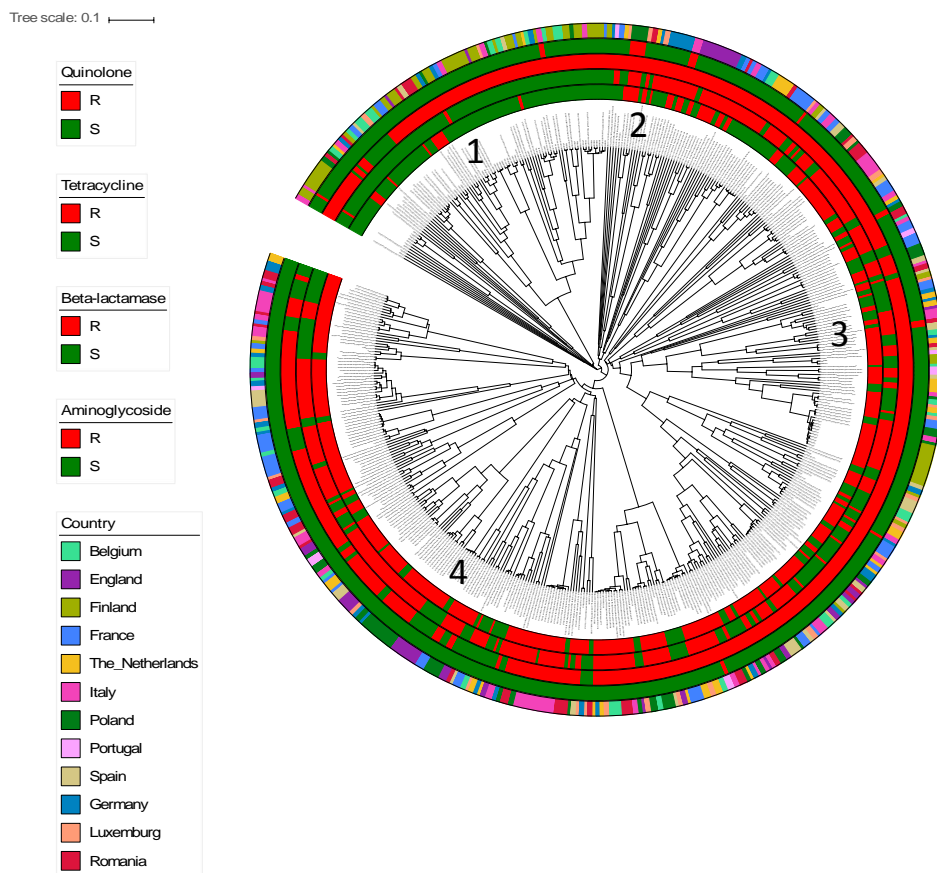


Y-axis is the number of strains.

Figure 1: Distribution of the top 20 ST-types (conventional seven multilocus sequence types)

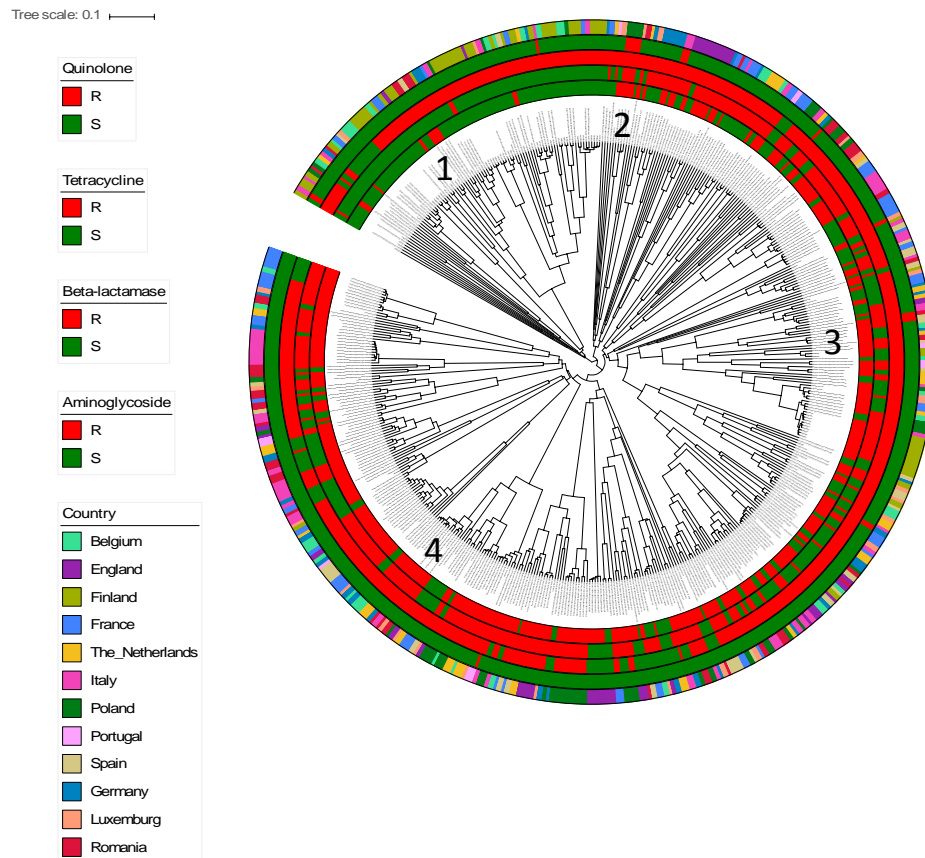
Phylogenetic tree and country association based on phylogeny

The cgMLST, wgMLST, SNP (Single Nucleotide Polymorphism), ND (Nucleotide Differences), rMLST and core genome trees were constructed (Figures 2-7). The SNP tree and ND tree are reference-based phylogeny, whereas cgMLST, wgMLST and rMLST are reference-free based methods and were constructed based on 1,343 loci, 1,643 loci and 52 loci schemes respectively. There were 101 SNPs left after recombination removed. Therefore, the SNP tree was constructed using the identified 32,392 SNPs before recombination detected. Additionally, the core genome tree was built based on 557 shared core genes.



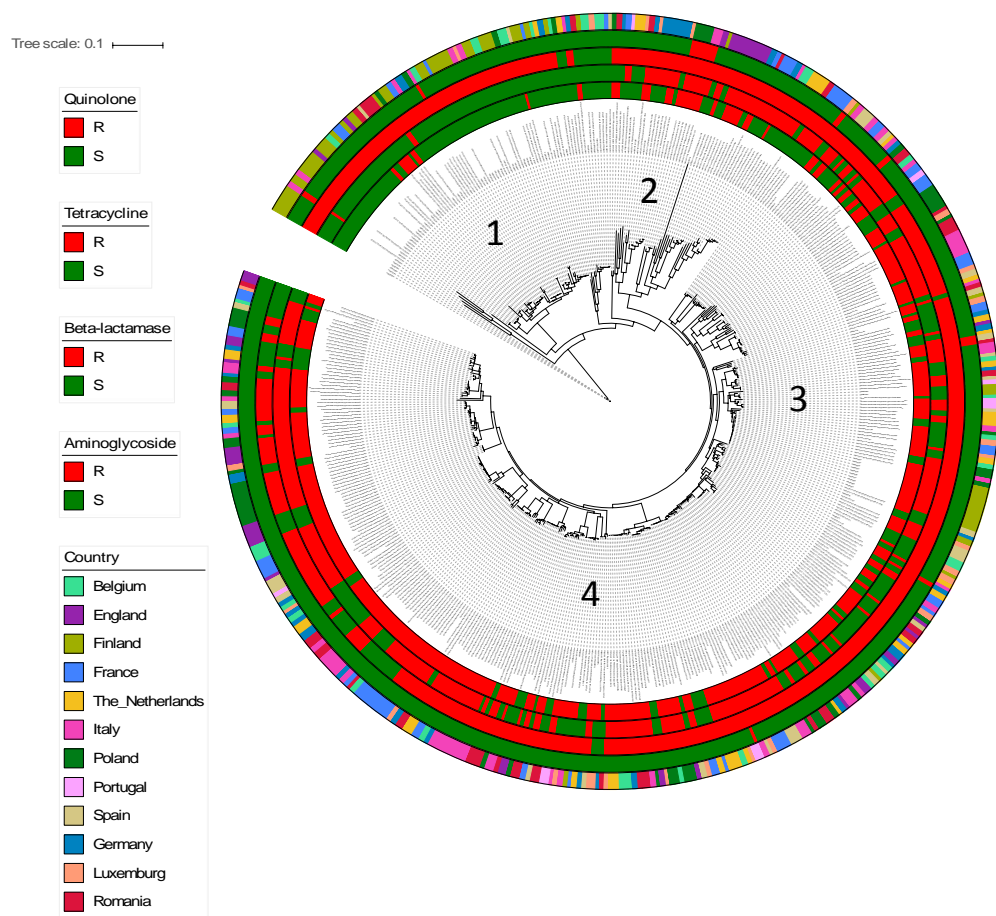
The order of the circle lanes from inner to outer was quinolone (R: resistance caused by *gyrA* mutations; S: no *gyrA* mutation detected), tetracycline (R: resistance caused by *tet(O)*; S: no *tet(O)* gene), beta-lactamase (R: resistance caused by *blaOXA* gene; S: no *blaOXA* gene), aminoglycoside (R: resistance caused by *aph(3')-III*, *aacE*; S: no *aph(3')-III*, *aacE* genes) and country.

Figure 2: Core genome MLST tree of the 502 isolates



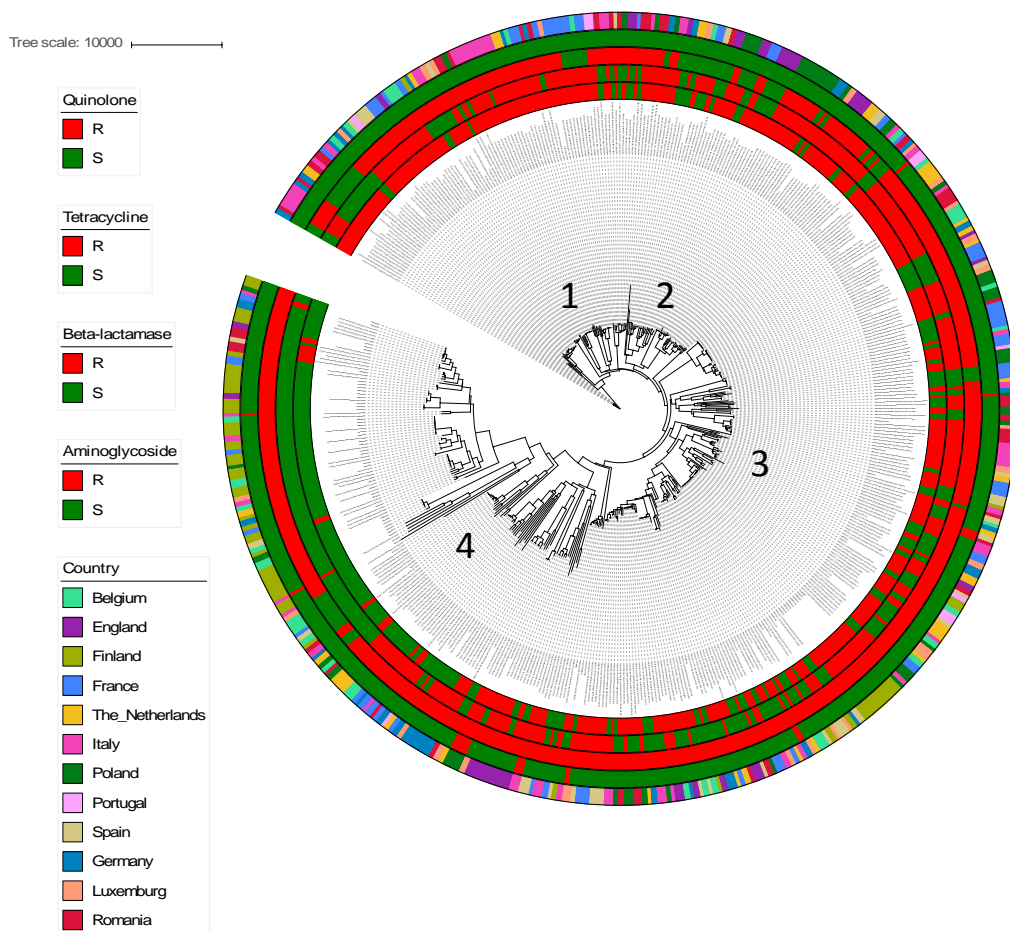
The order of the circle lanes from inner to outer was quinolone (R: resistance caused by *gyrA* mutations; S: no *gyrA* mutation detected), tetracycline (R: resistance caused by *tet(O)*; S: no *tet(O)* gene), beta-lactamase (R: resistance caused by *blaOXA* gene; S: no *blaOXA* gene), aminoglycoside (R: resistance caused by *aph(3')-III*, *aadE*; S: no *aph(3')-III*, *aadE* genes) and country.

Figure 3: Whole genome MLST tree of the 502 isolates



The order of the circle lanes from inner to outer was quinolone (R: resistance caused by *gyrA* mutations; S: no *gyrA* mutation detected), tetracycline (R: resistance caused by *tet(O)*; S: no *tet(O)* gene), beta-lactamase (R: resistance caused by *blaOXA* gene; S: no *blaOXA* gene), aminoglycoside (R: resistance caused by *aph(3')*-III, *aacE*; S: no *aph(3')*-III, *aacE* genes) and country.

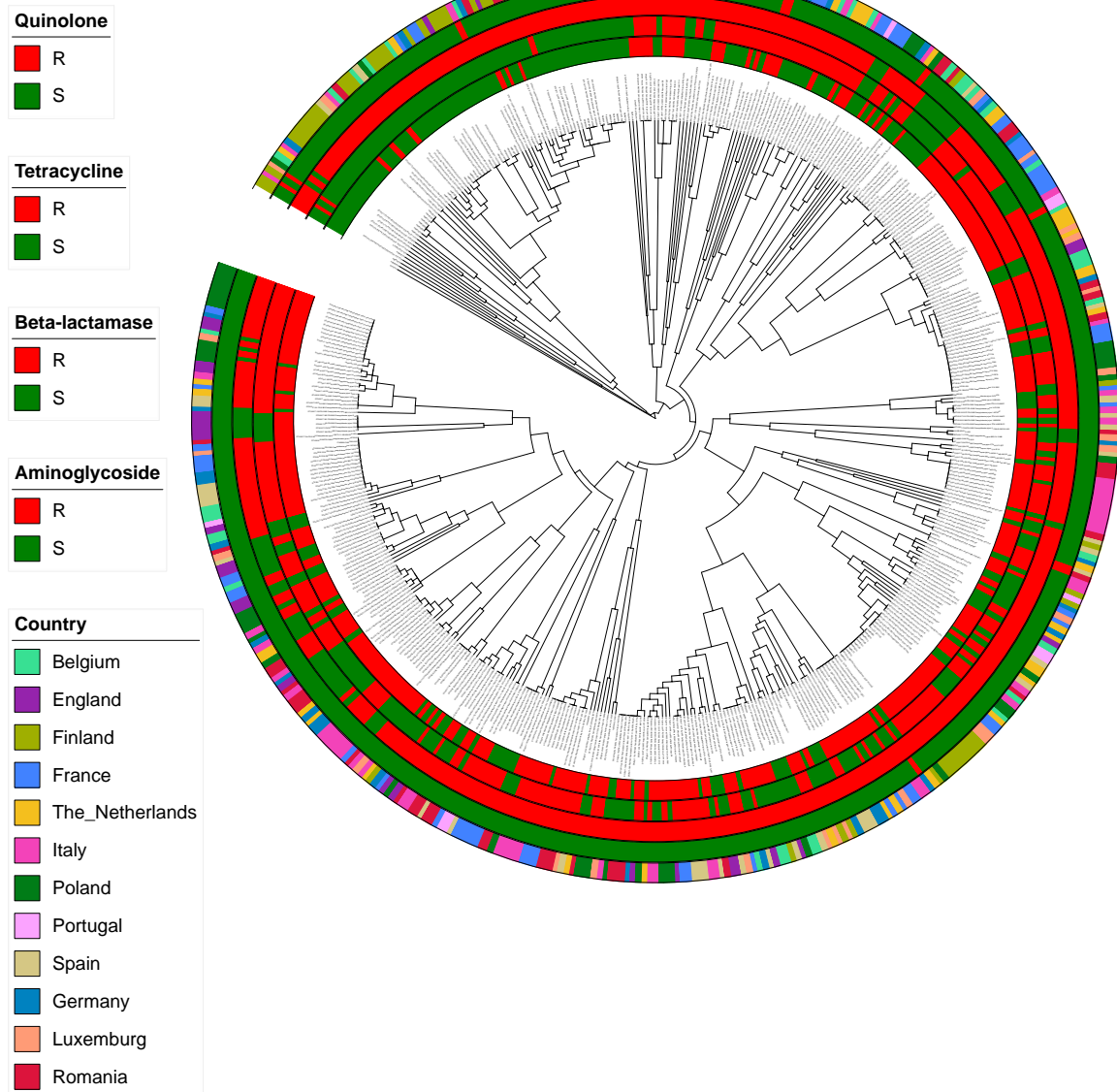
Figure 4: SNP tree of the 502 isolates



The order of the circle lanes from inner to outer was quinolone (R: resistance caused by *gyrA* mutations; S: no *gyrA* mutation detected), tetracycline (R: resistance caused by *tet(O)*; S: no *tet(O)* gene), beta-lactamase (R: resistance caused by *blaOXA* gene; S: no *blaOXA* gene), aminoglycoside (R: resistance caused by *aph(3')-III*, *aadE*; S: no *aph(3')-III*, *aadE* genes) and country.

Figure 5: ND tree of the 502 isolates

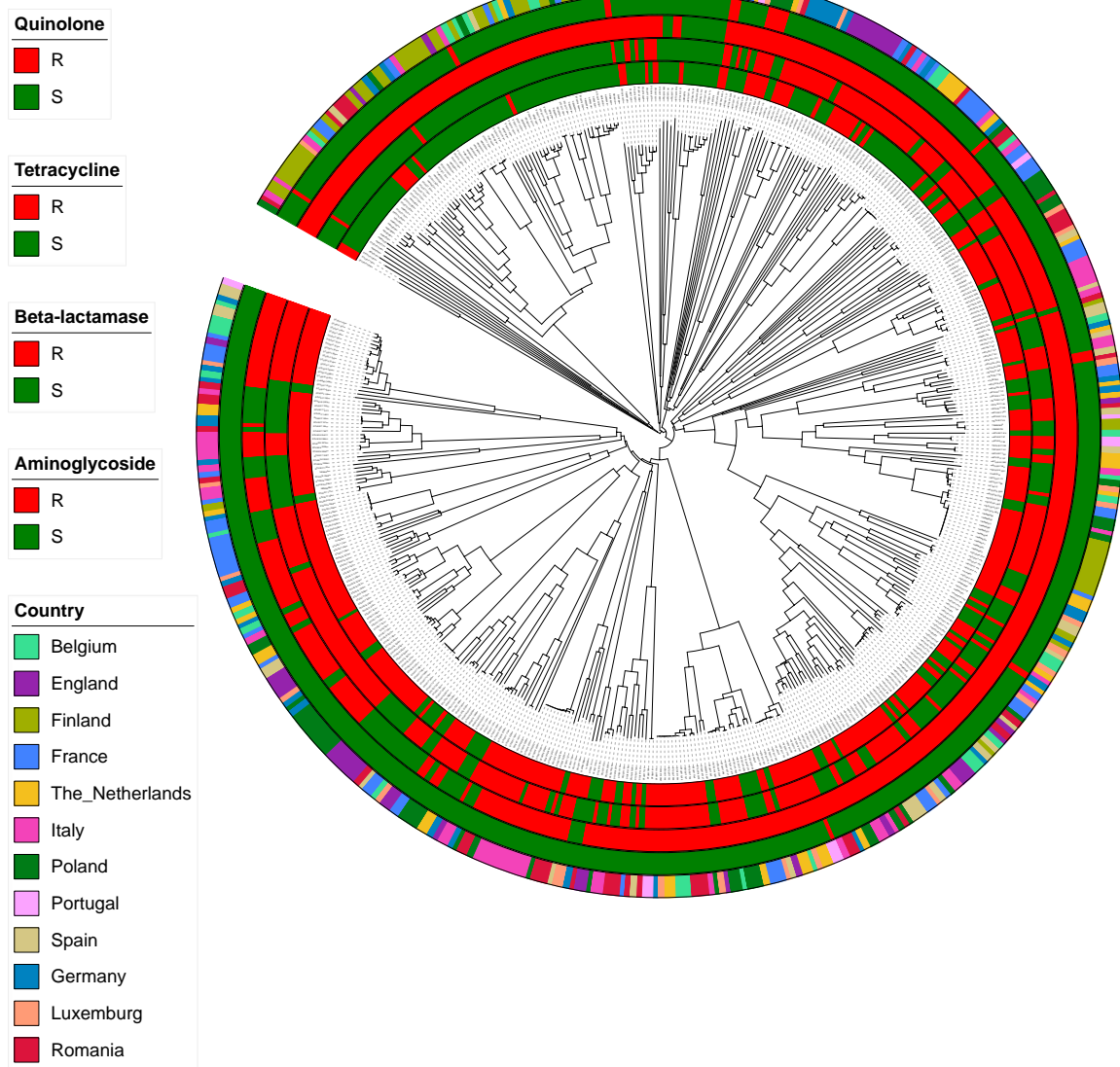
Tree scale: 0.1



The order of the circle lanes from inner to outer was quinolone (R: resistance caused by *gyrA* mutations; S: no *gyrA* mutation detected), tetracycline (R: resistance caused by *tet(O)*; S: no *tet(O)* gene), beta-lactamase (R: resistance caused by *blaOXA* gene; S: no *blaOXA* gene), aminoglycoside (R: resistance caused by *aph(3')*-III, *aacE*; S: no *aph(3')*-III, *aacE* genes) and country.

Figure 6: rMLST tree of the 502 isolates

Tree scale: 0.1



The order of the circle lanes from inner to outer was quinolone (R: resistance caused by *gyrA* mutations; S: no *gyrA* mutation detected), tetracycline (R: resistance caused by *tet(O)*; S: no *tet(O)* gene), beta-lactamase (R: resistance caused by *blaOXA* gene; S: no *blaOXA* gene), aminoglycoside (R: resistance caused by *aph(3')-III*, *aadE*; S: no *aph(3')-III*, *aadE* genes) and country.

Figure 7: Core genome tree of the 502 isolates

The trees were compared branch to branch using R. The mean relative similarity between trees was determined using R (Table 2). The comparison of trees showed that the trees were concordant at the internal nodes. The cgMLST and wgMLST were the most concordant as they are reference-free based method and constructed based on allele profiles of large set of multilocus sequences. Among the reference-free based trees, rMLST had the lowest relative similarity. The reference-free based trees and reference based trees (SNP and ND trees) showed low relative similarity.

Table 2: Mean relative similarity between reference-free and reference-based trees

	cgMLST	wgMLST	rMLST	Core genome	SNPs	ND
cgMLST	1	0.27	0.0020	0.24	0.0010	0.0010
wgMLST	0.27	1	0.0020	0.22	0.0010	0.0010
rMLST	0.0020	0.0020	1	0.0020	0.0030	0.0030
Core genome	0.24	0.22	0.0020	1	0.0010	0.0010
SNPs	0.0010	0.0010	0.0030	0.0010	1	0.26
ND	0.0010	0.0010	0.0030	0.0010	0.26	1

cgMLST: core genome multilocus sequence typing (MLST); ND: nucleotide differences; rMLST: ribosomal MLST; SNP: single nucleotide polymorphism; wgMLST: whole genome MLST.

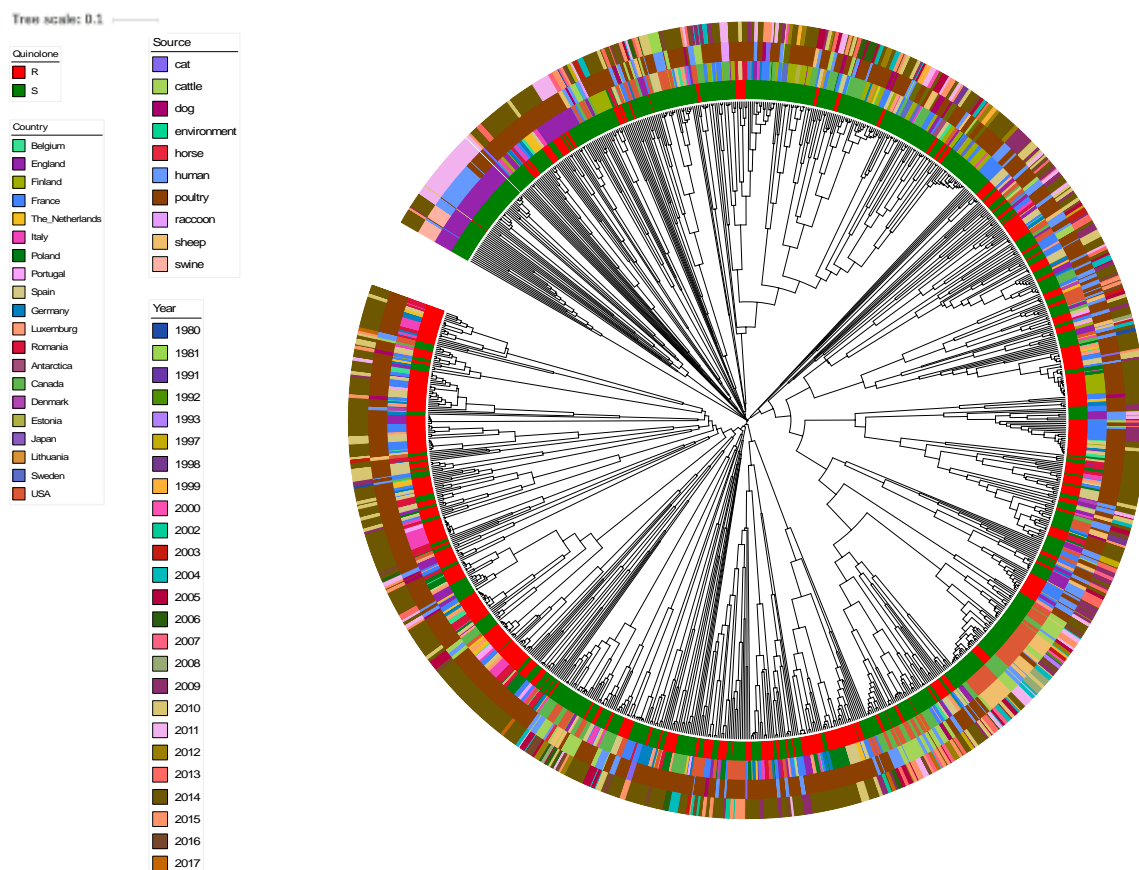
The cgMLST, wgMLST, SNP and ND trees were compared with each other to determine cluster similarity and agreement of clustering. Based on internal (ancestral) branching, the trees can be divided into four main clusters. The clusters from the cgMLST tree were used as reference clusters. The number of isolates found in the same clusters for each of the other trees is summarized in Table 3. The cgMLST and wgMLST trees had high agreement of clustering. The cgMLST and SNP tree were somehow in agreement except strains from cluster 2 that 46 out of 64 isolates from SNP tree were in agreement. The SNP tree was constructed from core SNPs. This might filter out some SNPs caused by horizontal genetic exchange and recombination. The ND tree was the least concordant with cgMLST in term of clustering agreement as the main cluster 1 and 2 from cgMLST were grouped into main cluster 4 of ND tree and the strains from cluster 4 of cgMLST were separated into three different clusters in ND tree (Table 3).

Table 3: Number of isolates found in the given clusters compared to the reference tree, cgMLST tree and number of total isolates in the given cluster

cgMLST tree	wgMLST tree	SNP tree	ND tree
Cluster 1 (90)	90/90 strains from cluster 1	72/72 strains from cluster 1	90/146 strains from cluster 4
Cluster 2 (46)	46/46 strains from cluster 2	46/64 strains from cluster 2	46/146 strains from cluster 4
Cluster 3 (174)	174/174 strains from cluster 3	174/174 strains from cluster 3	164/204 strains from cluster 3
Cluster 4 (192)	192/192 strains from cluster 4	192/192 strains from cluster 4	92/92 strains from cluster 1 60/60 strains from cluster 2 39/204 strains from cluster 3

In addition, the tree from rMLST using 52 loci and the core genome tree using 557 shared core genes provided a similar tree topology with 4 main clusters. All of the trees using either the gene by gene approaches (cgMLST, wgMLST, rMLST and core genome) or reference based approaches (SNP and ND), didn't cluster strains by geographical origin. All of the trees revealed that the isolates were clustered according to the presence/absence of *gyrA* mutations and ST-types.

In order to determine how the 502 poultry *C. jejuni* isolates attribute in a global phylogeny, a set of additional 536 available *C. jejuni* isolates were included in the cgMLST tree (Figure 8). The global tree (1,038 genomes) showed a large cluster of FQ-S isolates. There is no pattern of clustering according to country of origin and year of isolation. The majority of isolates were from poultry and they seem to be clustered together, whereas the human isolates spread sporadically throughout the tree.

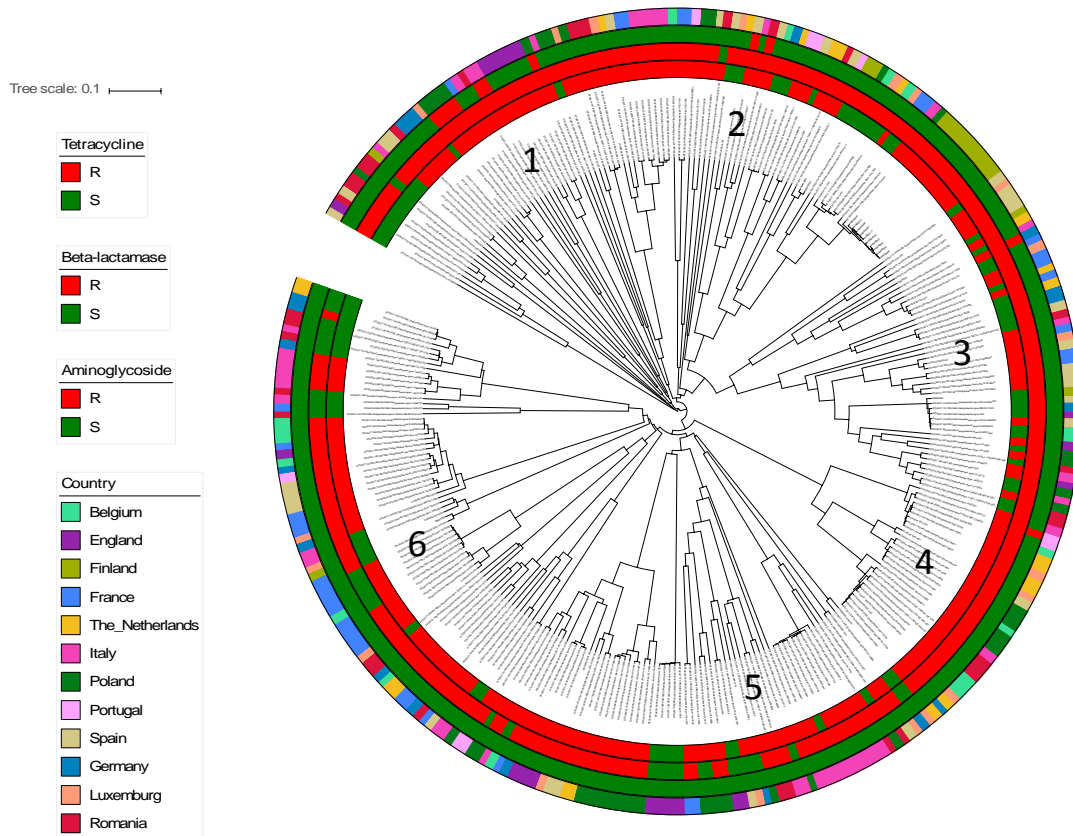


The order of the circle lanes from inner to outer was quinolone (R: resistance caused by *gyrA* mutations; S: no *gyrA* mutation detected), country, source and year.

Figure 8: cgMLST tree of global isolates

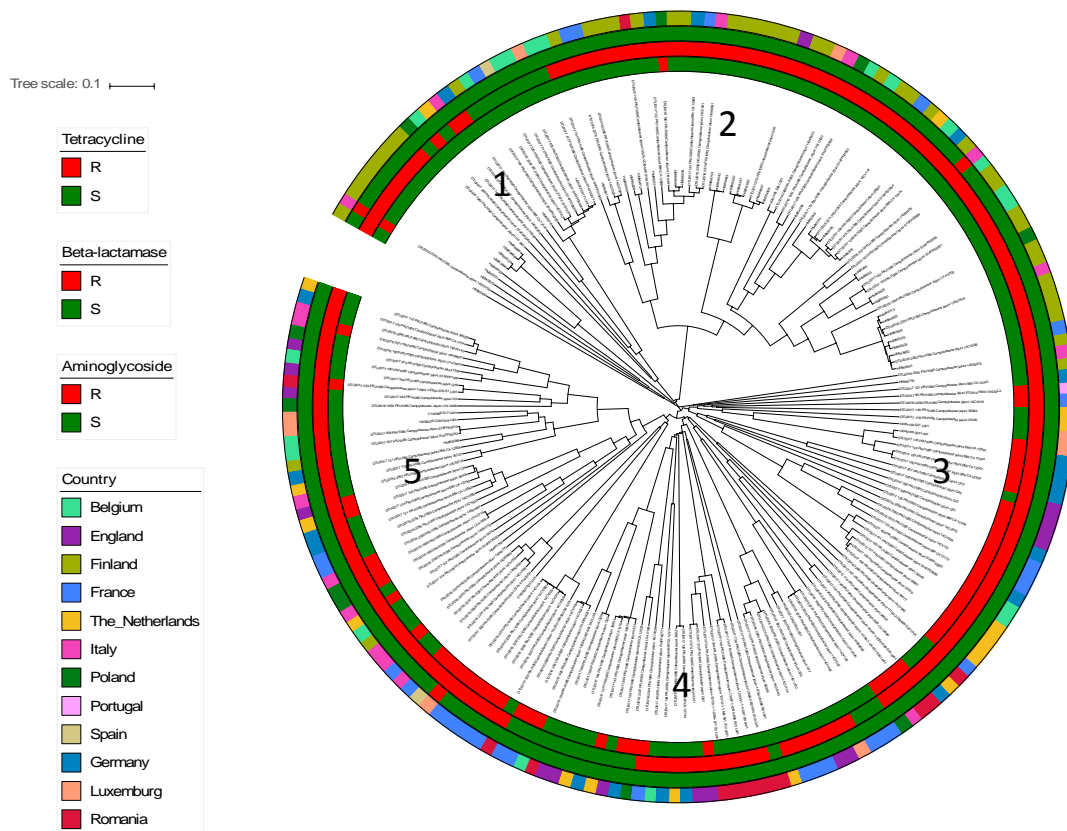
To understand whether the fluoroquinolone resistance among *C. jejuni* is associated to the individual countries without any effect from clustering due to the presence of *gyrA* mutations, we made two additional cgMLST trees from FQ-R isolates ($n=307$, presence of *gyrA* mutations) (Figure 9) and isolates without *gyrA* mutation ($n=195$) (Figure 10). The cgMLST tree for FQ-R isolates can be divided into six clusters and the cgMLST tree for non-*gyrA* mutation isolates can be separated into five clusters based on internal (ancestral) branching. From the clusters in the cgMLST trees, the distribution of countries was summarized as the number of isolates and percentage of isolates in Figures 11 and 12 for FQ-S and FQ-R isolates, respectively. Isolates from the same country, regardless of the presence of *gyrA* mutations, were associated with two or more clusters in the trees. The majority of FQ-R isolates from Finland were related

to a single cluster in FQ-R cgMLST tree and FQ-S isolates from Finland were associated with two clusters in FQ-S cgMLST tree, indicating that Finnish isolates were more clonal and restricted to Finland. All clusters from the cgMLST tree of FQ-R isolates contained strains from at least 9 to all 12 countries.



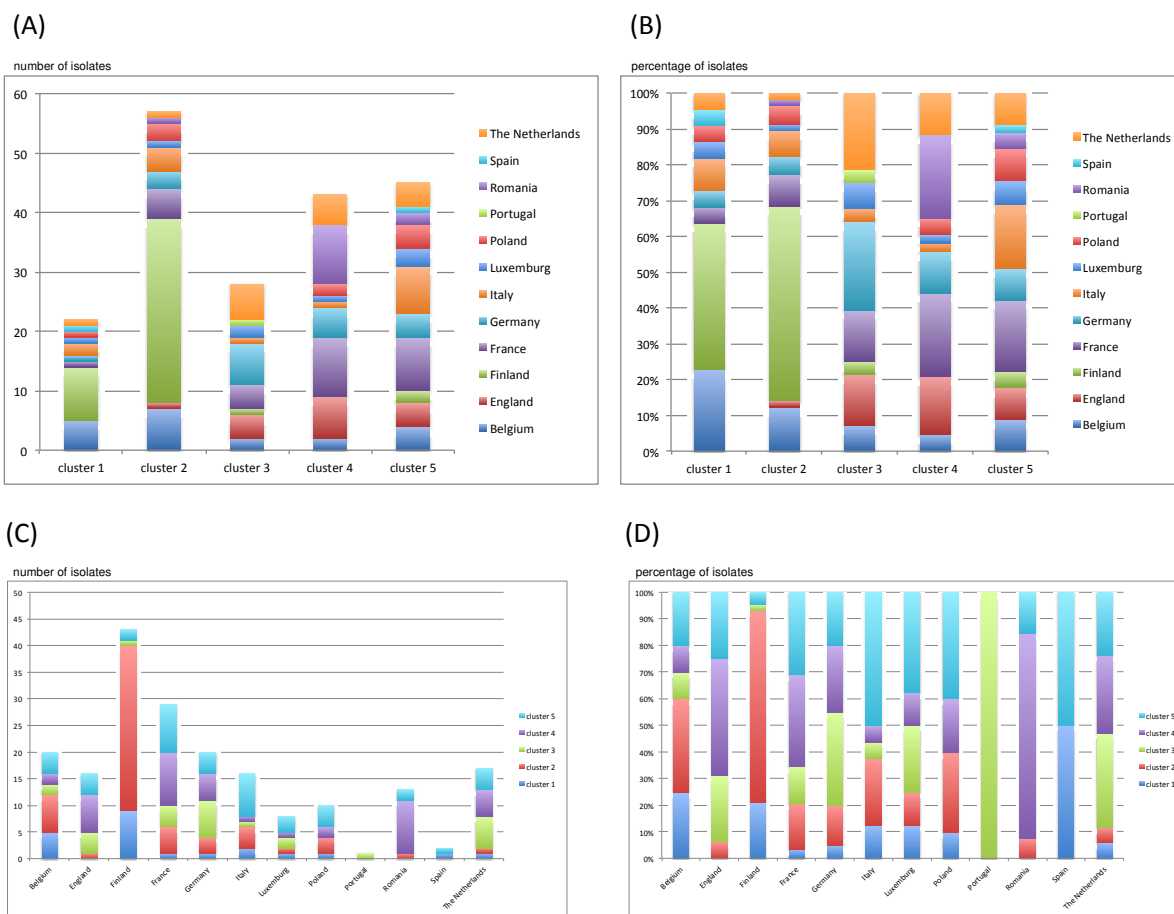
The order of the circle lanes from inner to outer was tetracycline (R: resistance caused by *tet(O)*; S: no *tet(O)* gene), beta-lactamase (R: resistance caused by *blaOXA* gene; S: no *blaOXA* gene), aminoglycoside (R: resistance caused by *aph(3')-III*, *aadE*; S: no *aph(3')-III*, *aadE* genes) and country.

Figure 9: cgMLST tree of the quinolone resistant isolates



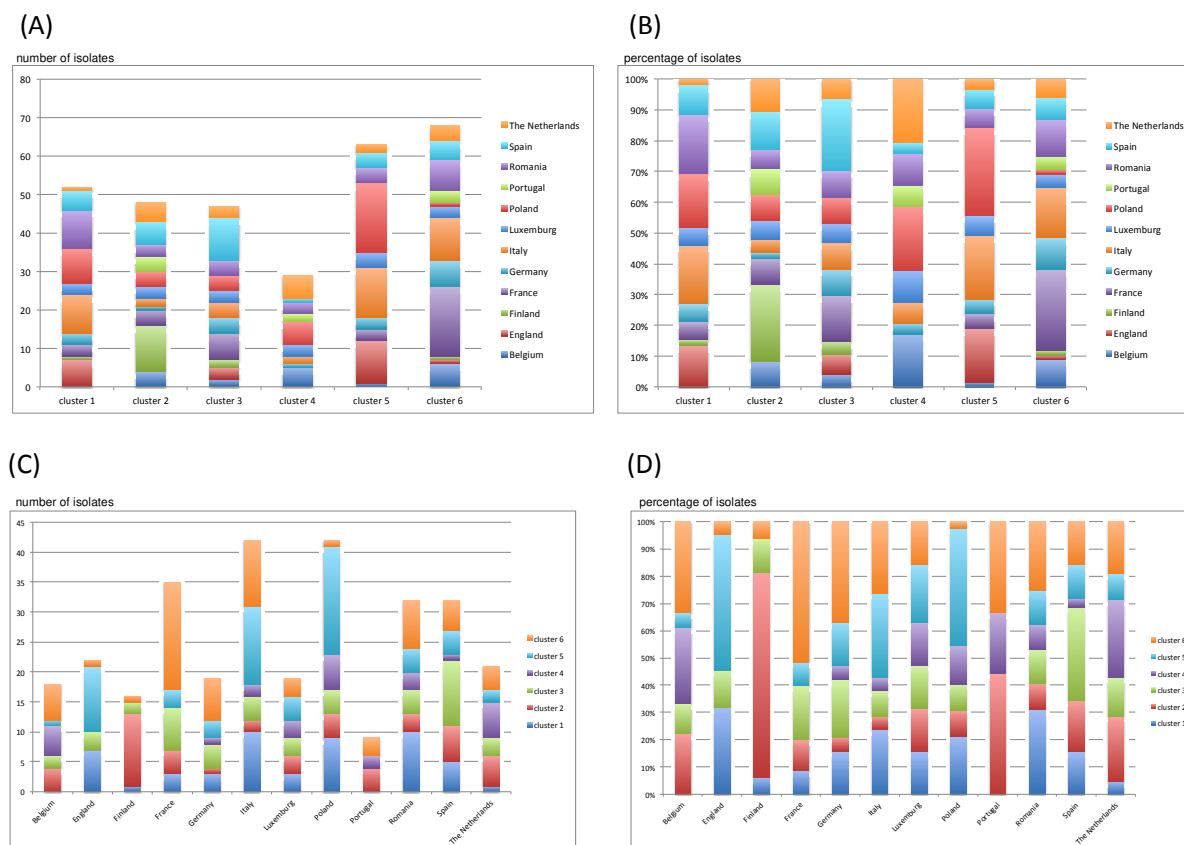
The order of the circle lanes from inner to outer was tetracycline (R: resistance caused by *tet(O)*; S: no *tet(O)* gene), beta-lactamase (R: resistance caused by *blaOXA* gene; S: no *blaOXA* gene), aminoglycoside (R: resistance caused by *aph(3')-III*, *aadE*; S = no *aph(3')-III*, *aadE* genes) and country.

Figure 10: cgMLST tree of the quinolone susceptible isolates



Y-axis in A and C are number of isolates. Y-axis in B and D are percentage of isolates.

Figure 11: Distribution of FQ-S isolates in the different clusters (A and B) and in different countries (C and D) defined by the cgMLST tree



Y-axis in A and C are number of isolates. Y-axis in B and D are percentage of isolates.

Figure 12: Distribution of the FQ-R isolates in the different clusters (A and B) and in different countries (C and D) defined by the cgMLST tree

Country association based on allele frequency of conventional seven genes multilocus sequence typing

In *C. jejuni* and *C. coli*, it is common that alleles are present in more than one population but at different frequencies (Pascoe et al., 2017). Based on this principle, the similar population of *Campylobacter* spp. should have similar allele frequency. Therefore, the matrix of allele frequency at host (excluding human isolates) and country levels from all isolates including global collection was subjected to principal coordinate analysis (PCoA) (Figure 13). Each dot in the PCoA plot represents a set of *C. jejuni* isolates from the same source and country. The PCoA showed that out of 708 total poultry isolates, the 617 isolates from Spain, USA, Luxemburg, France, The Netherlands, Portugal, Belgium, England, Germany, Italy, Poland and Romania were concentrated together, indicating a closely related population. The poultry isolates from Antarctica (n=3), Sweden (n=7) and Lithuania (n=2) were identical in PCoA plot because there were a few poultry isolates from those countries and they had identical alleles profiles. The poultry isolates from Finland (n=79) were distantly separated from other poultry isolates. The Finnish isolates also formed a separate cluster in the phylogenetic trees. This indicates distantly related *C. jejuni* populations in Finland compared to the *C. jejuni* populations in other European countries in this study. The PCoA showed similar allele frequency between Romania-Poland, Italy-Germany-England, Portugal-the Netherlands and USA-Luxemburg indicating similar *C. jejuni* populations between those countries.

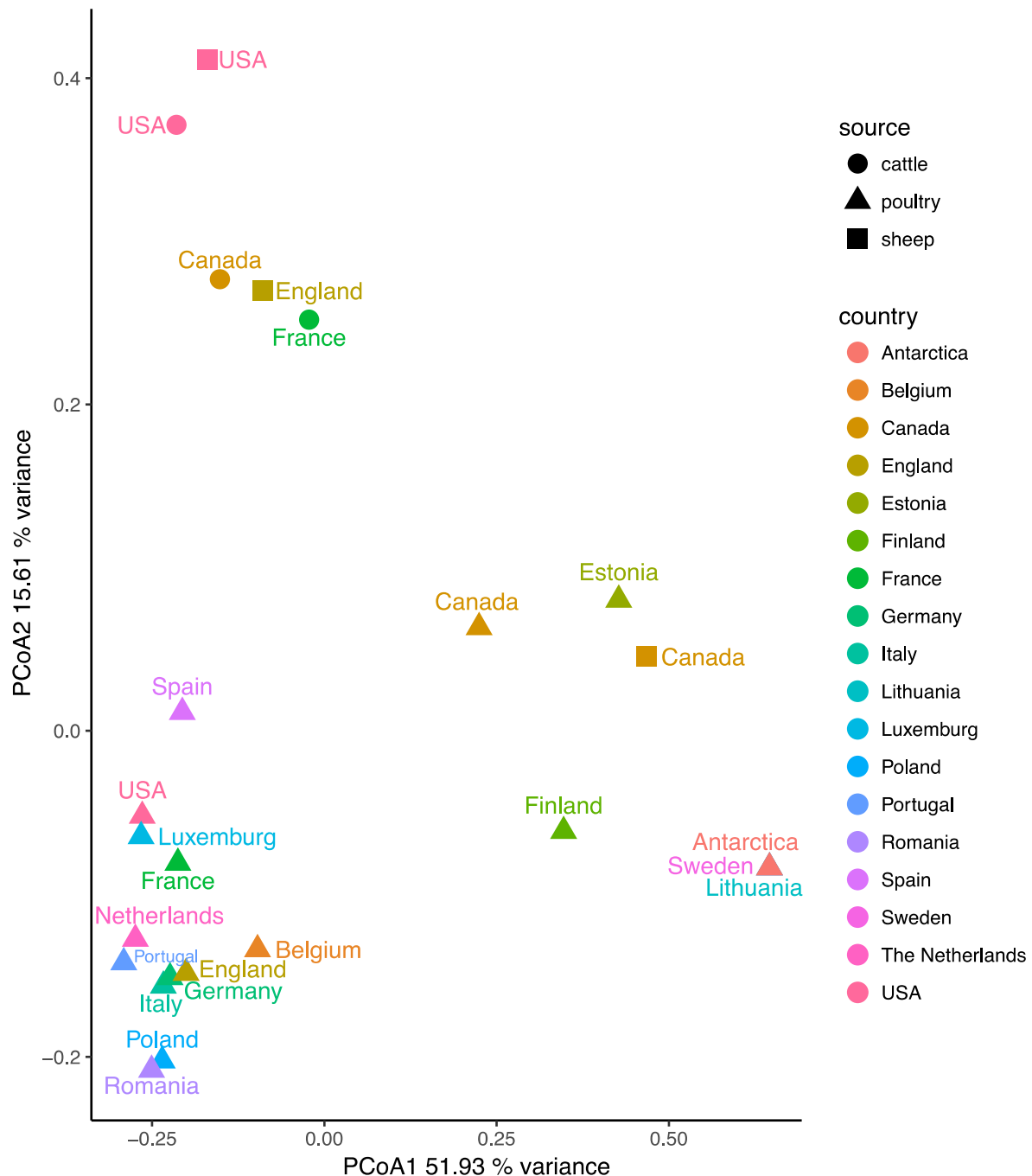
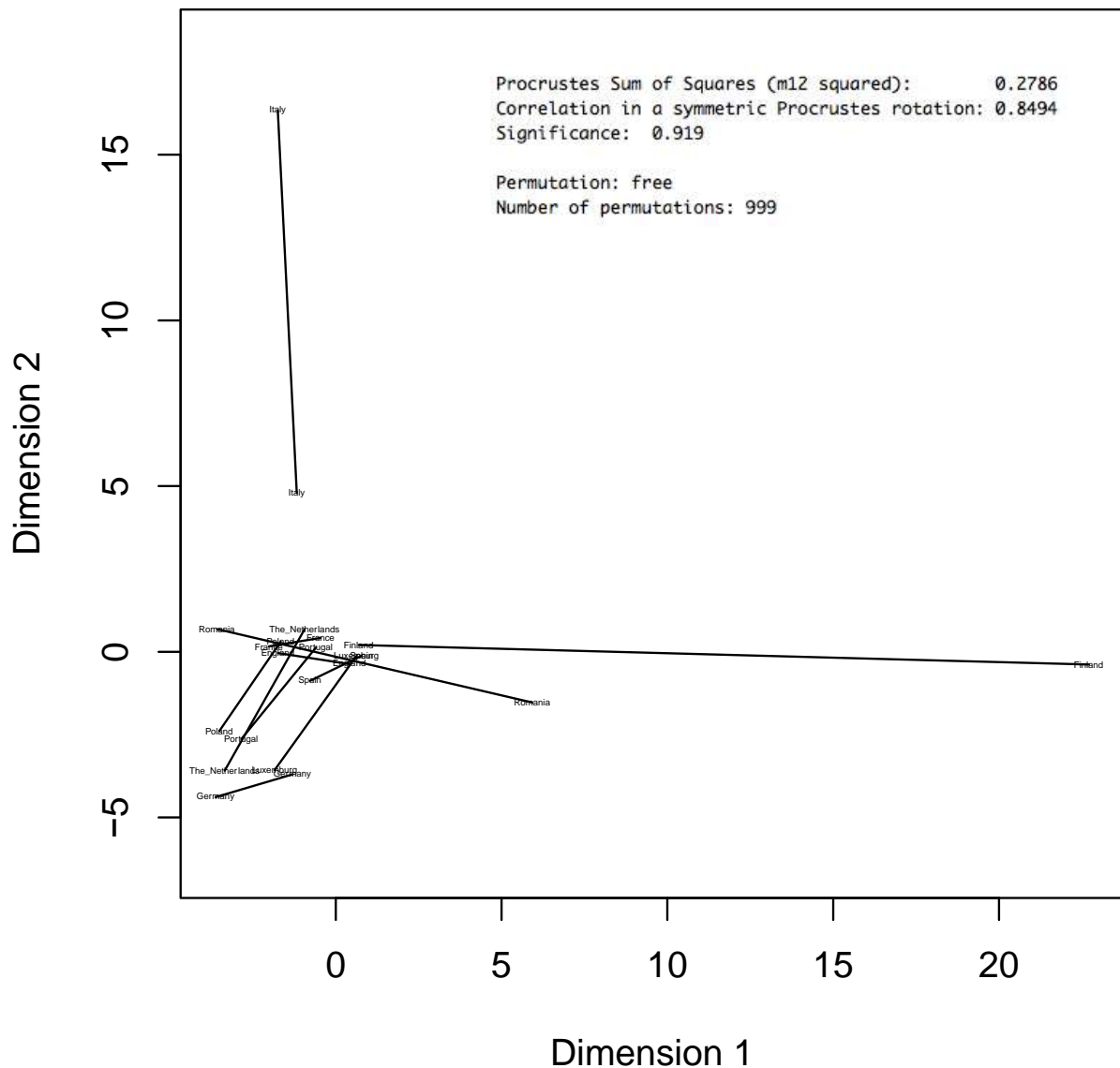


Figure 13: Principal coordinate analysis (PCoA) of allele frequency at country and host levels

Trade connectivity, antimicrobial use in livestock and allele frequency

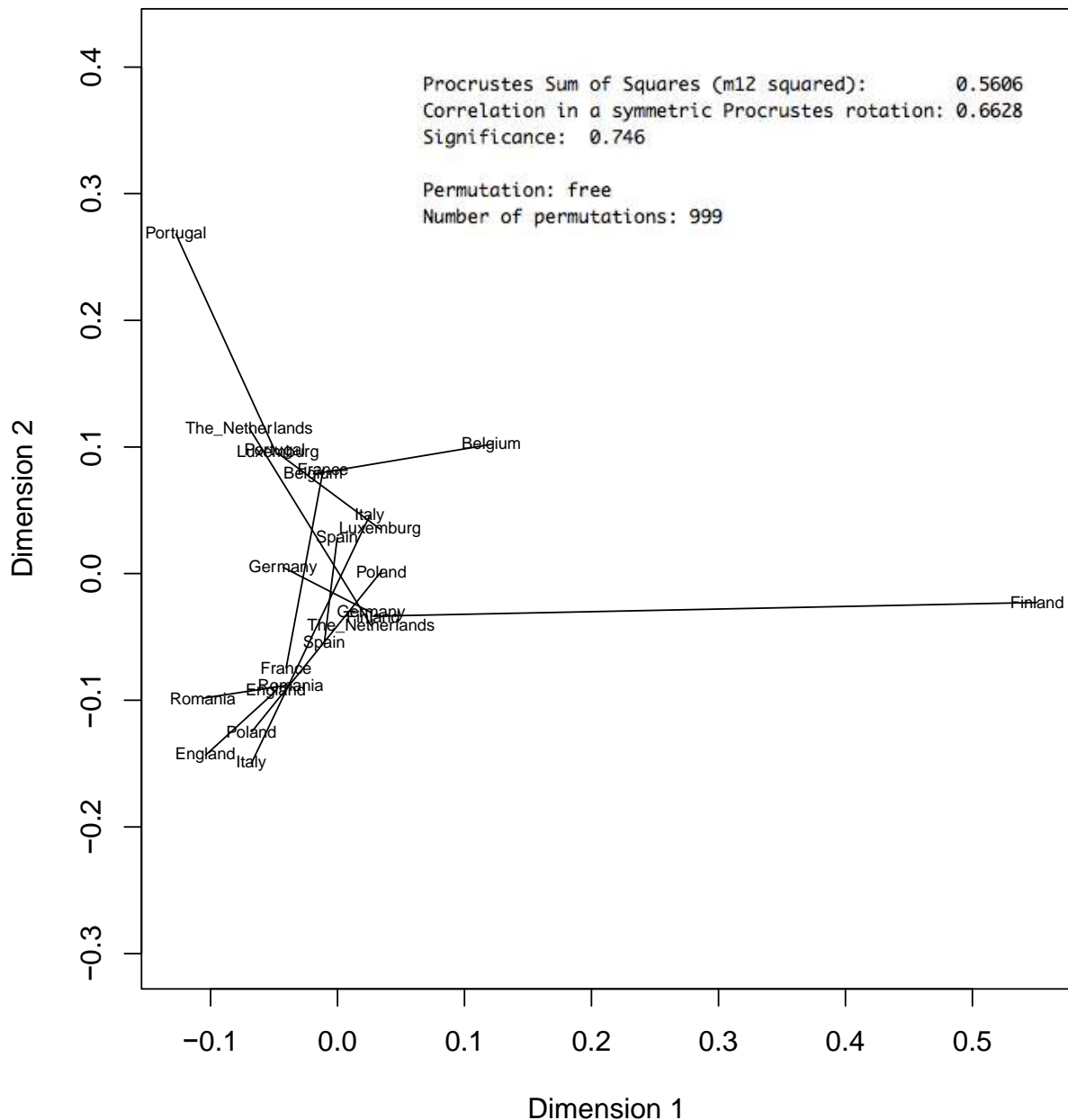
In order to determine whether trade connectivity and antimicrobial use in livestock are associated with allele frequency of *C. jejuni* or population of *C. jejuni*, the comparison of PCoA between trade connectivity against allele frequency (Figure 14) and antimicrobial use against allele frequency (Figure 15) were conducted using Procrustes analysis. The Procrustes results showed high correlation (0.8494 for trade connectivity and 0.6628 for antimicrobial usage). The strong correlation should be at a minimum 0.5. Nonetheless, the p-values (significance) were high (0.919 for trade connectivity and 0.746 for antimicrobial usage). The acceptable p-value should be lower than 0.05. Therefore, based on these trade

and antimicrobial use data, it is not significant enough to conclude the association between trade connectivity and population of *C. jejuni* and between antimicrobial use in livestock and *C. jejuni* population at country level.



Dimension 1 and 2 are PCoA dimensions. The shorter line indicates the more similarity of the two compared PCoA; trade pattern and allele frequency of each country.

Figure 14: The symmetric Procrustes correlation coefficients, p -value (significance) and plot between trade pattern and allele frequency of the 502 *C. jejuni* isolates



Dimension 1 and 2 are PCoA dimensions. The shorter line indicates the more similarity of the two compared PCoA; antimicrobial use and allele frequency of each country.

Figure 15: The symmetric Procrustes correlation coefficients, p-value (significance) and plot between antimicrobial use in livestock and allele frequency of the 502 *C. jejuni* isolates

5. Discussion and Conclusions

This study showed a high genomic diversity of FQ-R and FQ-S *C. jejuni* across the 10 largest poultry production countries in the EU. A high degree of agreement at internal nodes between the different typing methods was also observed. All trees clustered the 502 *C. jejuni* into four main clusters and had high number of isolates found in the same clusters. Comparing all nodes between trees, all trees showed low relative similarity score indicating that the degree of agreement at external nodes between trees was

low. It is currently not known which methods provide the most biologically correct clustering for internal and external nodes. Patterns of clustering according to the presence/absence of *gyrA* mutations and ST-types were observed. The isolates were not clustered according to the presence/absence of acquired resistance genes, country of origin and isolation year. The similar patterns of clustering were found in the cgMLST tree of additional published *C. jejuni* genomes. In addition, poultry isolates formed clusters unlike human isolates that spread through the tree. This indicates associated populations of *C. jejuni* in poultry origin and indicates high diversity of human *C. jejuni* isolates. The human isolates may originate from multiple host species. Isolation of bacteria in different host species can lead to population differentiation especially in the human *C. jejuni* isolates under this study.

The cgMLST trees of only FQ-R and FQ-S isolates also showed that isolates from the same country of origin were distributed into multiple clusters similarly to the trees combining FQ-R and FQ-S isolates. This result suggested that the presence/absence of *gyrA* mutations (T86I, T86A and T86V) for FQ did not influence the clustering of country of origin.

Using different genomic approaches either the gene by gene approach or the reference based approach and using different sets of genomic markers for constructing phylogeny ranging from whole genome MLST (1,643 loci), core genome MLST (1,343 loci), shared core genes (557 genes) and rMLST (52 loci), none of the phylogeny methods can cluster strains based on the country of origin. Phylogenetic signals can be weakened by genomic changes that occur within the reservoir host. This might make it difficult to attribute the country of origin based on phylogenetic analysis using either gene by gene or reference based approaches. A benchmark *C. jejuni* genomic dataset with known epidemiology or known phylogeny should be conducted for benchmarking of different phylogenetic methods. This would be critical for future outbreak detection of *C. jejuni* where we do not know which tool or method should be the most accurate one for outbreak investigation.

In agreement with previous studies (Kinana et al., 2006; Habib et al., 2009; Kovač et al., 2014; Cha et al., 2017) this study confirms an association between specific clonal types and the emergence of fluoroquinolone resistance. Based on the data available and the analysis performed in this study it cannot be concluded whether this relates to independent selection within countries or transmission of already resistant clones between countries. Also in agreement with previous studies (McTavish et al., 2008; Asakura et al., 2012), this study suggests phylogeographic structuring and proliferation of specific lineages in the different countries. In some countries the same types seem to predominate; however, the similarity between countries is not similar for FQ-R and FQ-S isolates and this requires further analysis. In *Campylobacter*, analyses of genomic data and molecular typing have been successful in identifying the reservoir host but not geographical origin (Pascoe et al., 2017). This is one of the main limitations of genomic analysis and cluster identification of *Campylobacter* genomes. Determination of alternative biogeographical markers in *Campylobacter* genomes and applying population genetics will contribute to the improvement of identification of geographical clustering.

Different genomic approaches have advantages and disadvantages in term of typeability. Only the conventional seven genes multilocus sequences and core genome multilocus sequences can give nomenclature of isolates as ST-type and cgST-type respectively. There are currently no allele profiles for wgMLST and rMLST. Nonetheless, the existing allele profiles for cgST-type are limited mainly to human *C. jejuni* isolates. Only 53% of the 502 poultry *C. jejuni* isolates in this study could be assigned to known cgST-types. SNP tree cannot give any nomenclature of isolates, however it can give SNP alignment that can be used for further analysis such as recombination analysis or temporal phylogenetic analysis. Similarly to core genes, it is a non-nomenclature approach. Nonetheless, it gives core gene sequences that could be used for identifying highly recombining genes as candidate markers of geographical attribution. In addition, other genes such as pan-genome and accessory genes could be candidate genes for further study on geographical origin. In addition, there was high level of recombination in this collection of *C. jejuni* (only 101 SNPs left after recombination detection of 32,392 identified SNPs).

The poultry isolates from Finland formed its own cluster in all phylogenetic trees, being different in the distribution of the top 20 ST-types and being placed distantly in the PCoA plot of allele frequency (Kovanen et al., 2014 and Llarena et al., 2017). This result indicates that the Finnish strains were from distantly related populations and different from the other *C. jejuni* populations from other European

countries included in this study. However, as the strains included in this study were from limited number of countries, the relationship between *C. jejuni* populations in other countries remains unknown, especially in Nordic countries.

The allele frequency indicated that the populations of *C. jejuni* isolates from poultry were more similar than the populations of *C. jejuni* isolates from other host species (Figure 13). The allele frequency also indicated the association of geographical origins of the poultry *C. jejuni* isolates. Analysis of allele frequency would be an alternative genomic approach for identification of geographical origin and source attribution of *Campylobacter* spp. Nevertheless, the country association from allele frequency need to be confirmed and supported by other analysis for instance the suggested analyses mentioned above and more detail on epidemiological data and trade data between those countries.

Trade connectivity and antimicrobial use in livestock were not significantly associated with allele frequency or populations of *C. jejuni*. Finland showed relatively lower association than other countries between trade and population of *C. jejuni*. This could be due to no influences of trading to the population structure of *C. jejuni* in Finland. Alternatively, the pattern of trading in general might not be able to link to population of *C. jejuni*. For instance when recombining *Campylobacter* transfer to a new animal host, it may acquire DNA from the resident population by horizontal gene transfer and become a different population than the original population in the exporting country.

Available antimicrobial usage data refer to sales of veterinary antimicrobials for all livestock species, not specifically poultry. The lack of precise data on the antimicrobial use in poultry from the poultry production countries and the lack of metadata, for instance the information of poultry farms, trade and the amount of antimicrobials used in each of antimicrobial classes from the country of isolation limit the ability to identify whether the emergence of antimicrobial resistance among *C. jejuni* is related to the selection through the antimicrobial use in these individual countries or to the transmission through countries.

In conclusion, the genomic diversity of FQ-R and FQ-S *C. jejuni* across the 10 largest poultry production countries in Europe was high. All phylogenetic approaches were concordant for internal nodes but not for external nodes. None of phylogenetic approaches could identify the country of origin of the isolates. There is a need to have a set of well characterized (benchmark) WGS data with known phylogeny for benchmarking different whole genome based phylogeny methods. Allele frequency could determine the association between countries of origin. Isolates were clustered based on *gyrA* mutations, ST-types and animal host species. No significant association between trade patterns, antimicrobial use in livestock and population of *C. jejuni* could be identified in this study, but trade and antimicrobial use data available were limited. Better trade and antimicrobial use data especially the antimicrobial use in poultry would be required to investigate this association.

References

- Aarestrup FM, 2015. The livestock reservoir for antimicrobial resistance: a personal view on changing patterns of risks, effects of interventions and the way forward. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370, 20140085.
- Asakura H, Brüggemann H, Sheppard SK, Ekawa T, Meyer TF, Yamamoto S and Igimi S, 2012. Molecular evidence for the thriving of *Campylobacter jejuni* ST-4526 in Japan. *PLoS One*, 7:e48394.
- Cha W, Mosci RE, Wengert SL, Venegas Vargas C, Rust SR, Bartlett PC, Grooms DL and Manning SD, 2017. Comparing the Genetic Diversity and Antimicrobial Resistance Profiles of *Campylobacter jejuni* Recovered from Cattle and Humans. *Frontiers in Microbiology*, 8:818.
- Clausen PTLC, Aarestrup FM and Lund O, 2017. Rapid and precise alignment of raw reads against redundant databases with KMA. Annual Danish Bioinformatics Conference, 24-25 August, 2017. Available online: <http://elixir-node.cbs.dtu.dk/wp-content/uploads/2017/09/Booklet-2017-small.pdf>

- Collignon P, Powers JH, Chiller TM, Aidara-Kane A and Aarestrup FM, 2009. World Health Organization ranking of antimicrobials according to their importance in human medicine: A critical step for developing risk management strategies for the use of antimicrobials in food production animals. *Clinical Infectious Diseases*, 49, 132-141.
- Cody AJ, Bray JE, Jolley KA, McCarthy ND and Maiden MCJ, 2017. A core genome multi-locus sequence typing scheme for stable, comparative analyses of *Campylobacter jejuni* and *C. coli* human disease isolates. *Journal of Clinical Microbiology*, pii: JCM.00080-17. doi: 10.1128/JCM.00080-17
- Denis M, Soumet C, Rivoal K, Ermel G, Blivet D, Salvat G and Colin P, 1999. Development of a m-PCR assay for simultaneous identification of *Campylobacter jejuni* and *C. coli*. *Letters in Applied Microbiology*, 29, 406-410.
- EFSA (European Food Safety Authority) and ECDC (European Centre for Disease Prevention and Control), 2016. The European Union summary report on antimicrobial resistance in zoonotic and indicator bacteria from humans, animals and food in 2014. *EFSA Journal* 2016;14(2):4380, 207 pp. doi:10.2903/j.efsa.2016.4380
- EMA (European Medicines Agency), 2016. European Surveillance of Veterinary Antimicrobial Consumption, 2016. Available online: http://www.ema.europa.eu/docs/en_GB/document_library/Report/2016/10/WC500214217.pdf
- Engberg J, Aarestrup FM, Taylor DE, Gerner-Smidt P and Nachamkin I, 2001. Quinolone and macrolide resistance in *Campylobacter jejuni* and *C. coli*: resistance mechanisms and trends in human isolates. *Emerging Infectious Diseases*, 7, 24-34.
- Habib I, Miller WG, Uyttendaele M, Houf K and De Zutter L, 2009. Clonal population structure and antimicrobial resistance of *Campylobacter jejuni* in chicken meat from Belgium. *Applied and Environmental Microbiology*, 75, 4264-4272.
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W and Gascuel O, 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology*, 59, 307-321.
- Hasman H, Saputra D, Sicheritz-Pontén T, Lund O, Svendsen CA, Frimodt-Møller N and Aarestrup FM, 2014. Rapid whole-genome sequencing for detection and characterization of microorganisms directly from clinical samples. *Journal of Clinical Microbiology*, 52, 139-146.
- Kaas RS, Leekitcharoenphon P, Aarestrup FM and Lund O, 2014. Solving the problem of comparing whole bacterial genomes across different sequencing platforms. *PLoS One*, 9: e104984. doi:10.1371/journal.pone.0104984
- Kinana AD, Cardinale E, Tall F, Bahsoun I, Sire JM, Garin B, Breurec S, Boye CS and Perrier-Gros-Claude JD, 2006. Genetic diversity and quinolone resistance in *Campylobacter jejuni* isolates from poultry in Senegal. *Applied and Environmental Microbiology*, 72, 3309-3313.
- Kovač J, Cadež N, Lušický M, Nielsen EM, Ocepek M, Raspor P and Možina SS, 2014. The evidence for clonal spreading of quinolone resistance with a particular clonal complex of *Campylobacter jejuni*. *Epidemiology and Infection*, 142, 2595-2603.
- Kovanen S, Kivistö R, Rossi M, Schott T, Kärkkäinen UM, Tuuminen T, Uksila J, Rautelin H and Hänninen ML, 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. *Journal of Clinical Microbiology*, 52, 4147-4154. doi:10.1128/JCM.01959-14
- Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Jelsbak L, Sicheritz-Pontén T, Usserya DW, Aarestrup FM and Lunda O, 2012. Multilocus sequence typing of total-genome-sequenced bacteria. *Journal of Clinical Microbiology*, 50, 1355-1361.

- Leekitcharoenphon P, Nielsen EM, Kaas RS, Lund O and Aarestrup FM, 2014. Evaluation of whole genome sequencing for outbreak detection of *Salmonella enterica*. PLoS One, 4;9(2):e87991. doi: 10.1371/journal.pone.0087991
- Legendre P and Gallagher ED, 2001. Ecologically meaningful transformations for ordination of species data. Oecologia, 129, 271–280.
- Li H and Durbin R, 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 25, 1754–1760. doi:10.1093/bioinformatics/btp324
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R and 1000 Genome Project Data Processing Subgroup, 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics, 25, 2078–2079. doi:10.1093/bioinformatics/btp352
- Llarena AK, Taboada E and Rossi M, 2017. Whole-Genome Sequencing in Epidemiology of *Campylobacter jejuni* Infections. Journal of Clinical Microbiology, 55, 1269-1275. doi: 10.1128/JCM.00017-17
- McTavish SM, Pope CE, Nicol C, Sexton K, French N and Carter PE, 2008. Wide geographical distribution of internationally rare *Campylobacter* clones within New Zealand. Epidemiology and Infection, 136, 1244-1252.
- Needleman SB and Wunsch CD, 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. Journal of Molecular Biology, 48, 443–453.
- Oksanen J, Guillaume Blanchet F, Friendly M, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH and Wagner H, 2016. vegan: Community Ecology Package. Available online: <https://rdrr.io/rforge/vegan/> <https://rdrr.io/rforge/vegan/>
- Pascoe B, Méric G, Yahara K, Wimalarathna H, Murray S, Hitchings MD, Sproston EL, Carrillo CD, Taboada EN, Cooper KK, Huynh S, Cody AJ, Jolley KA, Maiden MCJ, McCarthy ND, Didelot X, Parker CT and Sheppard SK, 2017. Local genes for local bacteria: Evidence of allopatry in the genomes of transatlantic *Campylobacter* populations. Molecular Ecology, 26, 4497–4508.
- Ruppitsch W, Pietzka A, Prior K, Bletz S, Fernandez HL, Allerberger F, Harmsen D and Mellmann A, 2015. Defining and Evaluating a Core Genome Multilocus Sequence Typing Scheme for Whole-Genome Sequence-Based Typing of *Listeria monocytogenes*. Journal of Clinical Microbiology, 53, 2869-2676.
- Sheppard SK and Maiden MC, 2015. The evolution of *Campylobacter jejuni* and *Campylobacter coli*. Cold Spring Harbor Perspectives in Biology, 7, a018119.
- Sheppard SK, Colles F, Richardson J, Cody AJ, Elson R, Lawson A, Brick G, Meldrum R, Little CL, Owen RJ, Maiden MC and McCarthy ND, 2010. Host association of *Campylobacter* genotypes transcends geographic variation. Applied and Environmental Microbiology, 76, 5269-5277.
- WHO (World Health Organization), 2014. Antimicrobial resistance: global report on surveillance. Available online: <http://www.who.int/drugresistance/documents/surveillancereport/en>
- Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM and Larsen MV, 2012. Identification of acquired antimicrobial resistance genes. Journal of Antimicrobial Chemotherapy, 67, 2640–2644, doi:10.1093/jac/dks261

Abbreviations

AMU	antimicrobial usage
BC	Bray-Curtis
BWA	Burrows-Wheeler Aligner
CCDA	<i>Campylobacter</i> Charcoal Differential Agar
CGE	Center for Genomic Epidemiology
cgMLST	core genome MLST
DTU Food	Technical University of Denmark, National Food Institute
ENA	European Nucleotide Archive
ESVAC	European Surveillance of Veterinary Antimicrobial Consumption
EUCAST	European Committee on Antimicrobial Susceptibility Testing
FQ-R	fluoroquinolone-resistant
FQ-S	fluoroquinolone-susceptible
MIC	minimal inhibitory concentration
MLST	multilocus sequence typing
MS	Member State
MTA	material transfer agreement
ND	nucleotide difference (tree)
NRLs	National Reference Laboratories
PCoA	principal coordinate analysis
PCR	polymerase chain reaction
PCU	population correction unit
rMLST	ribosomal MLST
SNP	single nucleotide polymorphism
ST	Sequence typing
UK	The United Kingdom
wgMLST	whole genome MLST
WGS	whole genome sequencing

Appendix A – Questionnaire to participants on selection and collection of isolates and related metadata

In the initial phase of the EFSA *C. jejuni* 2016 project, to get an overview of the available strains and data, we would like you to provide us with the information listed in the survey below.

If you have any questions or feedback for the survey, please contact Susanne Karlsmose Pedersen (suska@food.dtu.dk), at the Technical University of Denmark.

Note: An asterisk (*) indicates a question that requires an answer.

Further details of the project:

DTU-Food has recently signed a grant agreement with EFSA to carry out an exploratory study which might give further information regarding the genomic diversity of quinolone resistant and quinolone sensitive *C. jejuni* in broilers, and about the possible reasons for the high occurrence of quinolone resistance among these bacteria.

We therefore plan to collect isolates from different member states, perform whole genome sequencing and see whether we might identify clusters or clones associated with countries, trade and/or quinolone use.

Our focus will be on isolates from which data were provided for the 2014 EFSA report. We hope that you would be interested to support the project with data and strains.

Unfortunately, we cannot cover the shipping expenses. We anticipate jointly with you; the data providers, to publish the data/results in scientific journals after handing in the results – EFSA report. Please note that EFSA will have the possibility to use and publish data and results from the project indefinitely and without the need to notify us and data providers. The project has officially started on 18 August 2016, and is expected to be completed by 18 November 2017 (duration: 15 months), when DTU Food will provide the final report of the study, including details on the data collected and results of the analysis. We would like to stress that it will not be acceptable for us to delay the delivery of the final report of the study or its subsequent publication as an EFSA technical report due to potential national action/conflicting interests of the collaborating partners from the different countries. All produced genomes will be shared with European Nucleotide Archive (ENA).

* 1. About you: Your name?

* 2. About you: Your institute?

* 3. About you: Your country?

* 4. Do you have the *C. jejuni* isolates (broilers) available from which data were provided for 2014 to EFSA?

Yes, all isolates are available

Yes, some isolates are available

No, no isolates are available

Please indicate the number of available *C. jejuni* isolates from broilers:

* 5. Have any of the isolates already been whole genome sequenced (WGS'ed)?

Yes, all have been WGS'ed

Yes, some have been WGS'ed

No, none have been WGS'ed

Please indicate how many *C. jejuni* isolates from broilers have been WGS'ed:

* 6. If some of the *C. jejuni* isolates from broilers have not yet been WGS'ed, are there plans to do so?

Yes

No

Please add any comments:

* 7. Would you be willing to participate in a collaborative study on this?

Yes

No

Please add any comments:

* 8. Do you have further information on trace and origin of the isolates (e.g. farm of origin, trade connections in the farm of origin, details on antimicrobial usage in the farms of origin, etc.)?

Annex A – Supplementary Table

Annex A - Excel file: 'Annex A - Supplementary tables GenCamp.xlsx' - Supplementary Table with data.

Annex A can be found in the online version of this output ('Supporting information' section:

<http://onlinelibrary.wiley.com/doi/10.2903/sp.efsa.2018.EN-1398>