

## Supplementary Information for

### Host ecology regulates interspecies recombination in bacteria of the genus *Campylobacter*

Evangelos Mourkas<sup>1</sup>, Koji Yahara<sup>2</sup>, Sion C. Bayliss<sup>1</sup>, Jessica K. Calland<sup>1</sup>, Håkan Johansson<sup>3</sup>, Leonardos Mageiros<sup>1</sup>, Zilia Y. Muñoz-Ramirez<sup>4</sup>, Grant Futcher<sup>1</sup>, Guillaume Méric<sup>1#</sup>, Matthew D. Hitchings<sup>5</sup>, Santiago Sandoval-Motta<sup>4</sup>, Javier Torres<sup>4</sup>, Keith A. Jolley<sup>6</sup>, Martin C. J. Maiden<sup>6</sup>, Patrik Ellström<sup>7</sup>, Jonas Waldenström<sup>3</sup>, Ben Pascoe<sup>1</sup>, Samuel K. Sheppard<sup>1,6\*</sup>

<sup>1</sup>The Milner Centre for Evolution, Department of Biology and Biochemistry, University of Bath, Bath BA2 7AY, UK; <sup>2</sup>Antimicrobial Resistance Research Center, National Institute of Infectious Diseases, Tokyo, 162-8640, Japan; <sup>3</sup>Centre for Ecology and Evolution in Microbial Model Systems, Linnaeus University, Kalmar, 391 82, Sweden; <sup>4</sup>Unidad de Investigacion en Enfermedades Infecciosas, UMAE Pediatria, Instituto Mexicano del Seguro Social; <sup>5</sup>Swansea University Medical School, Swansea University, Singleton Park, Swansea, SA2 8PP, UK; <sup>6</sup>Department of Zoology, University of Oxford, South Parks Road, Oxford, OX1 3PS, UK; <sup>7</sup>Department of Medical Sciences, Zoonosis Science Centre, Uppsala University, Uppsala, Sweden;

\*Corresponding authors: Samuel K. Sheppard; [s.k.sheppard@bath.ac.uk](mailto:s.k.sheppard@bath.ac.uk)

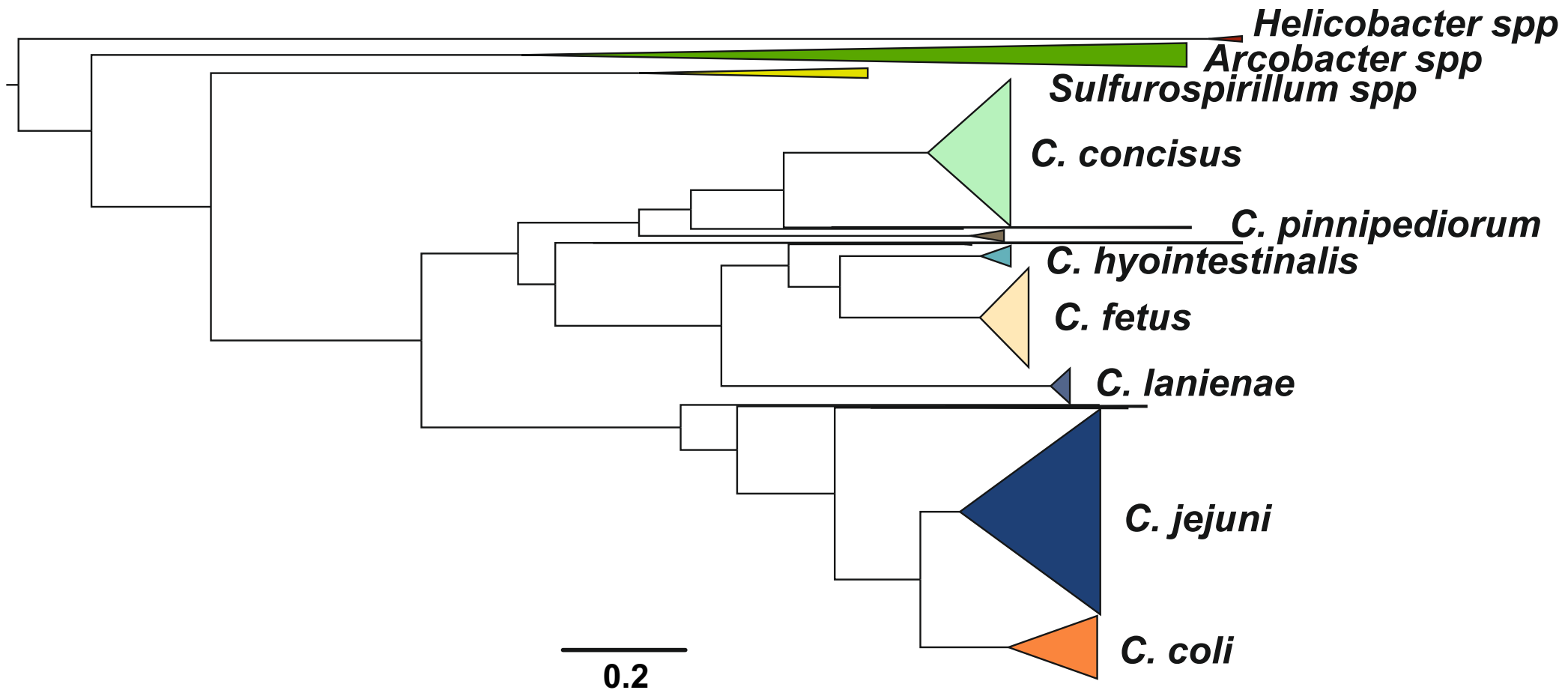
#Present address: Cambridge Baker Systems Genomics Initiative, Baker Heart and Diabetes Institute, 75 Commercial Rd, Melbourne 3004, Victoria, Australia; Department of Infectious Diseases, Central Clinical School, Monash University, Melbourne, Victoria 3004, Australia.

### This PDF includes

Figures legends and Figures S1 to S12  
Legends for Tables S1 to S5

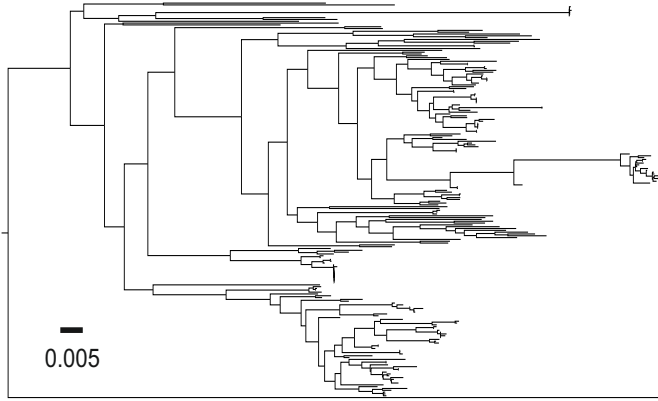
**Supplementary Figure 1. Population structure of the *Campylobacteraceae* family.**

Phylogenetic tree of 506 isolates that belong to the *Campylobacteraceae* family with *Helicobacter pylori* used as an outgroup. Different colors correspond to main species with number of isolates greater than three. The Tree was reconstructed using a gene-by-gene concatenated alignment of 799 core genes shared by >95% by all isolates and an approximation of the maximum-likelihood algorithm (ML) implemented in RAxML. The scale bar indicates the estimated number of substitutions per site.

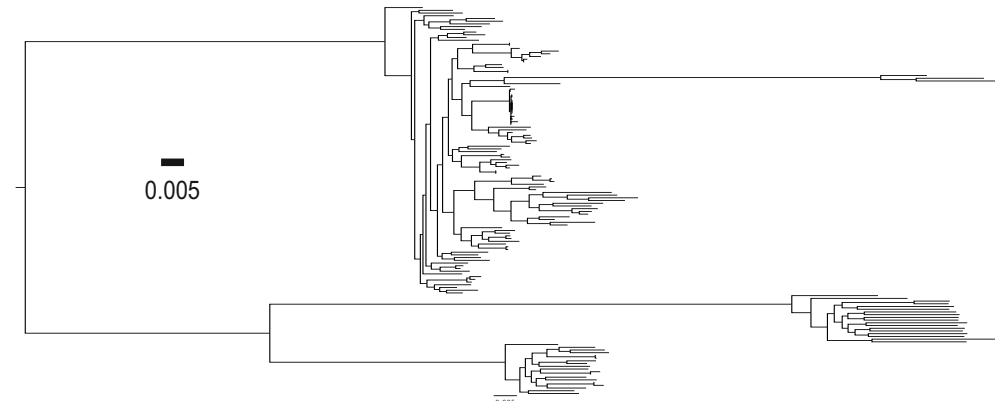


**Supplementary Figure 2.** Core genome species trees. Single-species trees for nine *Campylobacter* species with >4 isolates demonstrating the diversity for among species. The scale bars indicate the estimated number of substitutions per site. (\*) The scale for the tree corresponding to *C. hepaticus* is 10 times smaller than the rest.

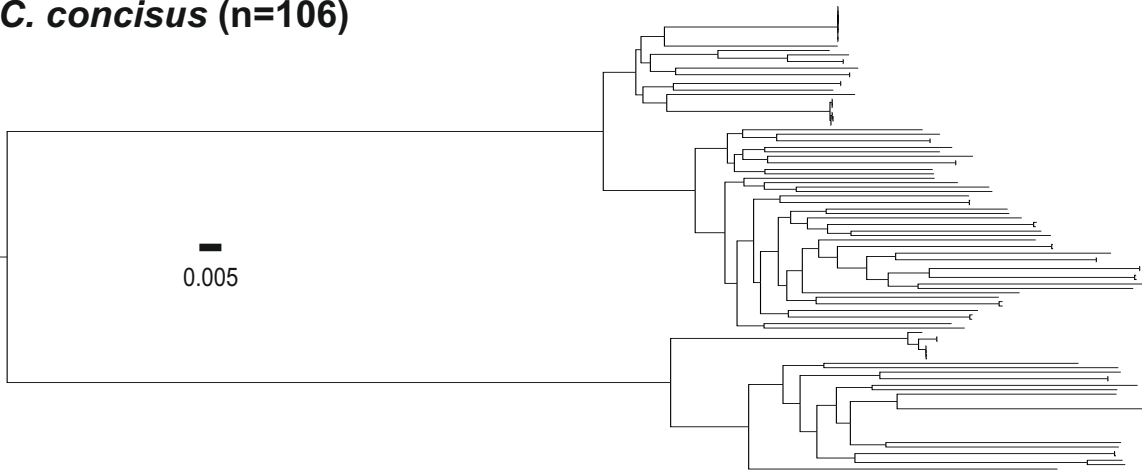
***C. jejuni* (n=218)**



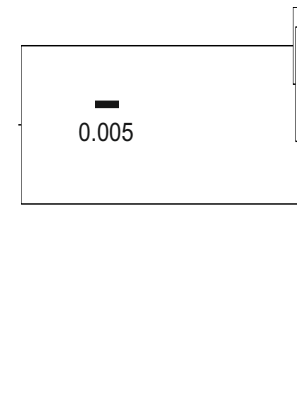
***C. coli* (n=143)**



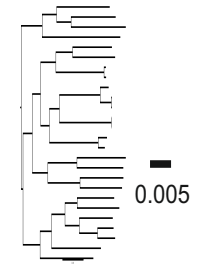
***C. concisus* (n=106)**



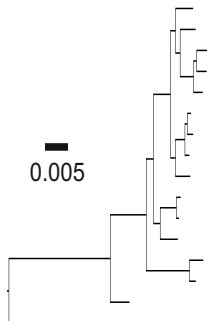
***C. fetus* (n=52)**



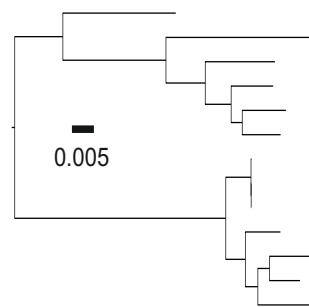
***C. lari* (n=26)**



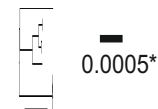
***C. hyointestinalis* (n=16)**



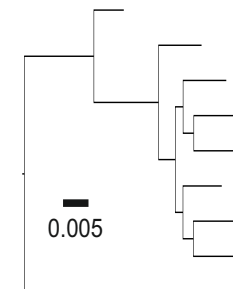
***C. lari* (n=13)**



***C. hepaticus* (n=10)\***



***C. pinnipediorum* (n=9)**



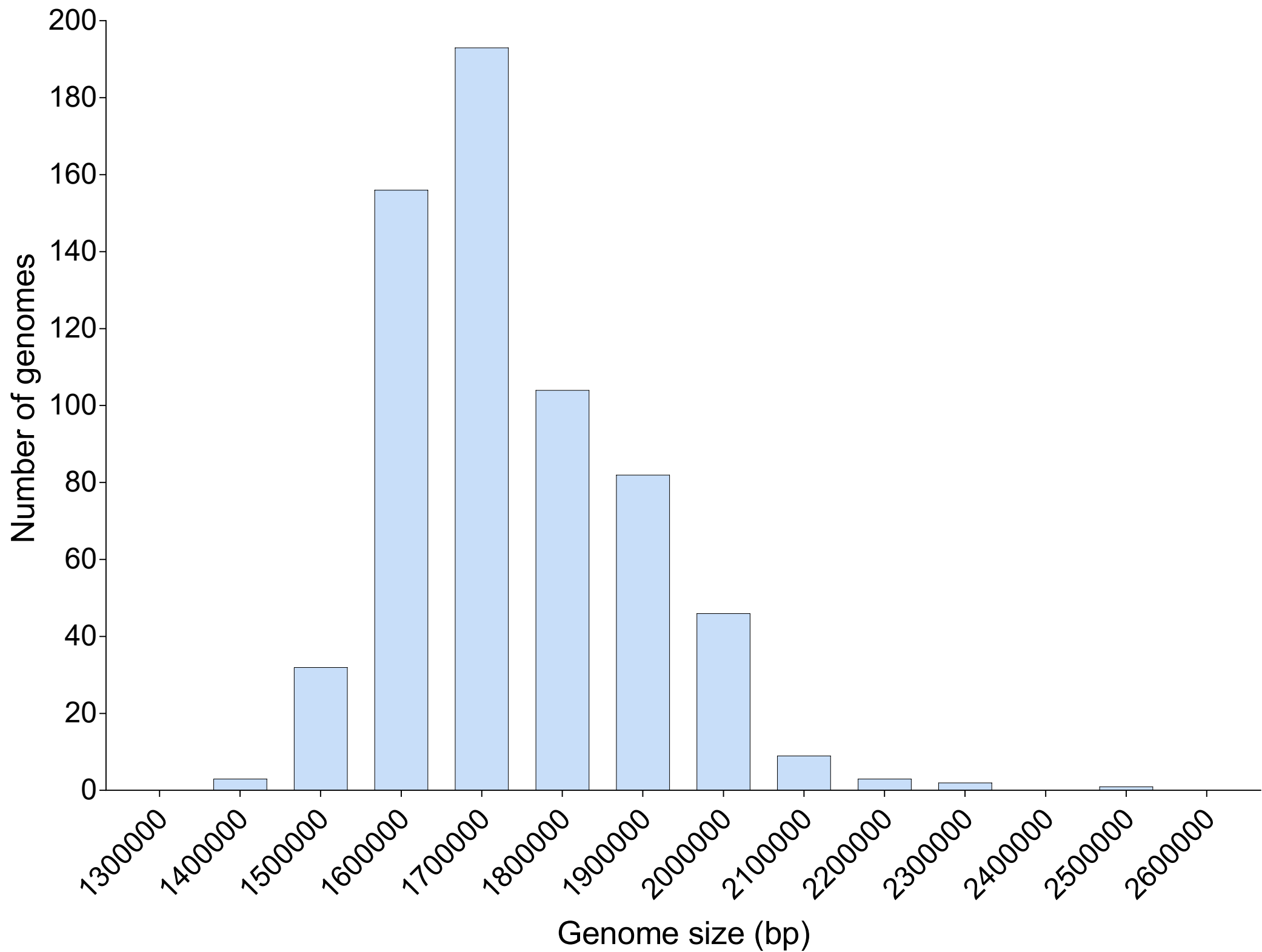
**Supplementary Figure 3. Overview of host-associations of *Campylobacter* species.**

Abundance and diversity of 631 *Campylobacter* isolates in each host and environment. Different colours correspond to main species with number of isolates  $\geq$  three. The number of isolates is shown on the y axis while the various isolation sources on the x axis.

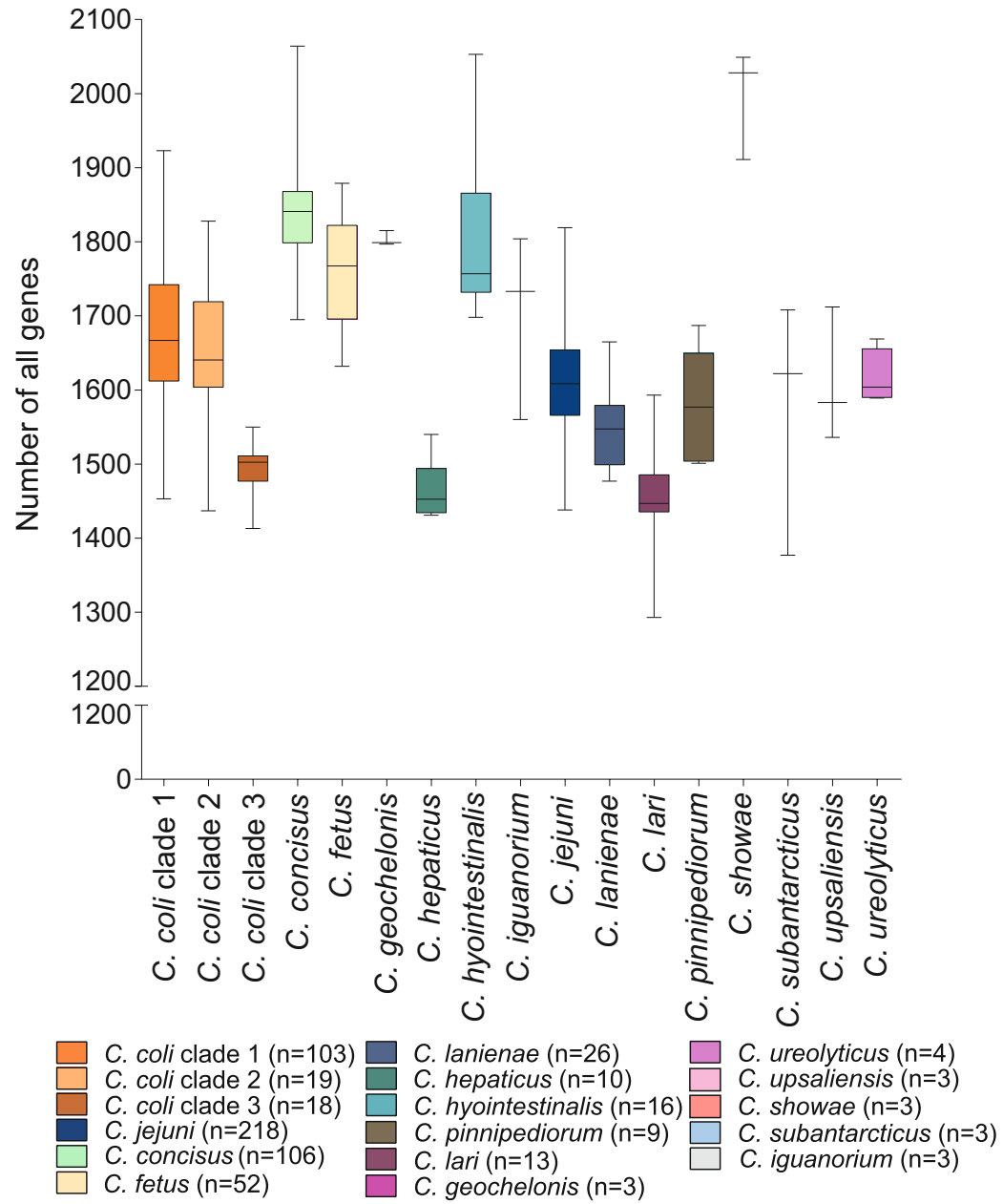


**Supplementary Figure 4. Genome size variation of the *Campylobacter* genus.** The frequency distribution of the genome size of all *Campylobacter* genomes used in this study is shown as a histogram. The number of genomes is shown on the y axis while the genome size (in bp) on the x axis.





**Supplementary Figure 5. Gene variation in the genus *Campylobacter*.** Overall distribution of the total number of genes per isolate for each *Campylobacter* species (where  $n \geq 3$  isolates). The number of genes is shown as boxplots (min to max).

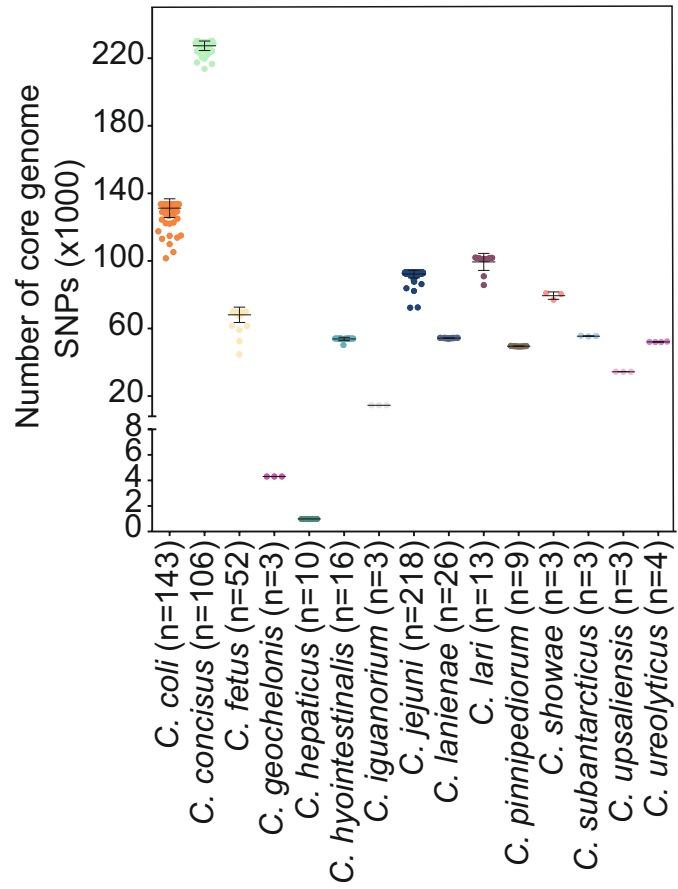
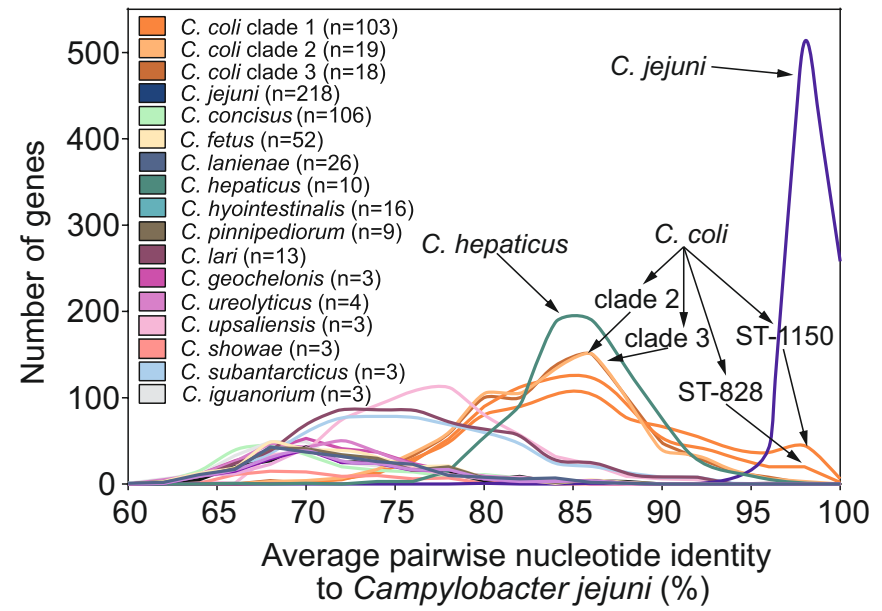


**Supplementary Figure 6. Accessory gene function in all main *Campylobacter* species.** The different gene functions are depicted on the y axis, while the number of shared accessory genes on the x axis. Different colours corresponding to different *Campylobacter* species.



**Supplementary Figure 7. Core genome allelic variation and the effect of recombination.**

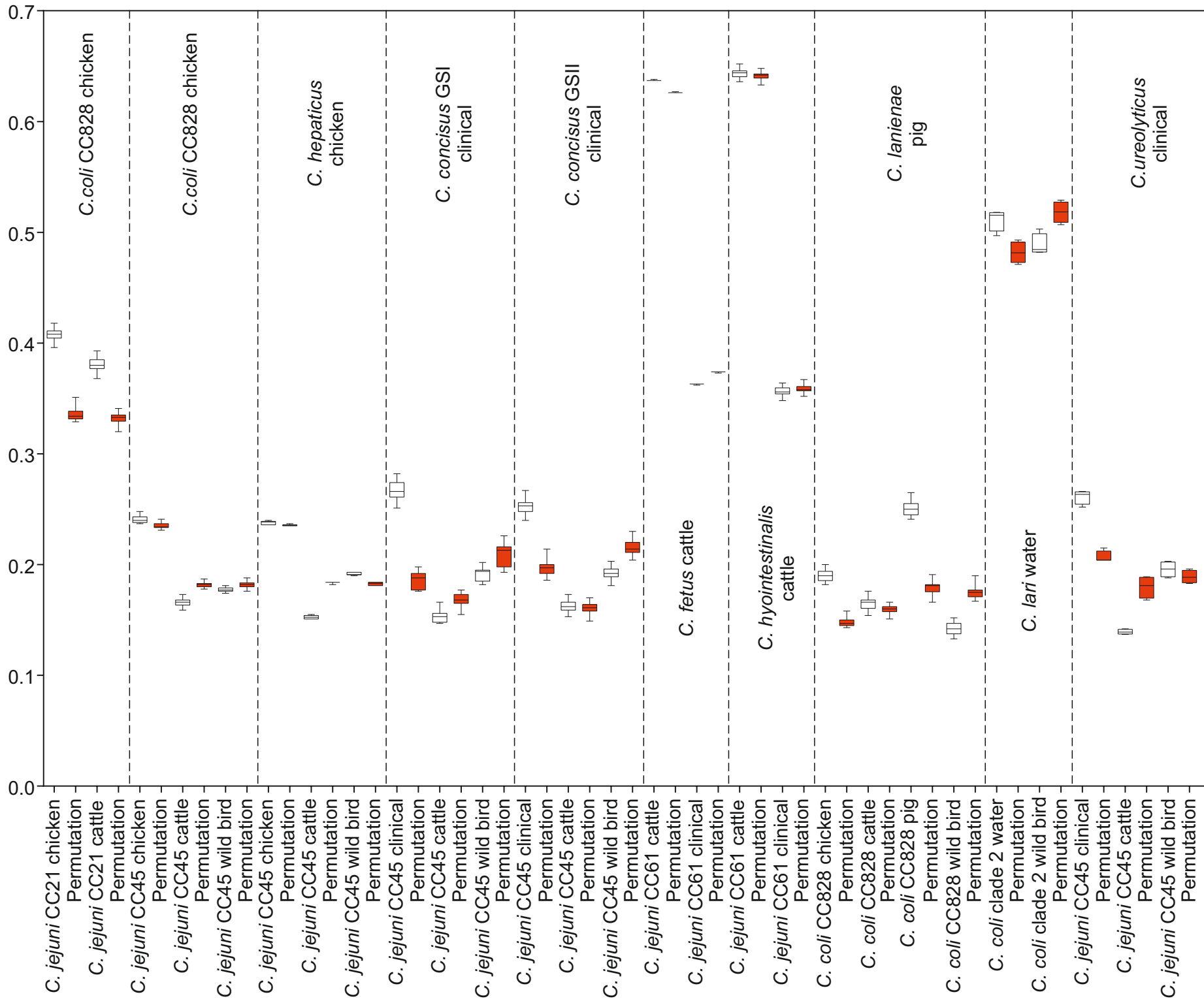
**a**, Number of SNPs per genome of the main *Campylobacter* species (where  $n \geq 3$  isolates) in the core genome alignment. The horizontal line in each plot represents the mean value while the upper and lower lines the standard deviation. **b**, Average nucleotide identity for pairwise comparisons of 820 core genes for 605 genomes of 15 main *Campylobacter* species. Different colours corresponding to different *Campylobacter* species.

**a****b**

**Supplementary Figure 8.** Probability of the recipient genomes sharing DNA with each donor groups is illustrated as box whiskers (white) for every donor-recipient comparison for all 10 pairs that supported our hypothesis. The analysis where the host data were randomized across all isolates is illustrated as box whiskers (red). The probability of copying DNA from a donor to a recipient genome is shown on the y axis. The midline in the box whiskers indicates the mean and the error bars the standard deviation.



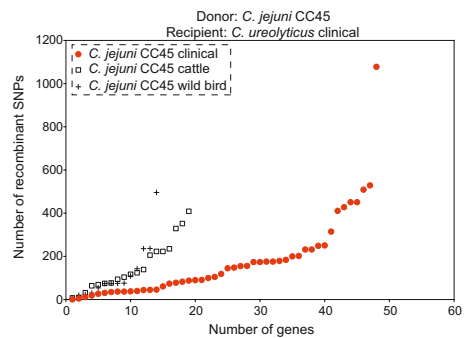
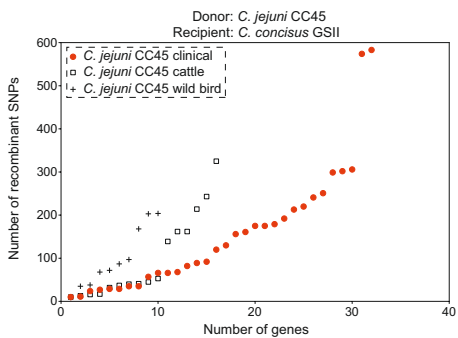
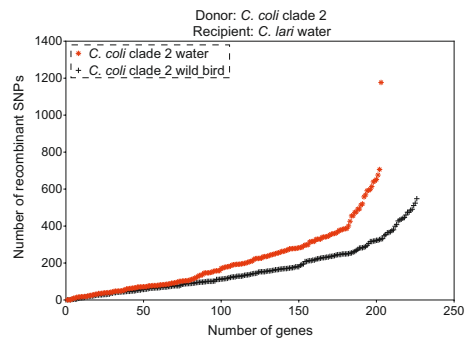
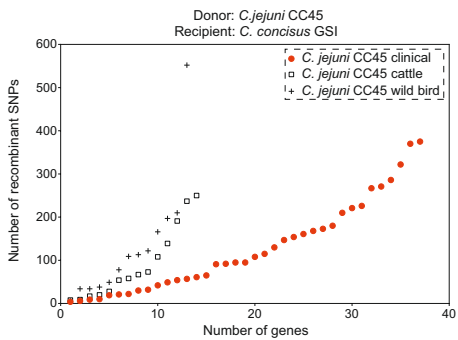
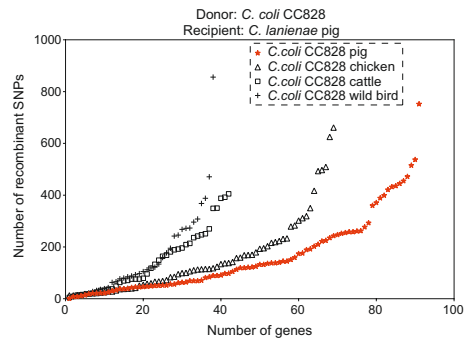
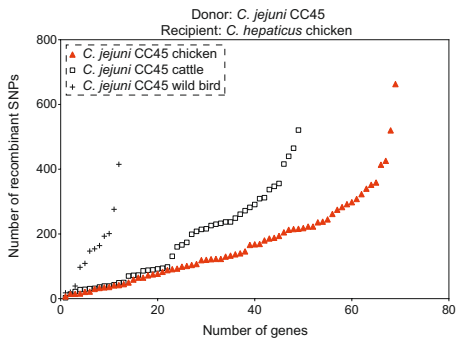
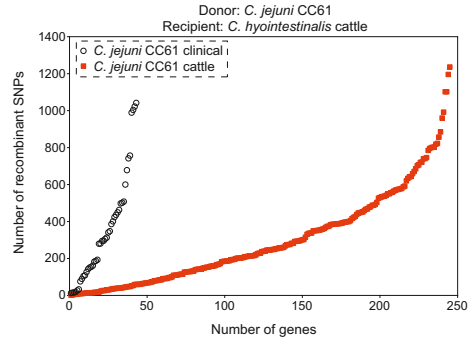
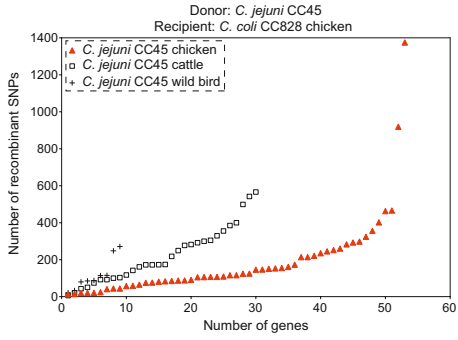
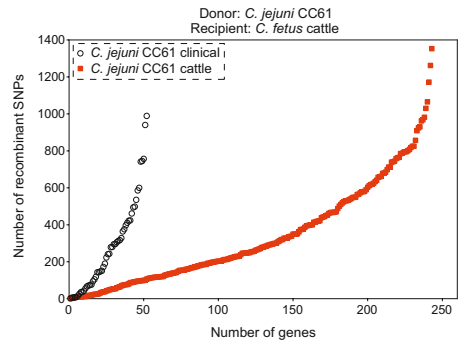
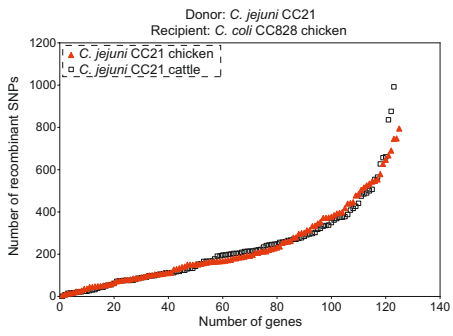
Probability of copying DNA from donor to recipient genome



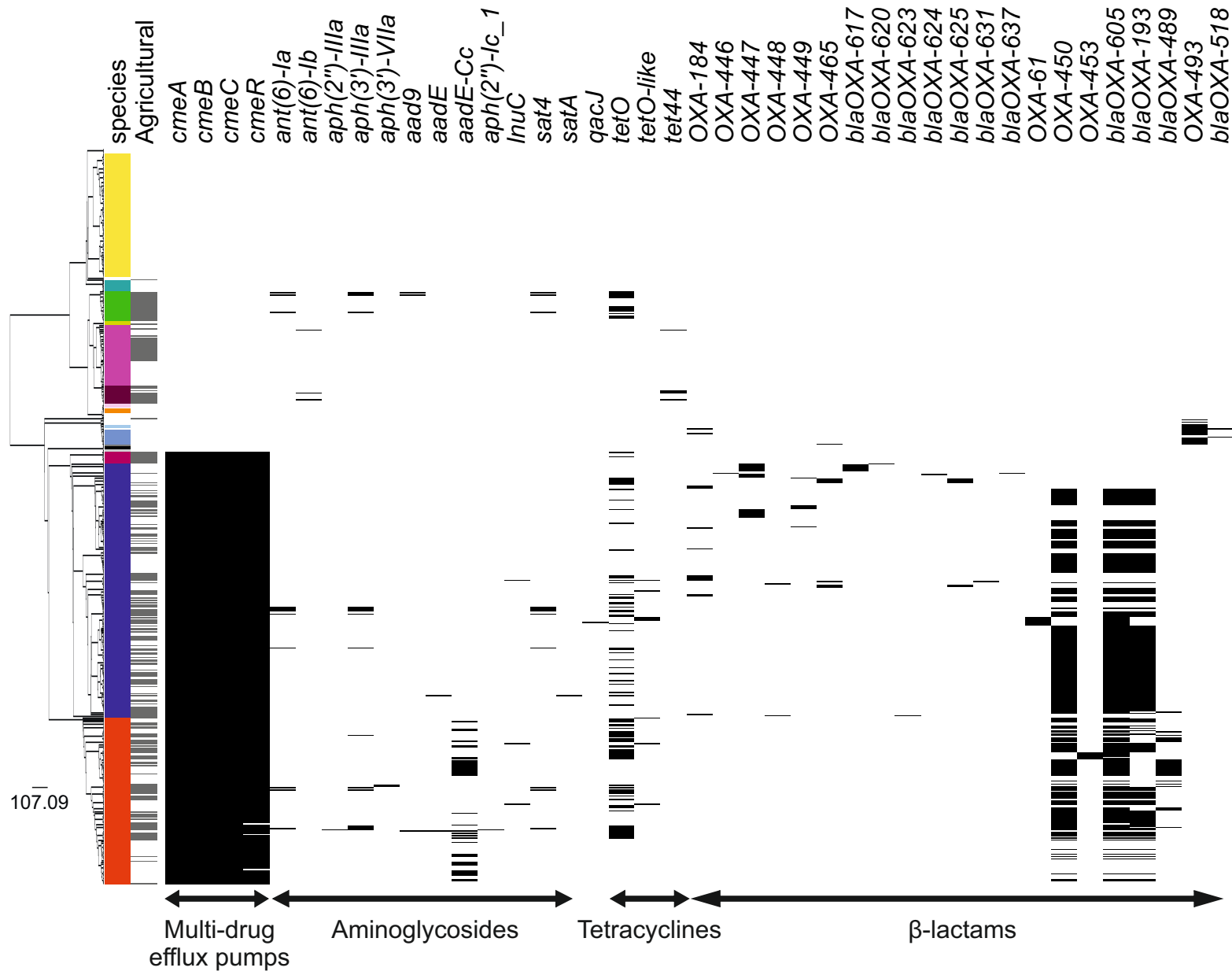
**Supplementary Figure 9. Genome position of genes containing recombining SNPs.** Genes and their corresponding number of recombining SNPs, inferred by Chromosome Painting analysis for all 10 species comparisons, and mapped to the NCTC11168 reference genome. Genes from within-host (red) and between-host (white) pair comparisons are shown for each comparison. Donors are isolates from chicken (triangle), cattle (square), wild bird (cross), pig (star), clinical (circle) and water (snowflake) samples. The dashed line indicates the 95<sup>th</sup> percentile for every individual group comparison.



**Supplementary Figure 10.** Genes ranked in ascending order of the number of recombining SNPs they contain as inferred by Chromosome Painting analysis for all ten species comparisons. Genes from within-host (red) and between-host (white) are shown for each comparison. Donors are isolates from chicken (triangle), cattle (square), wild bird (cross), pig (star), clinical (circle) and water (snowflake) samples.



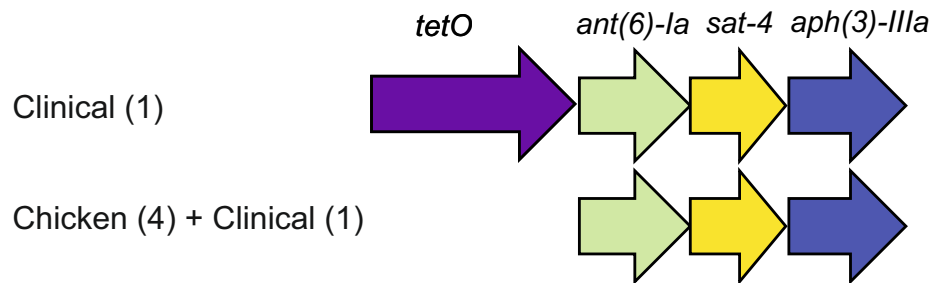
**Supplementary Figure 11.** Presence of antimicrobial resistance genes in the *Campylobacter* genus. The phylogenetic tree was reconstructed using a gene-by-gene concatenated alignment of 820 core and soft-core genes and an approximation of the maximum-likelihood algorithm (ML) implemented in RAxML. The designated colour scheme was used for each species in the first column. The second column indicates whether the strain is isolated from an agricultural animal (grey). Remaining columns indicate presence of AMR genes (black). The scale represents the number of substitutions per site.



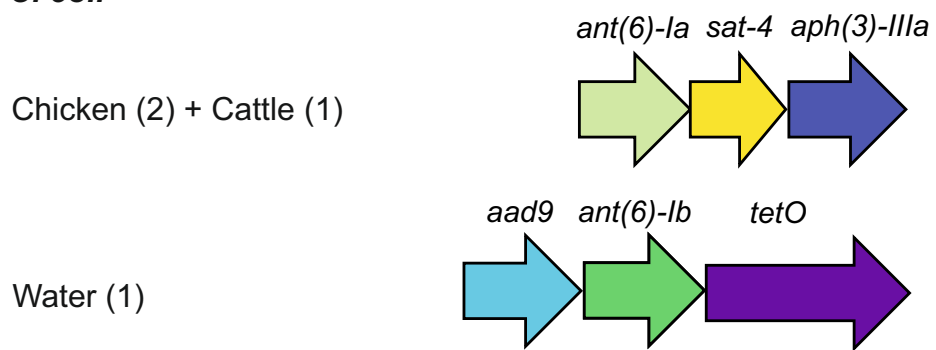
**Supplementary Figure 12. Genetic organization of AMR genes in *Campylobacter*.** The presence of each AMR gene, highlighted in different colours, is shown for representative genomes from *C. jejuni*, *C. coli*, *C. lanienae*, *C. hyointestinalis* and *C. fetus subspecies fetus* sampled from different agricultural animals. The number of isolate genomes containing each genomic arrangement is indicated in parenthesis.



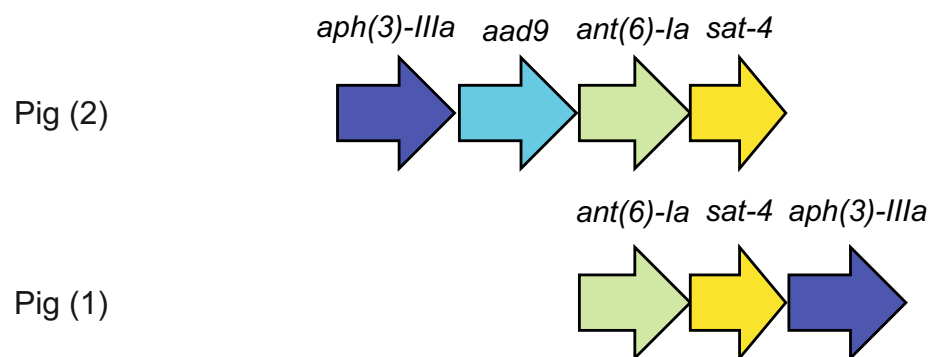
***C. jejuni***



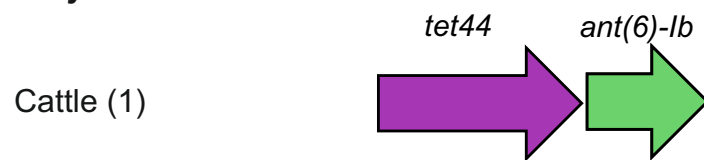
***C. coli***



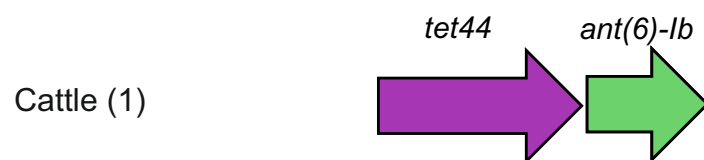
***C. lariena***



***C. hyointestinalis***



***C. fetus subs fetus***



**Table S1. Isolate information about the genomes used in this study.**

**Table S2. Within-host highly (>95<sup>th</sup> percentile) recombining genes.**

**Table S3. Recombination parameters as calculated by ClonalFrameML.**

**Table S4. Quantifying recombination between co-habiting species using ChromoPainter.**

**Table S5. Genes involved in interspecies recombination in 10 species comparisons.**