AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Complete Genome Sequences of Multidrug-Resistant *Campylobacter coli* Strains YH501, YH503, and YH504, from Retail Chicken

Yiping He,[a] Sue Reed,[a] Xianghe Yan,[b] Dandan Zhang,[c] Terence Strobaugh,[a] Joseph Capobianco,[a] Andrew Gehring[a]

[a]Molecular Characterization of Foodborne Pathogens Research Unit, Eastern Regional Research Center, Agricultural Research Service, U.S. Department of Agriculture, Wyndmoor, Pennsylvania, USA

[b]Environmental Microbial and Food Safety Laboratory, Beltsville Agricultural Research Center, Agricultural Research Service, U.S. Department of Agriculture, Beltsville, Maryland, USA

[c]Molecular Epidemiology, Inc., Seattle, Washington, USA

ABSTRACT *Campylobacter coli* is an important foodborne pathogen that can cause inflammation of the intestine and diarrhea in humans. The complete genomes, including megaplasmids, of *C. coli* strains YH501, YH503, and YH504 from retail chicken were sequenced and *de novo* assembled. Whole-genome analysis revealed a number of virulence and antibiotic resistance genes, suggesting significant potential for these poultry-originating isolates to cause human disease.

Campylobacter coli (taxonomy identification number 195) strains with multidrug resistance have been frequently identified in retail chicken (1). To understand the occurrence, transmission, and pathogenic properties of this microorganism, three antimicrobial-resistant *C. coli* strains were isolated from retail chicken and characterized. Here, the complete genomes and brief annotations of *C. coli* strains YH501, YH503, and YH504 are presented.

C. coli strains YH501, YH503, and YH504 were isolated from retail chicken using a previously described passive filtration method (2). Species identification was confirmed via multiplex quantitative PCR (qPCR) targeting the *hipO* and *cdtA* genes (3). The Genomic-tip 100/G kit (Qiagen, Valencia, CA) was used to extract genomic DNA (gDNA) from cultures grown microaerobically in Mueller-Hinton broth at 42°C for 24 h. The whole genome was sequenced using a Pacific Biosciences (PacBio) single-molecule real-time (SMRT) instrument with a large-insert (>10-kb) library constructed with the SMRTbell Express template preparation kit v. 2.0 and a Megaruptor for DNA shearing without size selection (PacBio, Menlo Park, CA) and a MiSeq instrument with a Nextera XT library preparation kit for 2 × 300-bp paired-end reads, including the Nextera XT kit for transposome-based DNA shearing without size selection (Illumina, San Diego, CA). For strain YH501, the Illumina reads were quality assessed by FastQC, filtered with NxTrim, and assembled using SPAdes v. 3.7.1 (4). For strains YH503 and YH504, the PacBio reads were trimmed and assembled using Canu v. 1.3 (5). The resulting contigs underwent removal of overlapping ends and generation of circular molecules using Circlator v. 1.5.5 (6) and were polished with the raw Illumina reads using Pilon v. 1.22 (7). Finally, the draft genomes were validated by mapping reads back using CLC Workbench v. 9.5 (Qiagen Bioinformatics, Redwood City, CA). The origin of the chromosome was manually identified and rotated to the *dnaA* sequence. The plasmid origin was determined via homology to known *Campylobacter* plasmid sequences. All software was used with default parameters unless otherwise noted. Table 1 summarizes the sequence data and genome statistics of the *C. coli* isolates.

Rapid Annotation using Subsystems Technology (RAST) (http://rast.nmpdr.org) (8) predicted multiple virulence and antimicrobial resistance genes in each strain. Over 30 genes were associated with cell motility and chemotaxis. Both plasmids pCOS503 and pCOS504 contain a 17-kb

Editor Frank J. Stewart, Montana State University

Address correspondence to Yiping He, yiping.he@usda.gov.

The authors declare no conflict of interest.

Received 15 March 2022
Accepted 26 June 2022
Published 13 July 2022

TABLE 1 Summarized sequence data and genome statistics for *C. coli* strains

	MiSeq data				PacBio data										
C. coli strain	No. of reads	No. of contigs	N_{50} (bp)	Coverage (×)	No. of reads	Read length (bp)	No. of contigs	Coverage (×)	Chromosome GenBank accession no.	Chromosome size (bp)	Plasmid GenBank accession no.	Plasmid size (bp)	GC content (%)	No. of coding sequences	No. of RNAs
YH501	3,086,128	66	86,563	227	NA[a]	NA	NA	NA	CP015528	1,668,523	NA	NA	31.5	1,742	52
YH503	4,164,474	75	125,860	493	5,047	14,693	2	38	CP025281	1,705,805	CP025282	108,453	31.4	1,786	50
YH504	2,310,583	98	35,591	206	12,779	14,145	2	95	CP091644	1,722,143	CP091645	110,357	31.3	1,804	50

[a] NA, not applicable.

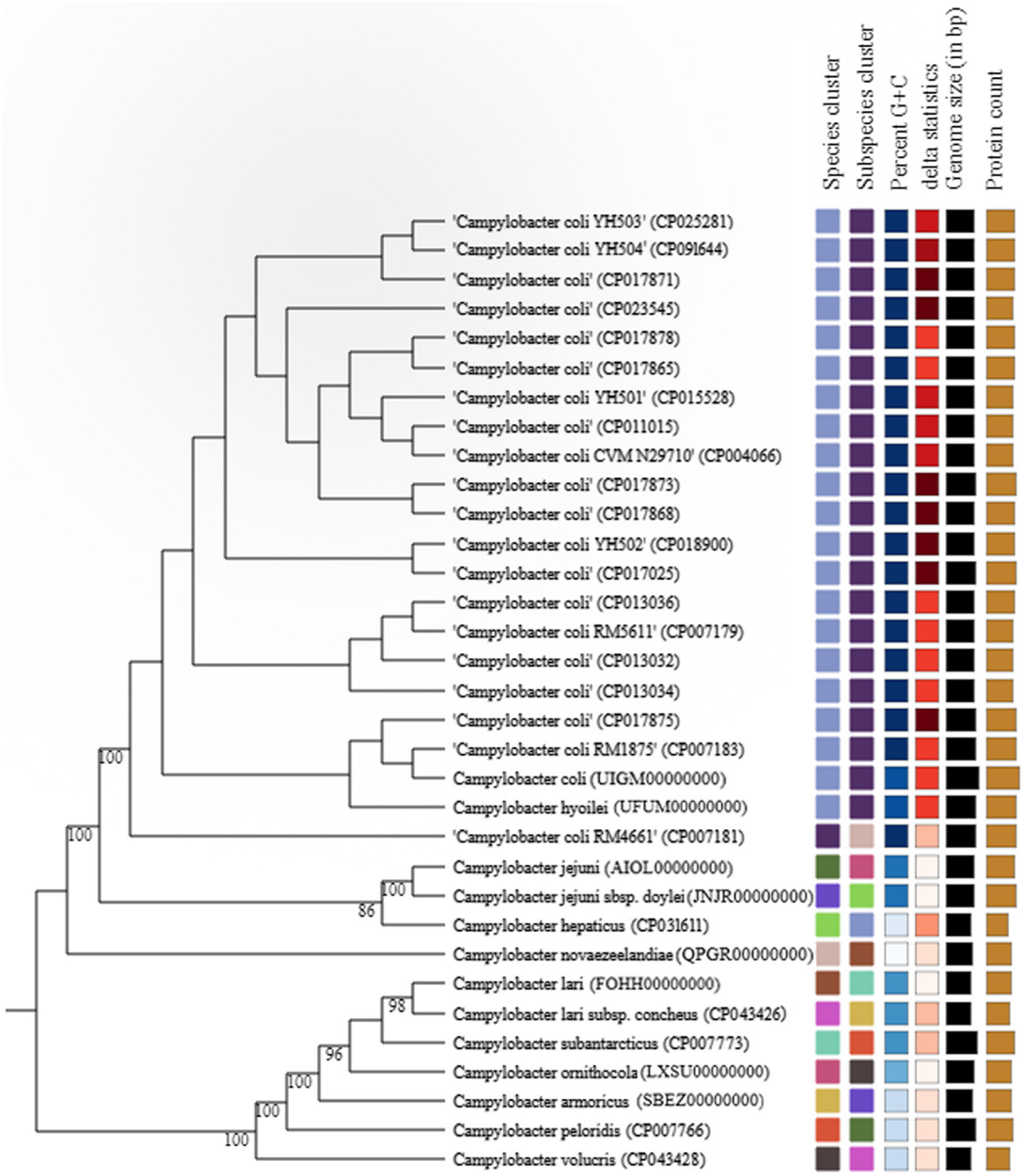

FIG 1 Phylogenomic tree of *C. coli* YH501, YH503, and YH504 and 30 other *Campylobacter* sp. strains based on GBDP. The tree was inferred using FastME 2.1.6.1 (11) with GBDP distances calculated from genome sequences. Branch lengths were scaled in terms of GBDP distances. Numbers above the branches are GBDP pseudo-bootstrap support values from 100 replications. Species and subspecies clusters are shown in color blocks. The variations of GC contents (27.39 to 31.49%) and δ statistics (0.138 to 0.285) for assessment of phylogenetic accuracy (lower δ values indicate higher accuracy) are also indicated in different colors. Genome sizes (1,439,924 to 1,938,580 bp) and protein contents (1,379 to 2,041 proteins) are shown in different colors from light (low number) to dark (high number). GenBank accession numbers of all the strains are included in the parentheses next to strain names.

gene cluster of a type VI secretion system, an important virulence factor capable of mediating hemolysis of host cells. Furthermore, the genomes were analyzed for pathogenic potential with the PathogenFinder tool (9). The predicted probability of being a human pathogen was >80.5%, indicating large potential for these poultry-originating isolates to cause disease in humans.

A whole-genome-based taxonomic analysis of the isolates was performed using the genome BLAST distance phylogeny (GBDP) method via the TYGS (https://tygs.dsmz.de) (10). Results in Fig. 1 showed that these *C. coli* food isolates were clustered together and were closely related to the reference genomes of the same *Campylobacter* species.

Data availability. The complete genome sequences of *C. coli* strains YH501, YH503, and YH504 were deposited in GenBank under the accession numbers CP015528, CP025281 and CP025282 (chromosome and plasmid pCOS503), and CP091644 and CP091645 (chromosome and plasmid pCOS504), respectively. Sequence reads for the strains are in the SRA database under the accession numbers SRX13999879 (Illumina), SRX14007924 (Illumina), SRX14007925 (PacBio), SRX14013558 (Illumina), and SRX14013559 (PacBio).

ACKNOWLEDGMENTS

This research was supported by the U.S. Department of Agriculture (USDA), Agricultural Research Service, National Program 108, Current Research Information System number 8072-42000-093.

We thank the University of Delaware DNA Sequencing and Genotyping Center for providing PacBio SMRT sequencing.

Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA.

REFERENCES

1. Silva J, Leite D, Fernandes M, Mena C, Gibbs PA, Teixeira P. 2011. *Campylobacter* spp. as a foodborne pathogen: a review. Front Microbiol 2:200. https://doi.org/10.3389/fmicb.2011.00200.
2. He YP, Reed S, Bhunia AK, Gehring A, Nguyen LH, Irwin PL. 2015. Rapid identification and classification of *Campylobacter* spp. using laser optical scattering technology. Food Microbiol 47:28–35. https://doi.org/10.1016/j.fm.2014.11.004.
3. He Y, Yao X, Gunther NW, IV, Xie Y, Tu S-I, Shi X. 2010. Simultaneous detection and differentiation of *Campylobacter jejuni*, *C. coli*, and *C. lari* in chickens using a multiplex real-time PCR assay. Food Anal Methods 3:321–329. https://doi.org/10.1007/s12161-010-9136-6.
4. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
5. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. Genome Res 27:722–736. https://doi.org/10.1101/gr.215087.116.
6. Hunt M, Silva ND, Otto TD, Parkhill J, Keane JA, Harris SR. 2015. Circlator: automated circularization of genome assemblies using long sequencing reads. Genome Biol 16:294. https://doi.org/10.1186/s13059-015-0849-0.
7. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963.
8. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res 42:D206–D214. https://doi.org/10.1093/nar/gkt1226.
9. Cosentino S, Voldby Larsen M, Moller Aarestrup F, Lund O. 2013. PathogenFinder: distinguishing friend from foe using bacterial whole genome sequence data. PLoS One 8:e77302. https://doi.org/10.1371/journal.pone.0077302.
10. Meier-Kolthoff JP, Göker M. 2019. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nat Commun 10:2182. https://doi.org/10.1038/s41467-019-10210-3.
11. Lefort V, Desper R, Gascuel O. 2015. FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program. Mol Biol Evol 32:2798–2800. https://doi.org/10.1093/molbev/msv150.