Complete Genome Sequences of Two Shiga Toxin-Producing Escherichia coli Strains Isolated from Crows

Michelle Q. Carter, a Antares Pham, a Diana K. Carychao, a Michael B. Cooley a

a Western Regional Research Center, Produce Safety, and Microbiology Research Unit, U.S. Department of Agriculture, Agricultural Research Service, Albany, California, USA

ABSTRACT Escherichia coli strains RM9088 and RM10410 were isolated from crows near a leafy greens-growing region in California in April and July 2009, respectively. Both strains carry genes encoding Shiga toxins and other virulence factors in enteric pathogens. Here, we report the complete genome sequences of RM9088 and RM10410.

Shiga toxin-producing Escherichia coli (STEC) is one of the main bacterial causal agents of foodborne illness outbreaks associated with fresh produce (1). Although STEC O157:H7 has been considered the most frequent cause of STEC-associated outbreaks, recent studies suggest that non-O157 STEC strains are causing a large number of human infections worldwide (2–4). STEC naturally resides in ruminant animals, primarily cattle; however, diverse STEC strains have been isolated from birds (5–7), implying that birds might be an environmental source of STEC transmission. To better understand the pathogenicity of avian STEC, we sequenced the genomes of the two isolates obtained from crows.

Strains RM9088 and RM10410 were isolated by cloacal swab, as described previously (8). Genomic DNA was extracted from the mid-exponential-phase cultures grown in LB broth, as described previously (9). Genomic libraries were prepared according to the PacBio 20-kb library standard protocol (10) using the SMRTbell DNA template prep kit 3.0, followed by size selection with the BluePippin size selection system (Sage Science, Inc.) and then template binding with the P6v2 kit. DNA sequencing was performed on an RS II instrument (Pacific Biosciences) with P6-C4 sequencing chemistry and a 360-min data collection protocol. The sequence reads were filtered with PreAssembler filter prior to de novo assembly with RS_HGAP_Assembly v.3. Detailed sequencing metrics and filter parameters for each strain are listed in Table 1. The closed genomes were annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (11).

The RM9088 genome is composed of a 5,270,611-bp chromosome and two plasmids, encoding a total of 5,500 coding DNA sequences (CDSs). The RM10410 genome is composed of only a 5,227,472-bp chromosome, encoding 5,114 CDSs (Table 1). The serotypes of RM9088 and RM10410 were determined to be O109:H48 and O113:H4, respectively, using SerotypeFinder 2.0 (12), with the default settings (thresholds for percent identity [%ID], 85%, and minimum length, 60%). The sequence types (STs) of RM9088 and RM10410 are ST339 and ST10, respectively, using the Warwick scheme (13). In silico phylo-typing using the Clermont method (14) placed both strains in phylogroup A.

The stx genes in strain RM9088 encode Stx1a, located on a 43,766-bp prophage (chromosome positions, base pairs 4689207 to 4732972) that was identified using PHASTER (15, 16). The p1RM9088 plasmid (167,256 bp) is a typical pEHEC (the large virulence plasmid of enterohemorrhagic E. coli [EHEC]) (17) containing genes (hlyCABD) encoding enterohemolysin. Interestingly, this plasmid also carries genes such as EAST1 (annotated as astA in RM9088) (GenBank accession number AB042002) and sta1 (GenBank accession number AJ555214) encoding heat-stable toxins, a common viru-
lence factor of enterotoxigenic *E. coli*. No known STEC virulence genes were identified on the second plasmid, p2RM9088 (86,529 bp). The two sets of *stx* genes in strain RM10410 encode Stx1a and Stx2d, respectively. The *stx*1a gene is located on a 94,727-bp prophage (chromosome positions, base pairs 1786478 to 1881204); the *stx*2d gene is located on a 59,845-bp prophage (chromosome positions, base pairs 2420694 to 2480538). A search of additional virulence factors using VirulenceFinder 2.0 (18), with the default settings (threshold for %ID, 90%; minimum length, 60%), failed to identify the pathogenicity island locus of enterocyte effacement (LEE) or genes encoding any type III secretion effectors in either of the two genomes. Since strain RM10410 has no plasmids, it lacks genes encoding enterohemolysin.

**Data availability.** The sequences described in this study are available under Bio-Project accession number PRJNA557687. The GenBank accession numbers are CP042298, CP042296, CP042297, and CP042350 for the RM9088 chromosome, plasmid p1RM9088, plasmid p2RM9088, and RM10410 chromosome, respectively. The raw reads are available under Sequence Read Archive (SRA) accession numbers SRR9953605 and SRR9953217 for RM9088 and RM10410, respectively.

**ACKNOWLEDGMENT**

This research work was supported by the U.S. Department of Agriculture Agricultural Research Service under grant CRIS 2030-42000-050-00D.

**REFERENCES**

1. Rangel JM, Sparling PH, Crowe C, Griffin PM, Swerdlow DL. 2005. Epidemiology of *Escherichia coli* O157:H7 outbreaks, United States, 1982–2002. Emerg Infect Dis 11:603–609. [https://doi.org/10.3201/eid1104.040739](https://doi.org/10.3201/eid1104.040739).
2. Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson MA, Roy SL, Carter et al.
3. Gould LH, Mody RK, Ong KL, Clogher P, Cronquist AB, Garman KN, Jones JL, Griffin PM. 2011. Foodborne illness acquired in the United States—major pathogens. Emerg Infect Dis 17:7–15. [https://doi.org/10.3201/eid1701.p11101](https://doi.org/10.3201/eid1701.p11101).
