JB Journals.ASM.org

GENOME ANNOUNCEMENT

Genome Analysis of Bovine-Mastitis-Associated *Escherichia coli* O32:H37 Strain P4

Shlomo Blum,[a] Noa Sela,[b] Elimelech D. Heller,[c] Shlomo Sela,[d] and Gabriel Leitner[a]

Department of Bacteriology, Kimron Veterinary Institute, Bet Dagan, Israel[a]; Department of Plant Pathology, Agricultural Research Organization, Volcani Center, Bet Dagan, Israel[b]; Department of Animal Sciences, Faculty of Agricultural, Food and Environmental Quality Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel[c]; and Microbial Food-Safety Research Unit, Department of Food Quality and Safety, Agricultural Research Organization, Volcani Center, Bet Dagan, Israel[d]

***Escherichia coli* is a major pathogen of bovine intramammary infections. Here we report the first draft of the genome sequence of the *E. coli* O32:H37 P4 strain, which is widely used in experimental bovine mastitis studies.**

Bovine mastitis, the inflammation of the mammary gland in cows, is a disease of important economic impact to the dairy industry. *Escherichia coli* is the most common Gram-negative bacterium causing mastitis in cows worldwide. Substantial research data support the idea that mastitis is caused by a subset of *E. coli* strains that are more adapted to the udder (2); however, no specific virulence traits or pathotypes have yet been defined for these strains.

E. coli strain P4 was isolated originally from a case of bovine mastitis, and intramammary inoculation of this strain in cows reproduced the disease (3). *E. coli* P4 has been used in a vast number of studies of bovine mastitis and has been largely accepted as a model mastitis strain. P4 is classified as O32:H37, ECOR phylogenetic group A (4), and multilocus sequence type ST10 (6). The genome of P4 was sequenced as part of a research project that aims to identify virulence factors of *E. coli* associated with mastitis.

Whole-genome sequencing was performed by Dynlabs (Zerifin, Israel) using Roche 454 GS FLEX Titanium. A total of 104,618,725 bases were obtained, consisting of 316,974 raw reads with a 330-base mean size. *De novo* assembling by Newbler 2.6 resulted in 87 contigs (63 longer than 500 bases) composed of 99.7% of the bases sequenced, with an N50 of 165,669 bases and largest contig size of 459,864 bases. The P4 genome size and G+C content were estimated to be about 5.2 Mb and 50.6%, respectively. megaBLAST comparison of *de novo* assembly results revealed high levels of similarity to *E. coli* K-12 derivative strains W3110 and DH10B along most of the sequence.

Annotation of *de novo*-assembled contigs was done using the RAST server (1). A total of 4,856 coding sequences were identified, of which 1,133 encoded hypothetical proteins, with 85% of these not assigned to a subsystem. Eighty-two RNA genes were found, 71 of which are tRNAs. Possible virulence factors identified so far are aerobactin and enterobactin siderophores, protein secretion system type II, yidE mediator of hyperadherence, CFA I fimbriae, type 1 pili, and curli. Interestingly, according to RAST, the closest neighbors of *E. coli* P4 are different strains of *E. coli* O104:H4 (highest score, 504 with strain GOS1).

Raw reads were also mapped against the genome sequence of *E. coli* K-12 MG1655 (NC_000913.2) using GS Mapper 2.6. Eighty-eight percent of reads were successfully mapped, covering 94% of the reference genome, with 20-fold coverage. Unmapped reads were assembled and analyzed by BLAST, revealing the presence of partial sequences of a plasmid F, bacteriophage P2, and a genomic island previously reported in this strain (5).

The *E. coli* P4 genome, together with whole-genome sequences of other bovine mastitis strains as well as nonpathogenic strains, will be further analyzed in order to identify potentially mastitis-specific virulence factors.

Nucleotide sequence accession numbers. This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under accession number AJQW00000000. The version described in this paper is the first version, AJQW01000000.

ACKNOWLEDGMENTS

This work was partially supported by research grants from the Israeli Dairy Board.

We acknowledge Dina Volodarsky and Inna Vulih-Shultsman (Dynlabs), who performed sequencing and initial bioinformatics analysis.

REFERENCES

1. **Aziz RK, et al.** 2008. The RAST server: rapid annotations using subsystems technology. BMC Genomics **9**:75. doi:10.1186/1471-2164-9-75.
2. **Blum S, et al.** 2008. Identification of a bovine mastitis *Escherichia coli* subset. Vet. Microbiol. **132**:135–148.
3. **Bramley AJ.** 1976. Variations in the susceptibility of lactating and nonlactating bovine udders to infection when infused with Escherichia coli. J. Dairy Res. **43**:205–211.
4. **Clermont O, Bonacorsi S, Bingen E.** 2000. Rapid and simple determination of the *Escherichia coli* phylogenetic group. Appl. Environ. Microbiol. **66**:4555–4558.
5. **Dufour D, Germon P, Brusseaux E, Le Roux Y, Dary A.** 2011. First evidence of the presence of genomic islands in *Escherichia coli* P4, a mammary pathogen frequently used to induce experimental mastitis. J. Dairy Sci. **94**:2779–2793.
6. **Wirth T, et al.** 2006. Sex and virulence in *Escherichia coli*: an evolutionary perspective. Mol. Microbiol. **60**:1136–1151.

Received 2 April 2012 Accepted 30 April 2012

Address correspondence to Shlomo Blum, shlomobl@moag.gov.il.

doi:10.1128/JB.00535-12