**RECORD OF GENOMIC ANALYSIS** Research and Development Section

Ottawa Laboratory (Carling), Ontario Laboratory Network

960 Carling Avenue, Bldg. 22, C.E.F.

Ottawa, Ontario K1A 0C6

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| OLC Reference: | | ROGA-YYYY-MM-DD VTEC | |  | Sequencing Analyst: | | |  |
| Date received at OLC: | | YYYY-MM-DD |  | | | Bioinformatics Analyst: |  | |
| Date sequenced: | | YYYY-MM-DD |  | | | Report author: |  | |
|  |  | |  | | | Verified by: |  | |
|  |  | |  | | | Location of main file: | Room B5 | |

|  |
| --- |
| **Identification Summary:**  Isolate DAR-FD-2017-MI-00109 was submitted for whole-genome sequence analysis and confirmed to be VTEC based on detection of probe sequences (e-probes) indicating the presence of the verotoxin gene *vtx2*.  Further analyses conducted using GeneSeekr probes and databases from the Center for Genomic Epidemiology (<https://cge.cbs.dtu.dk/services/SerotypeFinder/>) were used to predict serotypes.   * Strain was predicted to be serotype O8:H28.   **Quality Control Analysis:**   * Isolate is **ST-4496**. Isolates with this genotype have been recovered from linked flour samples in 2017 (See RDIMS#9360778). * These isolates do not match the OLC-795 nalidixic acid-resistant *E. coli* O157:H7 control strain used at the CFIA (ST-11, rST-2119). |

**Table 1. GeneSeekr Analysis**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Strain** | ***E. coli* species** | **Virulence profile***a* | | | | | | | **Multilocus Sequence Type (MLST)***b* | | | | | | | |
| **uidA** | **O-Antigen***c* | **H-Antigen***c* | **hlyA** | **eae** | **VT1** | **VT2** | **ST** | | **adk** | **fumC** | **gyrB** | **icd** | **mdh** | **purA** | **recA** |

● Indicates that marker is present; ND indicates marker not detected

*a*Virulence markers are detected based on detection of sequences (>80% identity) to genes from the reference genome of *Escherichia coli* O157:H7 Sakai (GenBank accession NC\_002695)[1].

*b*Multilocus sequence type (MLST) is based on the 7 gene schemes available at http://mlst.warwick.ac.uk/mlst/dbs/Ecoli (Achtman)[2]. Strains with the same sequence type (ST) have exact matches at all 7 alleles.

*c*Antigen determination based on databases available at the Center for Genomic Epidemiology (<https://cge.cbs.dtu.dk/services/SerotypeFinder/>). If not a 100% match, percent Identity relative to serotype marker is indicated in parentheses.

**Table 2. Sequence Data Quality**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Sample** | **SEQ ID** | **Total length*a*** | **Coverage*b*** | **% identity GDCS*c*** | **rST*d*** | **Pass/Fail** | **Pipeline version** |

*a*Size of an *E.coli* genome typically ranges between 4.5 and 5.5 Mb. Total length refers to the total number of nucleotides in the assembled sequence data.

*b*Coverage refers to sequencing redundancy. A minimum coverage of 20 indicates that, on average, each nucleotide in the genome has been covered by 20 sequence reads.

*c*To ensure complete coverage, sequence data is queried for Genomically Dispersed Conserved Sequences (GDCS) which currently include 53 ribosomal proteins distributed throughout the genome and conserved in all bacterial species (total of ~20000 nt sequence data). The % identity GDCS refers to the cumulative percentage of identical nucleotides to known alleles of these genes.

dRibosomal Multilocus sequence type (rST) is based on the 53 gene scheme available at http://pubmlst.org/rmlst/ [2]. Strains with the same sequence type (ST) have exact matches at all 53 alleles.