

***In silico* typing of *E. coli* using SerotypeFinder and VirulenceFinder - Results:**

Strain ID	O	H	Serotype	<i>eae</i> genes	<i>stx</i> genes	<i>stx</i> subtype	Repeated genes	Comments
STEC-EQA-11-BB-10	O187	H28	O187:H28	none	stx2g-Out-S-8	stx2g	gad , terC , traT	traT – multiple variants in database and/or multiple copies of genes
STEC-EQA-11-BH-1	O80	H2	O80:H2	eae-e07-xi	stx2d-O55-5905 stx2d-OR-TS06-08	stx2d	afaB, cia, etsC, gad , iss, mchF, nleB, nleC, ompT, shiA, terC , traJ	O80 is only detected in wzx – and with low coverage (812/1221) and ID% (96,43%). Additional analysis would be needed to determine the O-type. afaB – bug (same reference sequence) cia, etsC, iss, mchF nleB, nleC, ompT, shiA - multiple variants in database and/or multiple copies of genes traJ – same gene with different reference sequences – clean up in database
STEC-EQA-11-strain0017	O157	H7	O157:H7	eae-g01-gamma	stx1a-O157-FLY16, stx2c-O157-FLY16	stx1a, stx2c	astA, gad , nleB, tccP, terC	astA, nleB, tccP - multiple variants in database and/or multiple copies of genes

- **gad** and **terC** is found in almost all *E. coli* – disregard
- See the **extended output file for sequence alignments (results.txt)** – look at variants of same gene
- Close alleles – multiple hits of same gene, same position, different reference sequences
- Multiple copies of same gene/or pseudogenes – multiple hits of same gene, different position
- Bugs in database – 100% ID, same gene, same position, same reference sequence
- Clean up - 100% ID, same gene, same position, different reference sequences
- Assembly/fragmented genomes – same gene, different contigs, part of gene detected