

1 **TABLE S1** Genome assembly parameters

Contig measurements	TGH35	TGH64	TGH120	TGH33	TGH51
N75	80573	80573	104765	135548	125864
N50	141174	150740	490002	218943	205835
N25	236835	236932	604901	377544	373192
Minimum	1019	1019	1033	1004	1029
Maximum	353588	353716	998276	489189	489190
Average	72423	70165	108667	86928	78895
Count	57	59	37	48	54
Total	4128136	4139722	4020692	4172520	4260308

2 These genome assemblies were generated with CLC Genomics Workbench 11.0.1
 3 (Qiagen) with the following parameters: Word size 64, Minimum contig length 1000,
 4 Mismatch cost 2, No scaffolding.

5

6 **TABLE S2** Genomic features of the novel isolates

Strain	Genome size (bp)	GC%	CDS	Full length toxin-encoding
				genes
TGH35	4128136	28.61	3798	<i>tcdA, tcdB, cdtA, cdtB</i>
TGH64	4139722	28.59	3832	<i>tcdA, tcdB, cdtA, cdtB</i>
TGH120	4020692	28.33	3636	<i>tcdA, tcdB</i>
TGH33	4172520	28.67	3827	<i>tcdB</i>
TGH51	4260308	28.5	3989	<i>tcdB</i>

- 7 Rapid Annotation using Subsystems Technology (RAST) webserver was used to
- 8 annotate the genomes. CDS: coding sequences.