

**Supplementary Table 2: Preliminary validation results of *in silico* serotyping against 41 public genome assemblies
with reported serotypes**

Isolate	Accession	Gene	Reported	Blast+ result	SRST2 result	Coverage	Depth	Diffs	Uncertainty	Divergence
536	NC_008253	<i>fliC</i>	H31	H31	H31	100	72.645	47snp		2.818
536	NC_008253	<i>wzx</i>	O6	O6	O6	100	66.308			0
536	NC_008253	<i>wzy</i>	O6	O6	O6	100	67.846	1snp		0.074
12009	NC_013353	<i>fliC</i>	H2	H2	H2	100	63.414			0
12009	NC_013353	<i>wzx</i>	O103	O103	O103	100	67.102	2snp		0.158
12009	NC_013353	<i>wzy</i>	O103	O103	O103	100	61.95	2snp		0.174
11128	NC_013364	<i>wzx</i>	O111	O111	O111	100	66.801			0
11128	NC_013364	<i>wzy</i>	O111	O111	O111	100	57.491			0
11368	NC_013361	<i>fliC</i>	H11	H11	H11	100	60.43			0
11368	NC_013361	<i>wzx</i>	O26	O26	O26	100	58.613			0
11368	NC_013361	<i>wzy</i>	O26	O26	O26	100	57.426			0
55989	NC_011748	<i>fliC</i>	H4	H4	H4	100	68.39			0
55989	NC_011748	<i>wzx</i>	O104	O104	O104	100	66.659	1snp		0.078
55989	NC_011748	<i>wzy</i>	O104	O104	O104	100	68.451	4snp		0.359
APEC_O78	CP004009	<i>wzx</i>	O78	O78	O78	100	74.335	1snp		0.071
APEC_O78	CP004009	<i>wzy</i>	O78	O78	O78	100	70.872	5snp		0.446
APEC01	NC_008563	<i>wzx</i>	O1	O1	O1	100	56.318	2snp		0.163
APEC01	NC_008563	<i>wzy</i>	O1	O1	O1	100	59.354	2snp		0.199
ATCC8739	NC_010468	<i>wzx</i>	O146	O146	O146	100	69.192	10snp		0.65
ATCC8739	NC_010468	<i>wzy</i>	O146	O146	O146	100	73.081	4snp		0.339
B_str_REL606	NC_012967	<i>wzx</i>	O7	O7	O7	100	76.175	1snp		0.07
B_str_REL606	NC_012967	<i>wzy</i>	O7	O7	O7	100	73.612			0
BL21-DE3	NC_012971	<i>wzx</i>	O7	O7	O7	100	69.054			0
BL21-DE3	NC_012971	<i>wzy</i>	O7	O7	O7	100	72.365			0
BL21-Gold	NC_012947.1	<i>wzx</i>	O7	O7	O7	100	83.374			0
BL21-Gold	NC_012947.1	<i>wzy</i>	O7	O7	O7	100	77.093			0
CB9615	NC_013941	<i>fliC</i>	H7	H7	H7	100	67.35			0

Supplementary Table 2: Preliminary validation results of *in silico* serotyping against 41 public genome assemblies with reported serotypes

Isolate	Accession	Gene	Reported	Blast+ result	SRST2 result	Coverage	Depth	Diffs	Uncertainty	Divergence
CB9615	NC_013941	wzx	O55	O55	O55	100	67.629			0
CB9615	NC_013941	wzy	O55	O55	O55	100	65.309			0
CE10	NC_017646	wzx	O7	O7	O7	100	66.232			0
CE10	NC_017646	wzy	O7	O7	O7	99.915	66.563	1indel		0
CFT073	NC_004431	fliC	H1	H1	H1	100	71.246			0
CFT073	NC_004431	wzx	O6	O6	O6	100	70.055			0
CFT073	NC_004431	wzy	O6	O6	O6	100	64.177	2snp		0.149
K-12_substr_DH10B	NC_010473	wzx	O16	O16	O16	100	70.173			0
K-12_substr_DH10B	NC_010473	wzy	O16	O16	O16	100	74.333			0
E2348/69	NC_011601	fliC	H6	H6	H6	100	69.654	12snp		0.729
E2348/69	NC_011601	wzx	O127	O127	O127	100	67.362			0
E2348/69	NC_011601	wzy	O127	O127	O127	100	72.561			0
EC4115	NC_011353	fliC	H7	H7	H7	100	66.001			0
EC4115	NC_011353	wzx	O157	O157	O157	100	56.534			0
EC4115	NC_011353	wzy	O157	O157	O157	100	66.09	1snp		0.084
ED1a	NC_011745	wzx	O81	O81	O81	100	68.21			0
ED1a	NC_011745	wzy	O81	O81	O81	100	62.546			0
EDL933	NC_002655	fliC	H7	H7	H7	100	63.628	1snp		0.057
EDL933	NC_002655	wzx	O157	O157	O157	100	64.826			0
EDL933	NC_002655	wzy	O157	O157	O157	100	57.896	1snp		0.084
EL-2071	CP003301	fliC	H4	H4	H4	100	69.741			0
EL-2071	CP003301	wzx	O104	O104	O104	100	62.335			0
EL-2071	CP003301	wzy	O104	O104	O104	100	60.478	4snp		0.359
H10407	NC_017633	fliC	H11	H11	H11	100	69.405			0
H10407	NC_017633	wzx	O78	O78	O78	100	70.418	1snp		0.071
H10407	NC_017633	wzy	O78	O78	O78	100	60.13	5snp		0.446
HS	CP000802	wzm	O9	O9	O9	100	69.706	14snp		1.781

**Supplementary Table 2: Preliminary validation results of *in silico* serotyping against 41 public genome assemblies
with reported serotypes**

Isolate	Accession	Gene	Reported	Blast+ result	SRST2 result	Coverage	Depth	Diffs	Uncertainty	Divergence
HS	CP000802	<i>wzt</i>	O9	O9	O9	99.923	71.897	2snp1indel		3.241
IAI1	NC_011741	<i>wzm</i>	O8	O8	O8	100	64.781	5snp		0.629
IAI1	NC_011741	<i>wzt</i>	O8	O8	O8	100	68.502	5snp		0.412
IAI39	NC_011750	<i>wzx</i>	O7	O7	O7	100	65.838			0
IAI39	NC_011750	<i>wzy</i>	O7	O7	O7	100	64.266			0
IHE3034	NC_017628	<i>wzx</i>	O18	O18	O18	100	66.773	2snp		0.16
IHE3034	NC_017628	<i>wzy</i>	O18	O18	O18	100	66.731	1snp		0.084
K-12_substr_W3110	NC_007779	<i>wzx</i>	O16	O16	O16	100	72.556			0
K-12_substr_W3110	NC_007779	<i>wzy</i>	O16	O16	O16	100	71.129			0
K-12_substr_MG1655	NC_000913	<i>wzx</i>	O16	O16	O16	100	66.441			0
K-12_substr_MG1655	NC_000913	<i>wzy</i>	O16	O16	O16	100	73.294			0
NRG857C	NC_017634	<i>fliC</i>	H1	H1	H1	100	75.528			0
NRG857C	NC_017634	<i>wzx</i>	O83	O83	O83	100	66.081	1snp		0.069
NRG857C	NC_017634	<i>wzy</i>	O83	O83	O83	100	66.916			0
O42	NC_017626	<i>wzx</i>	O44	O44	O44	100	66.449			0
O42	NC_017626	<i>wzy</i>	O44	O17	O77	100	61.128			0
p12b	NC_017663	<i>fliC</i>	H17	H4	H4	100	72.548	7snp		0.667
p12b	NC_017663	<i>fliA</i>	H17	H17	H17	100	72.113			0
RM12579	NC_017656	<i>fliC</i>	H7	H7	H7	100	68.684			0
RM12579	NC_017656	<i>wzx</i>	O55	O55	O55	100	65.32			0
RM12579	NC_017656	<i>wzy</i>	O55	O55	O55	100	63.159			0
S88	NC_011742	<i>fliC</i>	H7	H7	H7	100	70.367			0
S88	NC_011742	<i>wzx</i>	O45	O45	O45	100	65.614			0
S88	NC_011742	<i>wzy</i>	O45	O45	O45	100	68.822			0
SAKAI	NC_002695	<i>fliC</i>	H7	H7	H7	100	59.861			0
SAKAI	NC_002695	<i>wzx</i>	O157	O157	O157	100	66.442			0
SAKAI	NC_002695	<i>wzy</i>	O157	O157	O157	100	62.395	1snp		0.084

Supplementary Table 2: Preliminary validation results of *in silico* serotyping against 41 public genome assemblies with reported serotypes

Isolate	Accession	Gene	Reported	Blast+ result	SRST2 result	Coverage	Depth	Diffs	Uncertainty	Divergence
SE11	NC_011415	<i>fliC</i>	H28	H28	H28	100	69.52	21snp		1.207
SE11	NC_011415	<i>wzx</i>	O152	O173	O173	99.833	63.844	2snp2hol	edge0.0	1.002
SE11	NC_011415	<i>wzy</i>	O152	O173	O173	100	62.898	11snp		1.019
SE15	NC_013654	<i>fliC</i>	H5	H5	H5	100	69.559	11snp		0.839
SE15	NC_013654	<i>wzx</i>	O150	O16	O16	100	69.567			0
SE15	NC_013654	<i>wzy</i>	O150	O16	O16	100	65.667			0
TW14359	NC_013008	<i>fliC</i>	H7	H7	H7	100	62.298			0
TW14359	NC_013008	<i>wzx</i>	O157	O157	O157	100	64.424			0
TW14359	NC_013008	<i>wzy</i>	O157	O157	O157	100	65.575	1snp		0.084
UMN026	NC_011751	<i>fliC</i>	H18	H18	H18	100	65.3			0
UMN026	NC_011751	<i>wzx</i>	O17	O17	O17	100	66.491			0
UMN026	NC_011751	<i>wzy</i>	O17	O17	O77	100	63.455			0
UMNF18	CP002890	<i>wzx</i>	O147	O147	O147	100	61.154			0
UMNF18	CP002890	<i>wzy</i>	O147	O147	O147	100	61.378			0
UMNK88	NC_017641	<i>wzx</i>	O149	O149	O149	99.916	58.293	1indel		0
UMNK88	NC_017641	<i>wzy</i>	O149	O149	O149	100	64.085			0
UT189	NC_007946	<i>wzx</i>	O18	O18	O18	100	65.882	2snp		0.16
UT189	NC_007946	<i>wzy</i>	O18	O18	O18	100	67.839	1snp		0.084
Xuzhou21	NC_017906	<i>fliC</i>	H7	H7	H7	100	63.155			0
Xuzhou21	NC_017906	<i>wzx</i>	O157	O157	O157	100	67.413			0
Xuzhou21	NC_017906	<i>wzy</i>	O157	O157	O157	100	60.221	1snp		0.084

Coverage = percentage of the gene length that was covered

Depth = mean read depth across the length of all alleles which were assigned a top scoring allele number

Diffs = differences between the detected allele and the best scoring allele.

Uncertainty = details of parts of the top scoring alleles for which the depth of coverage was too low to give confidence in the result

Supplementary Table 2: Preliminary validation results of *in silico* serotyping against 41 public genome assemblies with reported serotypes

Isolate	Accession	Gene	Reported	Blast+ result	SRST2 result	Coverage	Depth	Diffs	Uncertainty	Divergence
---------	-----------	------	----------	---------------	--------------	----------	-------	-------	-------------	------------

Divergence = divergence of allele detected to sequence database (ran with option to report <10% divergence)

Reported =reported phenotype of the isolate

SRST2 = call from short read data and the EcOH database

BLAST+ = call from BLAST analysis of Velvet assemblies and the EcOH database