Supporting Information for

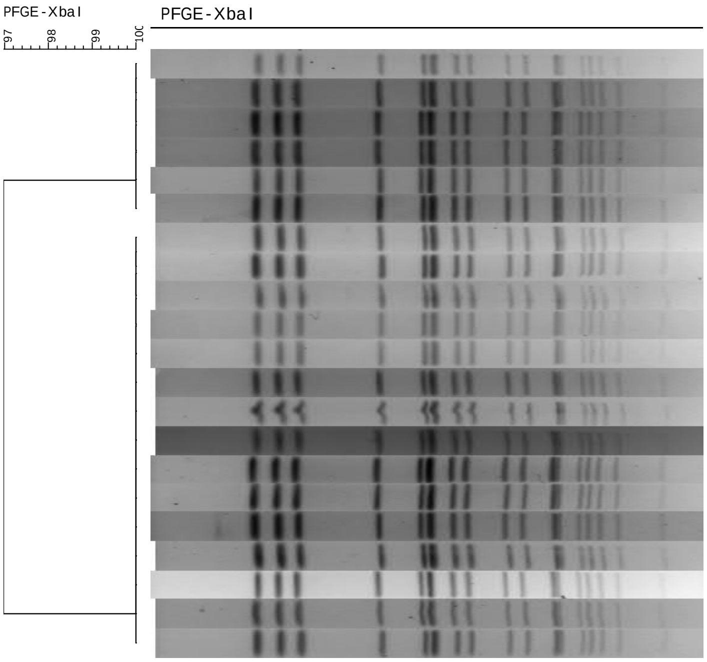
Comparative Genomic Analyses of Municipal Wastewater *Salmonella* Isolates Detects a Clinically Unreported Salmonellosis Outbreak

Sabrina Diemert and Tao Yan*

* Corresponding author (taoyan@hawaii.edu)

Department of Civil and Environmental Engineering, University of Hawaii at Manoa, Honolulu, HI 96822

- Number of pages: 4 (S1-S4)
- Number of tables: 5 (Tables S1-S5; Tables S3-S5 provided as Microsoft Excel spreadsheets)
- Number of figures: 1 (S1)



Key	Serotype	IsolatDate
Y13	Paratyphi B var. L(+) tartrate + (Java)	2010-05-17
Y30	Paratyphi B var. L(+) tartrate + (Java)	2010-05-24
Y33	Paratyphi B var. L(+) tartrate + (Java)	2010-06-01
Y34	Paratyphi B var. L(+) tartrate + (Java)	2010-06-01
Y81	Paratyphi B var. L(+) tartrate + (Java)	2010-07-06
Y27	Paratyphi B var. L(+) tartrate + (Java)	2010-05-24
Y1	Paratyphi B var. L(+) tartrate + (Java)	2010-05-10
Y11	Paratyphi B var. L(+) tartrate + (Java)	2010-05-17
Y2	Paratyphi B var. L(+) tartrate + (Java)	2010-05-10
Y21	Paratyphi B var. L(+) tartrate + (Java)	2010-05-24
Y22	Paratyphi B var. L(+) tartrate + (Java)	2010-05-24
Y29	Paratyphi B var. L(+) tartrate + (Java)	2010-05-24
Y3	Paratyphi B var. L(+) tartrate + (Java)	2010-05-10
Y38	Paratyphi B var. L(+) tartrate + (Java)	2010-06-01
Y385	Paratyphi B var. L(+) tartrate + (Java)	2011-03-28
Y386	Paratyphi B var. L(+) tartrate + (Java)	2011-04-04
Y395	Paratyphi B var. L(+) tartrate + (Java)	2011-05-02
Y4	Paratyphi B var. L(+) tartrate + (Java)	2010-05-10
Y401	Paratyphi B var. L(+) tartrate + (Java)	2011-05-09
Y5	Paratyphi B var. L(+) tartrate + (Java)	2010-05-10
Y9	Paratyphi B var. L(+) tartrate + (Java)	2010-05-10

Figure S1: PFGE dendrogram for *Salmonella enterica* serovar Paratyphi B var. Java (Paratyphi B var. L(+) tartrate (+), aka *S*. Java) isolates from Hawaii municipal wastewater (MW).

Differences in clusters are likely ascribed to annotation of a single band as singlet vs. doublet in BioNumerics. Note that samples are named beginning with “Y”, instead of “HIY” elsewhere in publication. Clinical isolates from the 2010 outbreak sharing the same PFGE pattern were assigned the PFGE XbaI pattern JKXX01.0194 via PulseNet (Pamela O’Brien, personal communication, November 2018).

Table S1: SPAdes genome assembly data for *S.* Java isolates from Hawaii municipal wastewater (MW)

Isolate name[a]	Biosample accession number[b]	SRA run number (for WGS data)[c]	Genome assembly coverage[d]	Number of contigs	Number of contigs >500 bp	Genome size (Mb)	Longest contig	N50[e]	L50[e]
HIY0001_2010-05-10	SAMN10496253	SRR5238264	109.8	62	50	4.69	378196	221551	8
HIY0002_2010-05-10	SAMN10496250	SRR5238166	80.6	62	52	4.69	604759	312296	6
HIY0003_2010-05-10	SAMN10496252	SRR5238249	137.2	64	37	4.70	630300	334985	6
HIY0004_2010-05-10	SAMN10496251	SRR5238235	111.6	75	64	4.69	397537	172158	9
HIY0005_2010-05-10	SAMN10496249	SRR5238148	83.8	65	56	4.69	414658	229518	7
HIY0009_2010-05-10	SAMN10496242	SRR3294513	53.8	72	57	4.69	393271	203825	8
HIY0011_2010-05-17	SAMN10496240	SRR3182472	73.5	54	43	4.69	397537	232688	8
HIY0013_2010-05-17	SAMN10496254	SRR5318398	61.8	51	37	4.69	687767	321416	6
HIY0021_2010-05-24	SAMN10496255	SRR5318402	31.8	47	38	4.69	630263	321416	6
HIY0022_2010-05-24	SAMN10496256	SRR5366629	90.5	53	42	4.69	630300	321416	6
HIY0027_2010-05-24	SAMN10496260	SRR5470976	78.7	68	53	4.70	604651	231012	7
HIY0029_2010-05-24	SAMN10496259	SRR5470973	73.4	61	47	4.70	630263	334985	6
HIY0030_2010-05-24	SAMN10496241	SRR3182490	232.0	54	31	4.70	687804	397250	5
HIY0033_2010-06-01	SAMN10496257	SRR5470970	49.7	52	41	4.69	630262	334985	6
HIY0034_2010-06-01	SAMN10496258	SRR5470972	73.1	63	45	4.70	630300	334985	6
HIY0038_2010-06-01	SAMN10496245	SRR3457661	67.9	53	42	4.69	630416	321416	6
HIY0081_2010-07-06	SAMN10496247	SRR3457664	76.7	52	43	4.69	630532	334985	6
HIY0385_2011-03-28	SAMN10496246	SRR3457662	51.5	43	36	4.69	630300	334985	6
HIY0386_2011-04-04	SAMN10496248	SRR3457665	48.6	47	39	4.69	630300	334985	6
HIY0395_2011-05-02	SAMN10496243	SRR3457658	54.8	48	39	4.69	687805	334985	6
HIY0401_2011-05-09	SAMN10496244	SRR3457660	46.6	47	38	4.69	630263	321416	6

Notes:

a. Isolate name is composed of (Sample ID)_(Isolation date).

b. BioSamples for assemblies are uploaded to NCBI under BioProject PRJNA507471.

c. WGS data are uploaded to NCBI SRA under BioProject PRJNA274995. This is linked with BioProject PRJNA507471 under umbrella BioProject PRJNA508196.

d. Assembly coverage assessed using BBMap of trimmed SRA paired-end reads against assembled contigs.

e. N50 and L50 assessed using QUAST(1); parameters reported only consider contigs ≥ 500 bp.

Table S2: Additional clinical S. Java isolate data (available from BioProject PRJNA230403)

Sample name	**Strain**	**Isolate type**	**BioSample**	**SRA accession number**	**Run number**	**Isolation date**	**Additional metadata**	**Submitter**
SRR7639130-USA-Unknown	2010K-0885	clinical	SAMN09761643	SRX4502673	SRR7639130	Unknown	Data missing	EDLB-CDC (Enteric Diseases Laboratory Branch)
SRR7639138-USA-Unknown	2010K-0883	clinical	SAMN09761656	SRX4502681	SRR7639138	Unknown	Data missing	EDLB-CDC (Enteric Diseases Laboratory Branch)
SRR3205681-USA-2010-05	Paratyphi B var. L(+) tartrate + (Java)	clinical	SAMN04485993	SRX1614841	SRR3205681	2010-05	Stool sample, host age 30-39, HHS region 9	CDC-PulseNet

All samples from Table S2 were obtained from exploration of the NCBI Pathogen Detection (Beta) Isolates Browser (2); the Browser provides access to data from the NCBI Pathogen Detection Pipeline: whole genome sequencing (WGS) data from the FDA, CDC, USDA, and other organizations is collected, clustered via maximum compatibility (3), and screened for the presence of antimicrobial resistance genes. These isolates were chosen from the same SNP cluster as the MW isolates (SNP cluster accession = PDS000027748.8).

References

1. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: Quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075.
2. NCBI. 2018. Pathogen Detection (Beta) Isolates Browser. https//www.ncbi.nlm.nih.gov/pathogens, retreived June 28, 2018.
3. Cherry JL. 2017. A practical exact maximum compatibility algorithm for reconstruction of recent evolutionary history. BMC Bioinformatics 18:1–12.