

1 **No isolate, no problem: Using a novel insertion sequence PCR to link rats to human**
2 **shigellosis cases in an underserved urban community**

3 Gordon Ritchie^{1,2}, Victor Leung^{1,2}, Chelsea G. Himsworth^{3,4}, Kaylee A. Byers^{3,4}, Lisa K.F.
4 Lee^{3,5}, Samuel D. Chorlton², Aleksandra Stefanovic^{1,2}, Marc G. Romney^{1,2}, Nancy Matic^{1,2*},
5 Christopher F. Lowe^{1,2*}

6 ¹Division of Medical Microbiology and Virology, St. Paul's Hospital, Vancouver, Canada

7 ²Department of Pathology and Laboratory Medicine, University British Columbia, Vancouver,
8 Canada

9 ³British Columbia Regional Centre, Canadian Wildlife Health Cooperative, Abbotsford, Canada

10 ⁴School of Population and Public Health, University of British Columbia, Vancouver, Canada

11 ⁵Department of Veterinary Pathology, Western College of Veterinary Medicine, Saskatoon,
12 Canada

13 *These senior authors contributed equally to this article

14 Running Title: *Shigella* strain typing with insertion sequence PCR

15 Word Count: 1877

16 Corresponding Author:

17 Christopher Lowe

18 St. Paul's Hospital, Providence Health Care

19 Virology Laboratory

20 1081 Burrard St.

21 Vancouver, BC, Canada

22 V6Z 1Y6

23 Email: clowe@providencehealth.bc.ca

24 Tel: 604-806-8422

25

26 Abstract

27 Introduction: During an investigation into a cluster of *Shigella flexneri* serotype 2a cases in an
28 underserved community, we assessed the relatedness of human and rat *S. flexneri* isolates
29 utilizing a novel PCR targeting insertion sites (IS-PCR) of mobile elements in the *Shigella*
30 genome characteristic of the cluster strain.

31 Methods: Whole genome sequences of *S. flexneri* (n=50) associated with the cluster were
32 analyzed. *de novo* genome assemblies were analyzed by a Geneious V10.2.6 motif search, and 2
33 unique IS were identified in all human *Shigella* sequences of the local cluster. Hydrolysis probe
34 PCR assays were designed to detect these sequences consisting of forward and reverse primers to
35 amplify across each insertion site, and a hydrolysis probe spanning the insertion site. IS-PCR
36 was performed for three *Shigella* PCR-positive culture-negative rat intestine specimens from this
37 community.

38 Results: Both insertion sites were detected in the *de novo* genome assemblies of all clinical *S.*
39 *flexneri* isolates (n=50). Two of the three PCR-positive culture-negative rat samples were
40 positive for both unique IS identified in the human *S. flexneri* isolates, suggesting that the rat
41 *Shigella* spp. strains were closely related to the human strains in the cluster. The cycle threshold
42 (Ct) values were >35, indicating that the bacterial load was very low in the rat samples.

43 Conclusions: Two unique IS were identified in clinical isolates from a community *S. flexneri*
44 cluster. Both IS targets were identified in PCR-positive (*Shigella* spp.), culture-negative rat
45 tissue and clinical isolates from humans, indicating relatedness.

46

47 **Introduction**

48 *Shigella* spp. have high potential for transmission from person-to-person through fecal-
49 oral contact due to their low infectious dose (10-100 organisms)(1). Humans are thought to be
50 the only natural reservoir for *Shigella* spp., and outbreaks/clusters have been reported in a variety
51 of settings, including daycares/schools, men who have sex with men, foodborne transmission,
52 travelling and regions with poor hygiene and sanitation(2, 3).

53 There are numerous methods available for outbreak analysis (pulsed-field gel
54 electrophoresis [PFGE], amplified fragment length polymorphism [AFLP], random amplified
55 polymorphic DNA [RAPD], variable number tandem repeat [VNTR], multi-locus sequence
56 typing [MLST], and core genome MLST [cgMLST]), but these methods are dependent on
57 growth of the bacterial isolate in culture(4). Culture-dependent methods are challenging when
58 working with fastidious organisms or samples with low burdens of pathogen. Cases of culture-
59 negative shigellosis have been increasingly recognized due to the use of molecular diagnostics(5,
60 6), rendering many of these typing methods ineffective. These methods are also resource-
61 intensive requiring specialized instrumentation and technical/bioinformatics expertise. Culture-
62 independent typing methods that are scalable and simple to perform are currently needed to
63 enhance outbreak investigations. Specifically for *Shigella* spp. which contain a high number of
64 mobile genetic elements and genomic rearrangements, strain relatedness can potentially be
65 analyzed through targeting of insertion sequences(3, 7).

66 In 2021, a cluster of *Shigella flexneri* serotype 2a cases occurred in an underserved
67 neighbourhood in Vancouver, British Columbia. A specific source (e.g. housing, food, sexual
68 exposures, travel) was not identified, but poor sanitary conditions in this community were
69 presumed to have contributed to ongoing transmission

70 (<http://www.vch.ca/Documents/26%20Feb%202021%20Physicians%20Update%20Shigella%20Outbreak%20Hep%20A%20Cases.pdf>). Rats were hypothesized to be one potential vector for
71 ongoing transmission based on previous observations(8). Although samples from rats (*Rattus*
72 *norvegicus*) recovered from this community were culture-negative for *Shigella* spp., three rat
73 samples tested positive for *Shigella*/Enteroinvasive *E. coli* (EIEC) using polymerase chain
74 reaction (PCR) [Biofire® FilmArray® Gastrointestinal Panel](C. Himsworth, unpub. data). We
75 utilized a novel PCR targeting insertion sites (IS-PCR) of mobile elements in the *Shigella*
76 genome characteristic of the outbreak strain for outbreak analysis.
77

78 **Methods:**

79 *S. flexneri* (n=50) isolates from the outbreak underwent Whole Genome Sequencing
80 (WGS) using the GridION. All genomes were serotype 2a (*in silico* serotyping with
81 ShigEiFinder) and MLST 245(9). Human *Shigella flexneri* isolates associated with the
82 community outbreak were grown overnight in Mueller Hinton broth, then inactivated at 99°C for
83 10mins. 250µL bacterial Lysis Buffer (Roche) was added to 250µL bacterial suspension,
84 followed by the addition of 25µL of Proteinase K (Sigma). The suspension was incubated at
85 65°C for 1 hour, then DNA was extracted on the MagNA Pure 24 (Roche Molecular Systems).
86 The DNA was prepared for sequencing with the SQK-LSK109 Ligation Sequencing Kit with
87 NBD104/114 barcoding. The DNA library was run on GridIon (Oxford Nanopore Technologies)
88 with R9 flowcells. Raw sequencing data was basecalled with Guppy (v4.1.0) and uploaded to
89 BugSeq for further automated analysis(10). Reads were assembled with metaFlye and polished
90 with Racon, Medaka (<https://github.com/nanoporetech/medaka>) and Homopolish(11–13). A
91 consensus FASTA sequence was constructed using bugseq.com. 109 copies of the IS1 mobile
92 element were found in the resulting *de novo* genome assemblies by a Geneious V10.2.6 motif

93 search. Sequences spanning these insertion sites were scanned by BLAST searches of Genbank
94 (performed August 2022 – at the time, >10,000 *S. flexneri* and >170,000 *E. coli* genome
95 sequences). One unique insertion site was present in all human *Shigella* sequences in the local
96 outbreak, but not present in any other *E. coli*/*Shigella* sequences in Genbank (ie all publicly
97 available *E. coli*/*Shigella* sequences) at the time of primer design. Another rare insertion site was
98 identified that was present in only 40 *S. flexneri* sequences within Genbank. This site was a result
99 of a fusion of the *acrA*-like gene with the *yegE* gene interrupted by an ISSfl2 insertion sequence.
100 Neither IS was identified in any other ST245 strain in Genbank. A BLAST search for other
101 gastrointestinal pathogens was conducted, but did not identify any matches with the IS-PCR
102 products.

103 Rats were collected from March 23 to April 9, 2021 from six city blocks in Vancouver's
104 urban Downtown Eastside Neighbourhoods. City blocks were selected with highest
105 concentration of positive human *S. flexneri* cases. We used lethal Snap-e rat traps (Kness
106 Manufacturing Co., Iowa, United States) inserted inside of tempered PROTECTA EVO Express
107 Bait Stations (Bell Laboratories, Wisconsin, United States). Six traps were placed in the
108 alleyway of each city block for a total of 36 active traps. We checked traps each morning, and for
109 trapped rats we recorded capture date and location. Following capture, rats were stored at -20°C
110 prior to undergoing a full necropsy. To collect samples for *Shigella* testing, we collected each
111 rat's whole intestine. This was stored at -80°C prior to *Shigella* spp. testing. Ten rat specimens
112 (intestine) from this community were recovered for IS-PCR (three PCR-positive culture-
113 negative). The rat intestine samples were suspended in 700uL of phosphate buffered saline (PBS)
114 and bead lysed with 0.2mm beads on the Qiagen TissueLyser LT for 3 minutes, then extracted on
115 the Roche MagNA pure compact and eluted in 50uL. To increase the specificity of the PCR,

116 hydrolysis probe PCR assays were designed to detect these sequences consisting of forward and
117 reverse primers to amplify across each insertion site, and a hydrolysis probe spanning the
118 insertion site (Table 1). PCR was performed on the Lightcycler® 480 (Roche Diagnostics, Laval,
119 QC) using the IDT Gene Expression Master Mix with 0.2uM primer, 0.1uM probe, and 1mg/mL
120 BSA. The cycling conditions were 95°C for 10mins, then 60 cycles of 97°C for 5s, 66°C
121 stepping down to 60°C at 0.2°C /cycle for 20s, 72°C for 2s, with fluorescence determination at
122 510nm. All primers and probes were synthesized by Integrated DNA Technologies. PCR positive
123 control was a 1/1000 dilution of one of the human outbreak isolates. PCR negative control was
124 PBS that was processed at the same time as the rat samples including bead lysing and extraction.
125 Bioanalyzer 2100 gel electrophoresis (Agilent Technologies, CA) was performed on the PCR
126 products. PCR products were barcoded, normalized and applied to the MinION (R9.4 FLO-
127 MIN106D) according to the Oxford Nanopore 1D Native barcoding protocol (version:
128 NBE_9065_v109_revE_23May2018). All DNA clean-ups were performed with 2X SPRI
129 Agencourt beads to ensure adequate DNA recovery from the low MW PCR products. Consensus
130 sequence reads were analyzed with Geneious 10.2.6 software.

131 This study was approved by the University of British Columbia's Animal Care
132 Committee (A20-0212) and adhered to guidelines outlined by the Canadian Council on Animal
133 Care.

134 **Results**

135 Both insertion sites were detected in the *de novo* genome assemblies of all clinical *S.*
136 *flexneri* isolates (n=50), which represented samples from the course of the outbreak.

137 Two of the three PCR-positive culture-negative rat samples were positive for both unique
138 IS identified in the human *S. flexneri* isolates, suggesting that the rat *Shigella* spp. strains were
139 closely related to the human strains in the outbreak. The cycle threshold (Ct) values for the PCR
140 were >35, indicating that the bacterial load was very low in the rat samples (Table 2). The 7
141 PCR-negative culture-negative rat samples were negative for both unique IS.

142 Agilent gel of IS-PCR products gave the expected amplicon sizes from the WGS of the
143 human *S. flexneri* isolates – 206bp and 212bp. Consensus sequence reads from the positive rat
144 samples were 100% identical to the human isolate sequences, confirming the presence of the IS
145 in the rat samples.

146 Discussion

147 We used a novel strain typing method (IS-PCR) to show that *Shigella* spp. from humans
148 and rats in the same community were related. In *Shigella* spp., much of the strain-to-strain
149 differences are due to transposase activity within mobile elements, which result in genome
150 rearrangement. This method uses the insertion site of mobile elements in the *Shigella* genome to
151 link bacterial strains, as insertion sites are generally stable through the course of an outbreak(3).
152 Two mobile element insertion sites were found in all the human *Shigella* outbreak isolates that
153 were rare in Genbank, and also detected in the rat samples by amplifying across the insertion
154 site, with enhanced discriminatory power via a hydrolysis probe spanning the insertion site. In
155 addition, gel electrophoresis showed identical sized amplicons for the rat samples and the human
156 isolates, and PCR products were sequenced and matched the targeted IS. The discovery of
157 identical rare insertion sites for IS1 in the human isolates and the rat samples demonstrates that
158 the *Shigella* strains are related. Although IS-PCR can infer relatedness, it cannot suggest
159 directionality of transmission. In addition, as rats are coprophagic, detection of low level *S.*

160 *flexneri* DNA in rat intestines may be secondary to contact or consumption of *Shigella*
161 contaminated food or objects rather than *Shigella* infection in the rats. Despite this, identification
162 of the human strain in rats can provide Public Health with a potential source to target in
163 mitigating ongoing transmission, though further research is required to understand the role rats
164 may play in the spread of human-associated pathogens such as *Shigella*, or conversely,
165 transmission of *Shigella* from humans to rats.

166 The utilization of IS-PCR enabled typing of culture-negative rat samples. There is a need
167 for culture-independent modalities to support outbreak investigations. Stool cultures may not
168 always be able to isolate *Shigella* spp. due to antibiotic treatment or non-viable bacteria. With the
169 evolving use of molecular testing for the detection of gastrointestinal pathogens, an isolate may
170 not be readily available for culture-dependent typing methods. PCR-based methods enable typing
171 of samples with low organism burden, which may be below the limit of detection for culture or
172 metagenomics(14). This methodology may also be amenable to outbreak/cluster investigations
173 for other microorganisms, such as *Mycobacteria* spp., *Corynebacterium* spp. or *Staphylococcus*
174 spp., where strain diversity is similarly achieved through mobile elements traversing throughout
175 the genome(15).

176 Unique to this study was the detection of *Shigella* spp. directly from rat intestine, and
177 evidence that it shared the same two IS sites as the human outbreak strains. A PCR-based
178 approach was able to identify the IS sites at low levels from non-sterile tissue. This may not be
179 possible with metagenomic sequencing as other more abundant gastrointestinal flora would be
180 preferentially sequenced, precluding recovery of complete *Shigella* spp. genomes.

181 This study is limited by sample size. We applied this novel IS-PCR to only one known
182 outbreak to date, and a potential limitation of the IS-PCR method would be the ability to detect

183 unique insertion sites for each outbreak, which would require WGS of the index outbreak strains.
184 A combination of IS-PCRs for rare insertion sites should also be considered, where the presence
185 of a particular combination of 2-3 insertion sites may be uniquely identified per outbreak with
186 good discriminatory power. Further work is also needed to assess the specificity of the IS-PCR,
187 including testing of historical samples within our geographical region. As IS are transposable
188 elements, there is a possibility for movement of the IS over time. For acute outbreaks over a
189 defined time, it would be unlikely to affect the IS-PCR design, but reaffirms the need to reassess
190 the appropriateness of the IS-PCR for subsequent outbreaks.

191 In conclusion, IS-PCR is a novel rapid method for determining relatedness of *Shigella*
192 strains, which we utilized to link clinical human cases to rats. As a PCR-based approach, it may
193 be better suited for direct PCR on tissue or samples where bacterial culture is not possible, and
194 can serve as a complimentary modality to existing laboratory methods such as WGS for
195 cluster/outbreak investigation.

196 **Conflicts of Interest**

197 SDC is a shareholder in BugSeq Bioinformatics Inc. The other authors report no relevant
198 conflicts of interest.

199 **Funding**

200 There was no financial support for this study.

201 **Acknowledgements**

202 We thank Harveen Atwal, University of Saskatchewan and Gustavo Gabaldon, Orkin Canada for
203 assistance with rat trapping.

204 **References**

205 1. Zaidi MB, Estrada-García T. 2014. Shigella: A Highly Virulent and Elusive Pathogen.
206 Curr Trop Med Reports 1:81–87.

207 2. Khalil IA, Troeger C, Blacker BF, Rao PC, Brown A, Atherly DE, Brewer TG, Engmann
208 CM, Houpt ER, Kang G, Kotloff KL, Levine MM, Luby SP, MacLennan CA, Pan WK,
209 Pavlinac PB, Platts-Mills JA, Qadri F, Riddle MS, Ryan ET, Shoultz DA, Steele AD,
210 Walson JL, Sanders JW, Mokdad AH, Murray CJL, Hay SI, Reiner RC. 2018. Morbidity
211 and mortality due to shigella and enterotoxigenic Escherichia coli diarrhoea: the Global
212 Burden of Disease Study 1990–2016. Lancet Infect Dis 18:1229–1240.

213 3. The HC, Thanh DP, Holt KE, Thomson NR, Baker S. 2016. The genomic signatures of
214 Shigella evolution, adaptation and geographical spread. Nat Rev Microbiol 14:235–250.

215 4. Sabat AJ, Budimir A, Nashev D, Sá-Leão R, van Dijl JM, Laurent F, Grundmann H,
216 Friedrich AW. 2013. Overview of molecular typing methods for outbreak detection and
217 epidemiological surveillance. Eurosurveillance 18:1–15.

218 5. Liu J, Platts-Mills JA, Juma J, Kabir F, Nkeze J, Okoi C, Operario DJ, Uddin J, Ahmed S,
219 Alonso PL, Antonio M, Becker SM, Blackwelder WC, Breiman RF, Faruque ASG, Fields
220 B, Gratz J, Haque R, Hossain A, Hossain MJ, Jarju S, Qamar F, Iqbal NT, Kwambana B,
221 Mandomando I, McMurry TL, Ochieng C, Ochieng JB, Ochieng M, Onyango C,
222 Panchalingam S, Kalam A, Aziz F, Qureshi S, Ramamurthy T, Roberts JH, Saha D, Sow
223 SO, Stroup SE, Sur D, Tamboura B, Taniuchi M, Tenant SM, Toema D, Wu Y, Zaidi A,
224 Nataro JP, Kotloff KL, Levine MM, Houpt ER. 2016. Use of quantitative molecular
225 diagnostic methods to identify causes of diarrhoea in children: a reanalysis of the GEMS

226 case-control study. *Lancet* 388:1291–1301.

227 6. Pavlinac PB, Platts-Mills JA, Tickell KD, Liu J, Juma J, Kabir F, Nkeze J, Okoi C,
228 Operario DJ, Uddin J, Ahmed S, Alonso PL, Antonio M, Becker SM, Breiman RF,
229 Faruque ASG, Fields B, Gratz J, Haque R, Hossain A, Hossain MJ, Jarju S, Qamar F,
230 Iqbal NT, Kwambana B, Mandomando I, McMurry TL, Ochieng C, Ochieng JB, Ochieng
231 M, Onyango C, Panchalingam S, Kalam A, Aziz F, Qureshi S, Ramamurthy T, Roberts
232 JH, Saha D, Sow SO, Stroup SE, Sur D, Tamboura B, Taniuchi M, Tennant SM, Roose A,
233 Toema D, Wu Y, Zaidi A, Nataro JP, Levine MM, Houpt ER, Kotloff KL. 2021. The
234 Clinical Presentation of Culture-positive and Culture-negative, Quantitative Polymerase
235 Chain Reaction (qPCR)-Attributable Shigellosis in the Global Enteric Multicenter Study
236 and Derivation of a Shigella Severity Score: Implications for Pediatric Shig. *Clin Infect*
237 *Dis* 73:E569–E579.

238 7. Hawkey J, Monk JM, Billman-Jacobe H, Palsson B, Holt KE. 2020. Impact of insertion
239 sequences on convergent evolution of *Shigella* species. *PLOS Genet* 16:e1008931.

240 8. Byers KA, Booker TR, Combs M, Himsworth CG, Munshi-South J, Patrick DM,
241 Whitlock MC. 2021. Using genetic relatedness to understand heterogeneous distributions
242 of urban rat-associated pathogens. *Evol Appl* 14:198–209.

243 9. Zhang X, Payne M, Nguyen T, Kaur S, Lan R. 2021. Cluster-specific gene markers
244 enhance *Shigella* and enteroinvasive *Escherichia coli* *in silico* serotyping. *Microb*
245 *Genomics* 7.

246 10. Fan J, Huang S, Chorlton SD. 2021. BugSeq: a highly accurate cloud platform for long-
247 read metagenomic analyses. *BMC Bioinformatics* 22:160.

248 11. Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, Kuhn K, Yuan
249 J, Polevikov E, Smith TPL, Pevzner PA. 2020. metaFlye: scalable long-read metagenome
250 assembly using repeat graphs. *Nat Methods* 17:1103–1110.

251 12. Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome
252 assembly from long uncorrected reads. *Genome Res* 27:737–746.

253 13. Huang Y-T, Liu P-Y, Shih P-W. 2021. Homopolish: a method for the removal of
254 systematic errors in nanopore sequencing by homologous polishing. *Genome Biol* 22:95.

255 14. Dennis TPW, Mable BK, Brunelle B, Devault A, Carter RW, Ling CL, Mmbaga BT,
256 Halliday JEB, Oravcova K, Forde TL. 2022. Target-enrichment sequencing yields
257 valuable genomic data for challenging-to-culture bacteria of public health importance.
258 *Microb Genomics* 8:1–7.

259 15. Wagner A, de la Chaux N. 2008. Distant horizontal gene transfer is rare for multiple
260 families of prokaryotic insertion sequences. *Mol Genet Genomics* 280:397–408.

261

262 Table 1. Primer and probe sequences for the Insertion Sequence PCR (IS-PCR) and *ipaH* PCR

IS-PCR1	F	GCC ACG CAA GCA CCT TAA AA
	R	ACC GAA AAG CAT CAC CAC CA
	Probe	FAM-CGA TTA TTA ATG GCG GGT AAT GCT GC-BHQ2
	Ampli con Size	206bp
IS-PCR2	F	ACT TTG AAC TGG CTG GGG TC
	R	CGC TTG AGC TTT CGC AGT TT
	Probe	FAM-ATA TCG CCG -ZEN- CAT CCA TTC CAG GAA GG-IABkFQ
	Ampli con Size	212bp
ipaH-PCR	F	CCT TTT CCG CGT TCC TTG A
	R	CGG AAT CCG GAG GTA TTG C
	Probe	FAM-CGC CTT TCC -ZEN-GAT ACC GTC TCT GCA-IABkFQ

263

264

265 Table 2. Cycle threshold (Ct) values from direct amplification of rat intestines associated with a
266 community *Shigella flexneri* outbreak for *ipaH* and two unique insertion sequences (IS-PCR1
267 and IS-PCR2)

Rat	<i>ipaH</i>	IS-PCR1	IS-PCR2
1	33	35	41
2	36	Negative	Negative
3	29	35	46

268