2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

Whole genome sequencing identifies independent outbreaks of Shigellosis in 2010 and 2011 in La Pampa Province, Argentina Isabel Chinen <sup>1</sup>, Marcelo Galas <sup>1</sup>, Ezequiel Tuduri <sup>1</sup>, Maria Rosa Viñas <sup>1</sup>, Carolina Carbonari <sup>1</sup>, Anabella Della Gaspera <sup>1</sup>, Daniela Nápoli <sup>1</sup>, David M Aanensen <sup>2,3</sup>, Silvia Argimón <sup>2</sup>, Nicholas R Thomson 4, Darren Hughes 5, Stephen Baker 6, Caterina Guzmán-Verri 7, Matthew TG Holden 8, Alejandra M Abdala 9, Lucia P Alvarez 9, Beatriz Alvez 10, Rosana Barros 11, Shirley Budall 12, Constanza Campano 13, Luciana S Chamosa 14, Paul Cheddie 15, Daniel Cisterna 1, Denise De Belder <sup>1</sup>, Milena Dropa <sup>16</sup>, David Durand <sup>17</sup>, Alan Elena <sup>14</sup>, Gustavo Fontecha <sup>18</sup>, Claudia Huber 19, Ana Paula Lemos 20, Luciano Melli 21, Roxana Elizabeth Paul 1, Lesly Suarez 22, Julian Torres Flores 22 and Josefina Campos 1\* <sup>1</sup> Instituto Nacional de Enfermedades Infecciosas, ANLIS, Buenos Aires, Argentina <sup>2</sup> Centre for Genomic Pathogen Surveillance, Hinxton, Cambridge, United Kingdom <sup>3</sup> Imperial College London, London, United Kingdom <sup>4</sup> The Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom <sup>5</sup> The Wellcome Genome Campus Advanced Courses, Hinxton, Cambridge, United Kingdom <sup>6</sup> The Hospital for Tropical Diseases, Wellcome Trust Major Overseas Programme, Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam <sup>7</sup> Universidad Nacional, 3000 Heredia, Costa Rica 8 The University of St. Andrews, St. Andrews, Scotland <sup>9</sup> National Institute for Agricultural Technology (INTA)CONICET, Argentina <sup>10</sup> Universidad Central de Venezuela, Caracas, Venezuela <sup>11</sup> Fluminense Federal University, Niterio, Brazil <sup>12</sup> University of the West Indies, Mona campus, Kingston, Jamaica

13 Institute of Public Health of Chile, Santiago, Chile

<sup>14</sup> Universidad de Buenos Aires, Buenos Aires, Argentina

15 University of Guyana, Georgetown, Guyana <sup>16</sup> University of Sao Paulo, Sao Paulo, Brazil <sup>17</sup> Universidad Peruana Cayetano Heredia, Lima, Perú <sup>18</sup> Microbiology Research Institute, Universidad Nacional Autónoma de Honduras, Tegucigalpa, Honduras 19 Laboratorio Central de Salud Pública, Asunción, Paraguay <sup>20</sup> Adolfo Lutz Institute, Sao Paulo, Brazil <sup>21</sup> Universidad Nacional de San Martín, Buenos Aires, Argentina <sup>22</sup> Cayetano Heredia University, Lima, Perú \* Corresponding Author. Josefina Campos, Genomics and Bioinformatics Platform, INEI-ANLIS Dr. Carlos G. Malbran, Avenida Dr Velez Sarsfield, C1282AFF, Buenos Aires, Argentina. Email: jcampos@anlis.gov.ar 

**Abstract** 

54

55

56

57

58

59

60

61

62

63

64

65

66

67

68

69

70

71

72

73

74

75

76

77

78

Shigella sonnei is an emergent cause of diarrheal disease in middle-income countries. The organism causes endemic disease and is also associated with sporadic outbreaks in susceptible populations. In 2010 and 2011 there were two suspected outbreaks of diarrheal disease caused by S. sonnei in La Pampa province in central Argentina. Aiming to confirm these as outbreaks and provide insight into the relationship of the strains causing these infections we combined antimicrobial susceptibility testing and pulsed field gel electrophoresis (PFGE) with whole genome sequencing (WGS). Antimicrobial susceptibility testing suggested the two events were unrelated; organisms isolated in 2010 exhibited resistance to trimethoprim sulphate whereas the 2011 S. sonnei were non-susceptible against ampicillin, trimethoprim sulphate and cefpodoxime. PFGE profiling confirmed the likelihood of two independent outbreaks, separating the isolates into two main XbaI restriction profiles. We additionally performed WGS on 17 isolates associated with these outbreaks. The resulting phylogeny confirmed the PFGE structure and separated the organisms into two comparatively distantly related clones. Antimicrobial resistant genes were common, and the presence of an OXA-1 was likely associated with resistance to cefpodoxime in the second outbreak. We additionally identified novel horizontally transferred genetic material that may impinge on the pathogenic phenotype of the infecting strains. Our study shows that even with a lack of supporting routine data WGS is an indispensible method for the tracking and surveillance of bacterial pathogens during outbreaks and is becoming a vital tool for the monitoring of antimicrobial resistant strains of S. sonnei.

80

81

82

83

84

85

86

87

88

89

90

91

92

93

94

95

96

97

98

99

100

101

102

103

Report Dysenteric diarrhea caused by members of the bacterial genus Shigella (comprised of the species S. flexneri, S. sonnei, S. boydii and S. dysenteriae) remains an on-going public health issue in many industrializing countries. It is estimated that the global burden of disease caused by *Shigella* spp. is  $\sim$ 125 million cases annually [1], the majority of these cases arise in children aged under five years. The Global Enteric Multicenter Study (GEMS), a casecontrol study of paediatric diarrheal disease conducted in Africa and Asia, found that enterotoxigenic Escherichia coli (ETEC) and Shigella were the two most common bacterial agents of diarrhea in sub-Saharan Africa and South Asia [2,3]. Notably, Shigella spp. were the most prevalent pathogen among children between 24 and 59 months old [3]. Of the four *Shigella* species *S. flexneri* and *S. sonnei* are responsible for the vast majority of the global burden of disease. Traditionally, S. sonnei has been the predominant cause of bacterial dysentery in industrialized countries, whereas S. flexneri has been considered to be associated with endemic disease and travel to lower income countries [4]. However, this trend is changing as S. sonnei is now emerging as a problem in lower middle-income countries, seemingly replacing S. flexneri as the leading cause of dysentery in these locations [5]. This trend has also been observed in parts of Latin America [6,7], roughly correlating with improvements in sanitation, water quality and, potentially, a fall in passive immunity against *S. sonnei* via a decline in other bacteria associated with poor water quality [8] Argentina is a middle-income country in South America with endemic Shigellosis [9], the number of officially reported cases of Shigellosis in 2014 was 4,116. This represents a comparatively high proportion of the 6,200 cases of confirmed the cases of bacterial diarrhea in 2014. Shigella outbreaks occur sporadically and rapidly and are frequently

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

120

121

122

123

124

125

126

127

128

associated with changes in antimicrobial susceptibility [10.11]. Aiming to better understand the dynamics of Shigellosis in Argentina we gathered bacterial isolates from two suspected outbreaks of Shigellosis investigated by the public health authorities in Argentina between 2010 and 2011. The two temporally independent events were attributed to S. sonnei and occurred within the La Pampa province in the central region of the country. In this retrospective study we combined available epidemiological data and microbiological data with Pulsed Field Gel Electrophoresis (PFGE) and whole genome sequencing (WGS) of the S. sonnei isolates from the suspected outbreaks in 2010 and 2011 to investigate their relatedness. We also analysed the discriminatory ability of WGS compared to PFGE, the current international gold-standard method for public health strain tracking. In December 2009 the Gobernador Centeno hospital in the city of General Pico reported an increased number of cases of diarrhea above the expected endemic rate, an outbreak was suspected. The first Shiaella (confirmed by standard microbiological methods to be S. sonnei) was isolated on the 7th January 2010; the last culture confirmed case of S. sonnei was on the 26th February 2010. The cases were distributed throughout (i.e. no apparent case clustering) General Pico. There were 26 reported cases, of which detailed microbiological data was available on nine. Of the 26 reported cases, 13 were in children aged between 0-5 years, and 13 cases were aged between 6-69 years (median: five years); 10 cases were female and 16 were male. No epidemiological association was recorded between cases, apart from a potential cluster in a single household (n=4 cases). The most common disease presentations were diarrhea with blood and mucus (13/26; 50%) and diarrhea with blood without mucus (11/26; 42%). Of the nine available S. sonnei organisms isolated during the potential 2010 outbreak, all had an identical antimicrobial susceptibility patterns by disc diffusion [12], exhibiting susceptibility against ampicillin, ciprofloxacin, nitrofurantoin,

130

131

132

133

134

135

136

137

138

139

140

141

142

143

144

145

146

147

148

149

150

151

152

153

fosfomycin, naladixic acid and cefpodoxime and non-susceptibility against trimethoprim sulphate. The second potential outbreak occurred the city of Castex, also in the province of La Pampa (60 Km from General Pico) in the summer of 2011 (3rd February 2011 to 30th March 2011). No supporting epidemiological data were available for this potential outbreak and six S. sonnei were isolated. An equal proportion of males and females were infected and the patient age range was 5-26 years (median: eight years). The antimicrobial susceptibility profile of the organisms demonstrated that all organisms were non-susceptible against ampicillin, trimethoprim sulphate and cefpodoxime. Additional laboratory testing suggested that all organisms in this second potential outbreak exhibited AmpC production. To confirm the likelihood of outbreaks and to investigate the temporal and spatial relationship between organisms we performed PFGE after Xbal digestion on 17 available isolates from 2010 (n=9) and 2011 (n=7) and one additional contextual strain isolated in General Pico in 2013 using standardized PulseNet protocols as previously described (Figure 1) [13]. The Xbal PFGE generated nine differing restriction patterns that could be grouped into two major groups (ARJ16X01.0086 and ARJ16X01.0318) that correlated precisely with their year of isolation; ARJ16X01.0086 is the most frequent pattern described in Argentina. These major restriction patterns differed by six fragments and had 85% pattern similarity. Additional PFGE with BlnI (again following standardized PulseNet protocols [13]) methods of on a limited subsample confirmed this grouping, signifying independent outbreaks likely caused by two differing clones of *S. sonnei* that could be distinguished by their antimicrobial susceptibility patterns. These cases clusters were additionally confirmed using the SatScan function in WHONET software.

155

156

157

158

159

160

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

PFGE does not provide sufficient resolution for phylogenetic inference, to understand the genetic relationship between the organisms from the two outbreaks we performed WGS on nine S. sonnei isolated in 2010, seven S. sonnei isolated in 2011 and the single contextual isolate from 2013. We identified single nucleotide polymorphisms (SNPs) in comparison to a Chinese reference strain (Ss046, accession number CP000038[14]), as previously described [11,15,16], and constructed a maximum likelihood phylogeny of the 18 strains, identifying approximately 1,500 variable nucleotide sites (Figure 2). Our data confirmed that the outbreaks were associated with two differing, distantly related clones of *S. sonnei*. The first clone (-10 suffix in Figure 2) was comprised of the 2010 isolates; six were highly related, with three remaining isolates in the same group but located on longer branches. The second clone (-11 suffix in Figure 2) contained all six isolates from 2011; these isolates were almost identical, containing less than 10 nucleotide substitutions across their genomes. An additional isolate from 2011 (1193-11) lay outside this group and was deemed not to part of the same clonal outbreak. Notably, the 2011 clone could be distinguished by non-susceptibility against cefpodoxime (yellow nodes in Figure 2) (Data accessible and viewable at http://microreact.org/project/EkJeuWfx-). We next assembled the genome sequences from the two independent outbreaks to identify additional horizontally transferred genetic material that may be associated with each of the clones and to classify the genes associated with changes in antimicrobial susceptibility. We found that the -10 clone contained a dfrA1 gene, which is associated with resistance against trimethoprim sulphate. Further, we identified genes associated with resistance to additional antimicrobials that were not tested, including streptomycin (strAB), tetracycline (tetAR) and sulphonomides (sulII). The -11 clone also contained a dfr gene (A5) and genes associated

180

181

182

183

184

185

186

187

188

189

190

191

192

193

194

195

196

197

198

199

200

201

202

203

with resistance against chloramphenicol (catA1), streptomycin (strAB), tetracycline (tetB) and sulphonomides (sullI). We also identified several AmpC β-lactamase genes potentially explaining the non-susceptibility against cefpodoxime including CMY and the Extended Spectrum Beta Lactamase (ESBL) gene, OXA-1, Further, in the -11 clone we identified and assembled a large (>90 Kb) plasmid that exhibited substantial homology and synteny to the recently described 96 Kb p12-4374\_96 plasmid in Salmonella Heidelberg [17] (accession number: CP012929). This plasmid, not previously described in Shigella, encoded a multitude of potentially interesting functions including a conjugation system, a type IVb pilus and the ethanolamine utilization protein, EutE. Here we have combined traditional methods for tracking bacterial pathogens (antimicrobial susceptibility testing and PFGE) and combined them with WGS to evaluate two potential outbreaks of Shigellosis in a single province in 2010 and 2011 in Argentina. Our data suggests that there were two independent outbreaks of S. sonnei induced diarrhea in 2010 and 2011, finding that the organisms causing these case clusters were distantly related to each other. This was somewhat unexpected given the geographical proximity of these two locations and signifies that multiple clones of S. sonnei are likely circulating in Argentina, several of which have outbreak potential. We found that the antimicrobial susceptibility profile was sufficient to distinguish between these outbreaks, providing an almost perfect temporal correlation with cefpodoxime resistance. This relationship was further confirmed by PFGE, the current gold standard for strain tracking in such scenarios in Argentina [18]. However, PFGE additionally over predicted the variability within the genomic structures, identifying several banding patterns within the specific clones. In this particular investigation WGS augmented the findings of the conventional approaches

205

206

207

208

209

210

211

212

213

214

215

216

217

218

219

220

221

222

223

224

225

226

227

228

and provided new insight into these outbreaks. Firstly, the phylogenetic inference, which has become standard for WGS of S. sonnei [11,15,16], permitted us an exquisite view of the relationship between and within the outbreaks, eventually confirming the two outbreaks. Further, assembly of the genome sequences identified the presence of the range antimicrobial resistance genes, predicting resistance to additional antimicrobials that were not susceptibility tested. These data permitted us to detect the ESBL gene OXA-1 [19], which we hypothesized to be associated with resistance against the third generation cephalosporin, cefpodoxime. ESBL genes are becoming more commonly reported in Shigella in Asia [20]. Our data predict that this concerning phenomenon is additionally occurring in Latin America via differing determinants. Oral third generation cephalosporins are one of the current mainstays of treatment for Shigellosis in Argentina, we recommend further genomic surveillance in this region to detect circulating beta lactamase genes. The assembled genome sequences additionally identified further novel sequences encoding potential virulence associated loci that may have a phenotypic effect during infection. These novel genes require conformation and additional experimentation to confirm their role in disease. Our study contains some limitations including a lack of strain diversity for better phylogenetic inference and a lack of epidemiological data. However, our investigation of these outbreaks represents a "real life" scenario, where limited data hamper contextualization. Here we show that even with a lack of supporting routine data WGS becomes an indispensible method for the tracking and surveillance of bacterial pathogens during outbreaks.

230

231

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250

251

252

253

Acknowledgements This work was conducted as a component of the genomics and epidemiological surveillance of bacterial pathogens course held from the 17-22 April 2011 in the Dr. Carlos G. Malbran Institute in Buenos Aires, Argentina, We wish to acknowledge the Wellcome Genome Campus Advanced Courses for providing financial and administrative support for the course that supported this investigation, including travel scholarships for attending students. We additionally wish to acknowledge the Dr. Carlos G. Malbran Institute for providing access to these data for training and publication purposes. **Declaration of interests** The authors declare no competing interests. References Bardhan P. Faruque a SG. Naheed A. Sack D a (2010) Decrease in shigellosis-related 1. deaths without Shigella spp.-specific interventions, Asia. Emerg Infect Dis 16: 1718-1723. 2. Kotloff KL, Blackwelder WC, Nasrin D, Nataro IP, Farag TH, et al. (2012) The Global Enteric Multicenter Study (GEMS) of diarrheal disease in infants and young children in developing countries: epidemiologic and clinical methods of the case/control study. Clin Infect Dis 55 Suppl 4: S232-S245. 3. Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH, et al. (2013) Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. Lancet 382: 209-222. 4. Keusch GT (2009) Bacterial Infections of Humans. Brachman PS, Abrutyn E, editors

254		Boston, MA: Springer US.
255	5.	Thompson CN, Duy PT, Baker S (2015) The Rising Dominance of Shigella sonnei: An
256		Intercontinental Shift in the Etiology of Bacillary Dysentery. PLoS Negl Trop Dis 9:
257		e0003708.
258	6.	Fullá N, Prado V, Durán C, Lagos R, Levine MM (2005) Surveillance for antimicrobial
259		resistance profiles among Shigella species isolated from a semirural community in
260		the northern administrative area of Santiago, Chile. Am J Trop Med Hyg 72: 851–854.
261	7.	Sousa MÂB, Mendes EN, Collares GB, Péret-Filho LA, Penna FJ, et al. (2013) Shigella in
262		Brazilian children with acute diarrhoea: prevalence, antimicrobial resistance and
263		virulence genes. Memorias Inst Oswaldo Cruz 108: 30–35.
264	8.	Shepherd JG, Wang L, Reeves P (2000) Comparison of O-antigen gene clusters of
265		Echerichia coli (Shigella) Sonnei and Plesiomonas shigelloides O17: Sonnei gained its
266		current plasmid-borne O-antigen genes from P. shigelloides in a recent event. Infect
267		Immun 68: 6056-6061.
268	9.	Casabonne C, González A, Aquili V, Balagué C (2016) Prevalence and virulence factors
269		of Shigella spp. isolated from patients with diarrhoea in Rosario, Argentina. Jpn J
270		Infect Dis.
271	10.	Huang I, Chiu C, Wang M, Wu C, Hsieh K, et al. (2005) Outbreak of Dysentery
272		Associated with Ceftriaxone-Resistant Shigella sonnei: First Report of Plasmid-
273		Mediated CMY-2-Type AmpC \L -Lactamase Resistance in S . sonnei. Society 43: 2608−
274		2612.
275	11.	Holt KE, Thieu Nga TV, Thanh DP, Vinh H, Kim DW, et al. (2013) Tracking the
276		establishment of local endemic populations of an emergent enteric pathogen. Proc
277		Natl Acad Sci U S A 110: 17522–17527.
278	12.	Clinical and Laboratory Standards Institute (2014) Performance Standards for

279		Antimicrobial Susceptibility Testing: Twenty-Fourth Informational Supplement.
280		Wayne, PA, USA: CLSI document M100-S24.
281	13.	Viñas MR, Tuduri E, Galar A, Yih K, Pichel M, et al. (2013) Laboratory-based
282		prospective surveillance for community outbreaks of Shigella spp. in Argentina. PLoS
283		Negl Trop Dis 7:
284	14.	Liang S, Watanabe H, Terajima J, Li C, Liao J, et al. (2007) Multilocus Variable-Number
285		Tandem-Repeat Analysis for Molecular Typing of Shigella sonnei. Society 45: 3574–
286		3580.
287	15.	Holt KE, Baker S, Weill F-X, Holmes EC, Kitchen A, et al. (2012) Shigella sonnei
288		genome sequencing and phylogenetic analysis indicate recent global dissemination
289		from Europe. Nat Genet 44: 1056–1059.
290	16.	Baker KS, Dallman TJ, Ashton PM, Day M, Hughes G, et al. (2015) Intercontinental
291		dissemination of azithromycin-resistant shigellosis through sexual transmission: a
292		cross-sectional study. Lancet Infect Dis.
293	17.	Labbé G, Edirmanasinghe R, Ziebell K, Nash JHE, Bekal S, et al. (2016) Complete
294		Genome and Plasmid Sequences of Three Canadian Isolates of Salmonella enterica
295		subsp. enterica Serovar Heidelberg from Human and Food Sources. Genome Announc
296		4.
297	18.	Pichel M, González Fraga S, Terragno R, Mulki J, Gentile A, et al. (2007) Short report:
298		analysis of clonal relationship among Shigella sonnei isolates circulating in
299		Argentina. Epidemiol Infect 135: 681–687.
300	19.	Paterson DL, Bonomo RA (2005) Extended-Spectrum -Lactamases: a Clinical Update.
301		Clin Microbiol Rev 18: 657–686.
302	20.	Nguyen NTK, Ha V, Tran NVT, Stabler R, Pham DT, et al. (2010) The sudden
303		dominance of blaCTX-M harbouring plasmids in Shigella spp. Circulating in Southern

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

Vietnam. PLoS Negl Trop Dis 4: e702. Figure 1. The relationship between Shigella sonnei isolated in independent outbreaks in La Pampa province. Argentina in 2010 and 2011 Dendogram based on PFGE profile after XbaI digestion of Shigella sonnei isolated from stool samples in 2010 (General Pico) and 2011 (Castex). One additional isolate from 2013 (General Pico) was included as contextual strain. Information regarding the strain ID, the date of isolation and the digestion pattern (according to PulseNet Latinoamerica) are provided. **Figure 2**. The phylogenetic relationship of *Shigella sonnei* isolated in independent outbreaks in La Pampa province, Argentina in 2010 and 2011 Unrooted maximum likelihood tree constructed using approximately 1,500 variable nucleotide sites from nine organisms isolated in 2010, seven isolates from 2010 and a single isolate from 2013 in La Pampa province, Argentina. S. sonnei Ss046 was added as the reference strain. Tree was viewed in microreact and nodes are labelled with the strain name and the year of isolation suffix collared according to susceptibly against cefpodoxime susceptibility (orange; susceptible, yellow, non-susceptible).



