Whole genome sequencing identifies independent outbreaks of Shigellosis in 2010 and 2011 in La Pampa Province, Argentina

Isabel Chinen [1], Marcelo Galas [1], Ezequiel Tuduri [1], Maria Rosa Viñas [1], Carolina Carbonari [1], Anabella Della Gaspera [1], Daniela Nápoli [1], David M Aanensen [2,3], Silvia Argimón [2], 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 [1], Milena Dropa [16], David Durand [17], Alan Elena [14], Gustavo Fontecha [18], 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*]

[1] Instituto Nacional de Enfermedades Infecciosas, ANLIS, Buenos Aires, Argentina

[2] Centre for Genomic Pathogen Surveillance, Hinxton, Cambridge, United Kingdom

[3] Imperial College London, London, United Kingdom

[4] The Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom

[5] The Wellcome Genome Campus Advanced Courses, Hinxton, Cambridge, United Kingdom

[6] The Hospital for Tropical Diseases, Wellcome Trust Major Overseas Programme, Oxford University Clinical Research Unit, Ho Chi Minh City, Vietnam

[7] Universidad Nacional, 3000 Heredia, Costa Rica

[8] The University of St. Andrews, St. Andrews, Scotland

[9] National Institute for Agricultural Technology (INTA)CONICET, Argentina

[10] Universidad Central de Venezuela, Caracas, Venezuela

[11] Fluminense Federal University, Niterio, Brazil

[12] University of the West Indies, Mona campus, Kingston, Jamaica

[13] Institute of Public Health of Chile, Santiago, Chile

[14] Universidad de Buenos Aires, Buenos Aires, Argentina

[15] University of Guyana, Georgetown, Guyana

[16] University of Sao Paulo, Sao Paulo, Brazil

[17] Universidad Peruana Cayetano Heredia, Lima, Perú

[18] Microbiology Research Institute, Universidad Nacional Autónoma de Honduras, Tegucigalpa, Honduras

[19] Laboratorio Central de Salud Pública, Asunción, Paraguay

[20] Adolfo Lutz Institute, Sao Paulo, Brazil

[21] Universidad Nacional de San Martín, Buenos Aires, Argentina

[22] 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

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*.

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 ~125 million cases annually [1], the majority of these cases arise in children aged under five years. The Global Enteric Multicenter Study (GEMS), a case-control 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

104 associated with changes in antimicrobial susceptibility [10,11]. Aiming to better understand
105 the dynamics of Shigellosis in Argentina we gathered bacterial isolates from two suspected
106 outbreaks of Shigellosis investigated by the public health authorities in Argentina between
107 2010 and 2011. The two temporally independent events were attributed to *S. sonnei* and
108 occurred within the La Pampa province in the central region of the country. In this
109 retrospective study we combined available epidemiological data and microbiological data
110 with Pulsed Field Gel Electrophoresis (PFGE) and whole genome sequencing (WGS) of the *S.*
111 *sonnei* isolates from the suspected outbreaks in 2010 and 2011 to investigate their
112 relatedness. We also analysed the discriminatory ability of WGS compared to PFGE, the
113 current international gold-standard method for public health strain tracking.
114
115 In December 2009 the Gobernador Centeno hospital in the city of General Pico reported an
116 increased number of cases of diarrhea above the expected endemic rate, an outbreak was
117 suspected. The first *Shigella* (confirmed by standard microbiological methods to be *S.*
118 *sonnei*) was isolated on the 7th January 2010; the last culture confirmed case of *S. sonnei* was
119 on the 26th February 2010. The cases were distributed throughout (i.e. no apparent case
120 clustering) General Pico. There were 26 reported cases, of which detailed microbiological
121 data was available on nine. Of the 26 reported cases, 13 were in children aged between 0-5
122 years, and 13 cases were aged between 6-69 years (median: five years); 10 cases were
123 female and 16 were male. No epidemiological association was recorded between cases,
124 apart from a potential cluster in a single household (n=4 cases). The most common disease
125 presentations were diarrhea with blood and mucus (13/26; 50%) and diarrhea with blood
126 without mucus (11/26; 42%). Of the nine available *S. sonnei* organisms isolated during the
127 potential 2010 outbreak, all had an identical antimicrobial susceptibility patterns by disc
128 diffusion [12], exhibiting susceptibility against ampicillin, ciprofloxacin, nitrofurantoin,

129 fosfomycin, naladixic acid and cefpodoxime and non-susceptibility against trimethoprim
130 sulphate.
131
132 The second potential outbreak occurred the city of Castex, also in the province of La Pampa
133 (60 Km from General Pico) in the summer of 2011 (3rd February 2011 to 30th March 2011).
134 No supporting epidemiological data were available for this potential outbreak and six *S.*
135 *sonnei* were isolated. An equal proportion of males and females were infected and the
136 patient age range was 5-26 years (median: eight years). The antimicrobial susceptibility
137 profile of the organisms demonstrated that all organisms were non-susceptible against
138 ampicillin, trimethoprim sulphate and cefpodoxime. Additional laboratory testing suggested
139 that all organisms in this second potential outbreak exhibited AmpC production.
140
141 To confirm the likelihood of outbreaks and to investigate the temporal and spatial
142 relationship between organisms we performed PFGE after XbaI digestion on 17 available
143 isolates from 2010 (n=9) and 2011 (n=7) and one additional contextual strain isolated in
144 General Pico in 2013 using standardized PulseNet protocols as previously described (Figure
145 1) [13]. The XbaI PFGE generated nine differing restriction patterns that could be grouped
146 into two major groups (ARJ16X01.0086 and ARJ16X01.0318) that correlated precisely with
147 their year of isolation; ARJ16X01.0086 is the most frequent pattern described in Argentina.
148 These major restriction patterns differed by six fragments and had 85% pattern similarity.
149 Additional PFGE with BlnI (again following standardized PulseNet protocols [13]) methods
150 of on a limited subsample confirmed this grouping, signifying independent outbreaks likely
151 caused by two differing clones of *S. sonnei* that could be distinguished by their antimicrobial
152 susceptibility patterns. These cases clusters were additionally confirmed using the SatScan
153 function in WHONET software.

154

155 PFGE does not provide sufficient resolution for phylogenetic inference, to understand the
156 genetic relationship between the organisms from the two outbreaks we performed WGS on
157 nine *S. sonnei* isolated in 2010, seven *S. sonnei* isolated in 2011 and the single contextual
158 isolate from 2013. We identified single nucleotide polymorphisms (SNPs) in comparison to
159 a Chinese reference strain (Ss046, accession number CP000038[14]), as previously
160 described [11,15,16], and constructed a maximum likelihood phylogeny of the 18 strains,
161 identifying approximately 1,500 variable nucleotide sites (Figure 2). Our data confirmed
162 that the outbreaks were associated with two differing, distantly related clones of *S. sonnei*.
163 The first clone (-10 suffix in Figure 2) was comprised of the 2010 isolates; six were highly
164 related, with three remaining isolates in the same group but located on longer branches.
165 The second clone (-11 suffix in Figure 2) contained all six isolates from 2011; these isolates
166 were almost identical, containing less than 10 nucleotide substitutions across their
167 genomes. An additional isolate from 2011 (1193-11) lay outside this group and was deemed
168 not to part of the same clonal outbreak. Notably, the 2011 clone could be distinguished by
169 non-susceptibility against cefpodoxime (yellow nodes in Figure 2) (Data accessible and
170 viewable at http://microreact.org/project/EkJeuWfx-).

171

172 We next assembled the genome sequences from the two independent outbreaks to identify
173 additional horizontally transferred genetic material that may be associated with each of the
174 clones and to classify the genes associated with changes in antimicrobial susceptibility. We
175 found that the -10 clone contained a *dfrA1* gene, which is associated with resistance against
176 trimethoprim sulphate. Further, we identified genes associated with resistance to additional
177 antimicrobials that were not tested, including streptomycin (*strAB*), tetracycline (*tetAR*) and
178 sulphonomides (*sulII*). The -11 clone also contained a *dfr* gene (A5) and genes associated

179 with resistance against chloramphenicol (*catA1*), streptomycin (*strAB*), tetracycline (*tetB*)
180 and sulphonomides (*sulII*). We also identified several AmpC β-lactamase genes potentially
181 explaining the non-susceptibility against cefpodoxime including CMY and the Extended
182 Spectrum Beta Lactamase (ESBL) gene, OXA-1. Further, in the -11 clone we identified and
183 assembled a large (>90 Kb) plasmid that exhibited substantial homology and synteny to the
184 recently described 96 Kb p12-4374_96 plasmid in *Salmonella* Heidelberg [17] (accession
185 number: CP012929). This plasmid, not previously described in *Shigella*, encoded a
186 multitude of potentially interesting functions including a conjugation system, a type IVb
187 pilus and the ethanolamine utilization protein, EutE.
188
189 Here we have combined traditional methods for tracking bacterial pathogens
190 (antimicrobial susceptibility testing and PFGE) and combined them with WGS to evaluate
191 two potential outbreaks of Shigellosis in a single province in 2010 and 2011 in Argentina.
192 Our data suggests that there were two independent outbreaks of *S. sonnei* induced diarrhea
193 in 2010 and 2011, finding that the organisms causing these case clusters were distantly
194 related to each other. This was somewhat unexpected given the geographical proximity of
195 these two locations and signifies that multiple clones of *S. sonnei* are likely circulating in
196 Argentina, several of which have outbreak potential. We found that the antimicrobial
197 susceptibility profile was sufficient to distinguish between these outbreaks, providing an
198 almost perfect temporal correlation with cefpodoxime resistance. This relationship was
199 further confirmed by PFGE, the current gold standard for strain tracking in such scenarios
200 in Argentina [18]. However, PFGE additionally over predicted the variability within the
201 genomic structures, identifying several banding patterns within the specific clones.
202
203 In this particular investigation WGS augmented the findings of the conventional approaches

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.

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

1. Bardhan P, Faruque a SG, Naheed A, Sack D a (2010) Decrease in shigellosis-related deaths without Shigella spp.-specific interventions, Asia. Emerg Infect Dis 16: 1718–1723.
2. Kotloff KL, Blackwelder WC, Nasrin D, Nataro JP, 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

Boston, MA: Springer US.

5. Thompson CN, Duy PT, Baker S (2015) The Rising Dominance of Shigella sonnei: An Intercontinental Shift in the Etiology of Bacillary Dysentery. PLoS Negl Trop Dis 9: e0003708.

6. Fullá N, Prado V, Durán C, Lagos R, Levine MM (2005) Surveillance for antimicrobial resistance profiles among Shigella species isolated from a semirural community in the northern administrative area of Santiago, Chile. Am J Trop Med Hyg 72: 851–854.

7. Sousa MÂB, Mendes EN, Collares GB, Péret-Filho LA, Penna FJ, et al. (2013) Shigella in Brazilian children with acute diarrhoea: prevalence, antimicrobial resistance and virulence genes. Memorias Inst Oswaldo Cruz 108: 30–35.

8. Shepherd JG, Wang L, Reeves P (2000) Comparison of O-antigen gene clusters of Echerichia coli (Shigella) Sonnei and Plesiomonas shigelloides O17: Sonnei gained its current plasmid-borne O-antigen genes from P. shigelloides in a recent event. Infect Immun 68: 6056–6061.

9. Casabonne C, González A, Aquili V, Balagué C (2016) Prevalence and virulence factors of Shigella spp. isolated from patients with diarrhoea in Rosario, Argentina. Jpn J Infect Dis.

10. Huang I, Chiu C, Wang M, Wu C, Hsieh K, et al. (2005) Outbreak of Dysentery Associated with Ceftriaxone-Resistant Shigella sonnei : First Report of Plasmid-Mediated CMY-2-Type AmpC $^{N}_{L}$ -Lactamase Resistance in S . sonnei. Society 43: 2608–2612.

11. Holt KE, Thieu Nga TV, Thanh DP, Vinh H, Kim DW, et al. (2013) Tracking the establishment of local endemic populations of an emergent enteric pathogen. Proc Natl Acad Sci U S A 110: 17522–17527.

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

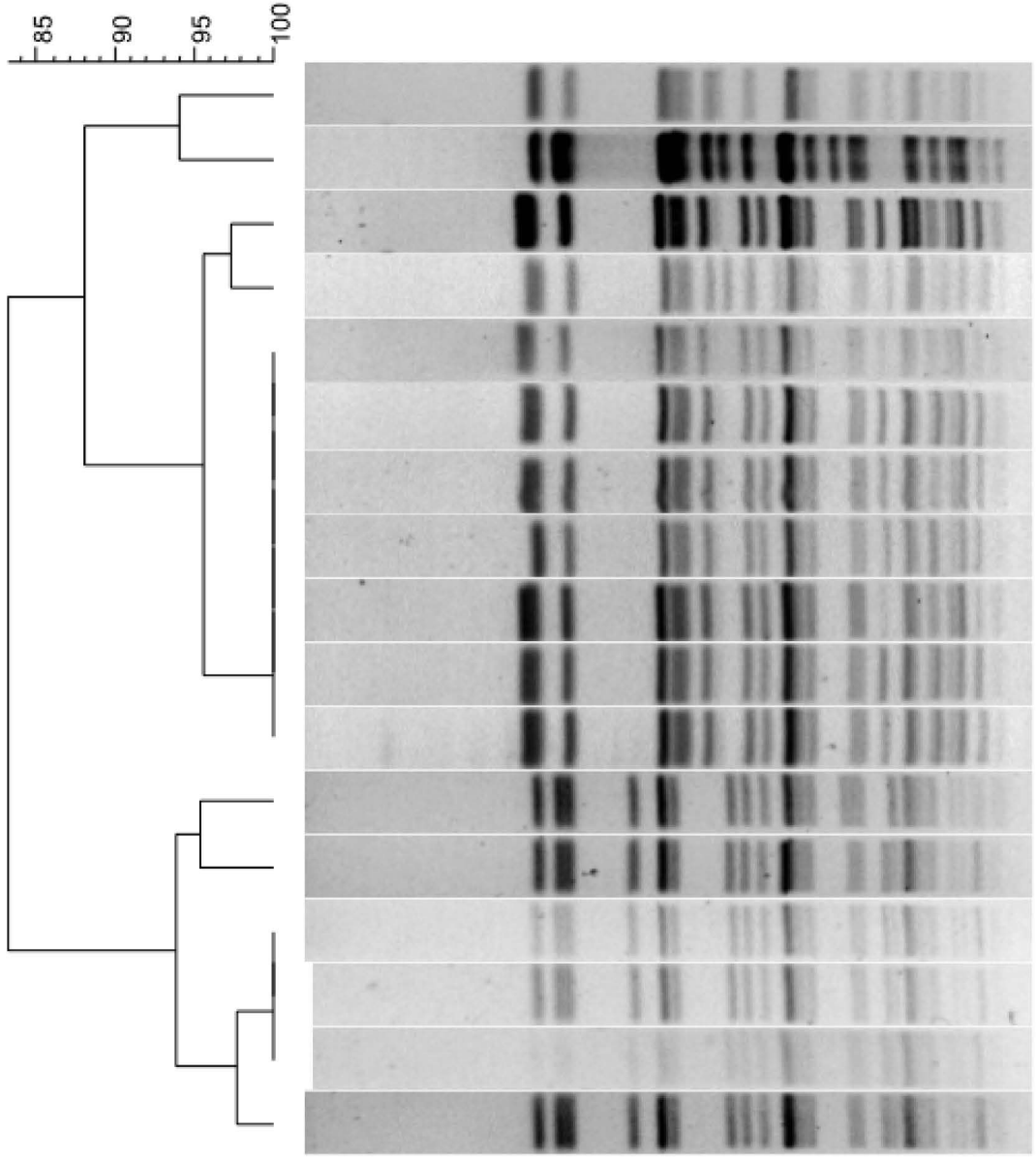
Vietnam. PLoS Negl Trop Dis 4: e702.

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).



Sample ID	Isolation date	XbaI pattern
SS1193/11	2011-06-13	ARJ16X01.0271
SS843/13	2013-04-03	ARJ16X01.0410
SS393/10	2010-02-03	ARJ16X01.0083
SS409/10	2010-02-08	ARJ16X01.0084
SS392/10	2010-01-27	ARJ16X01.0086
SS397/10	2010-01-26	ARJ16X01.0086
SS400/10	2010-01-27	ARJ16X01.0086
SS401/10	2010-01-27	ARJ16X01.0086
SS404/10	2010-02-02	ARJ16X01.0086
SS405/10	2010-01-27	ARJ16X01.0086
SS408/10	2010-02-02	ARJ16X01.0086
SS995/11	2011-03-30	ARJ16X01.0334
SS996/11	2011-03-30	ARJ16X01.0245
SS480/11	2011-02-03	ARJ16X01.0318
SS481/11	2011-02-04	ARJ16X01.0318
SS482/11	2011-02-23	ARJ16X01.0318
SS994/11	2011-03-22	ARJ16X01.0338


SS405-10
SS409-10
SS401-10
SS408-10
SS392-10
SS404-10
SS393-10
SS397-10
SS400-10
SS1193-11
Ss046
SS843-13
SS995-11
SS996-11
SS480-11
SS994-11
SS481-11
SS482-11