Comparative genome analysis of 12 *Shigella sonnei* strains: virulence, resistance, and their interactions

Running Title

Virulence and resistance in Shigella sonnei

Zuobin Zhu[1#], Liang Wang[2,3#], Feng Gu[4#], Ying Li[5], Heng Zhang[6], Ying Chen[5], Jiajia Shi[7], Ping Ma[5, 8], Bing Gu[5, 8*]

[1]Department of Genetics, Research Facility Center for Morphology, Xuzhou Medical University, Xuzhou 221004, China

[2]Department of Bioinformatics, School of Medical Informatics, Xuzhou Medical University, Xuzhou 221004, China

[3]Key Laboratory of New Drug Research and Clinical Pharmacy of Jiangsu Province, Xuzhou Medical University, Xuzhou 221000, China

[4]Clinical laboratory, Xuzhou tumor hospital, Xuzhou 221004, China

[5]Medical Technology School of Xuzhou Medical University, Xuzhou 221004, China

[6]Jining Psychiatric Hospital, Jining 272051, China

[7]School of Transportation, Nantong University, Nantong 226019, China

[8]Department of Laboratory Medicine, Affiliated Hospital of Xuzhou Medical University, Xuzhou 221006, China

[#]These authors contribute equally to the study.

[*]For all correspondence, please contact Prof. Bing Gu. Department of Laboratory Medicine, Affiliated Hospital of Xuzhou Medical University, Xuzhou 221006, China. E-mail: gb20031129@163.com.

Abstract

Shigellosis is a highly infectious disease that are mainly transmitted via faecal-oral contact of the bacteria *Shigella*. Four species have been identified in *Shigella* genus, among which *S. flexneri* is used to be the most prevalent species globally and commonly isolated from developing countries. However, it is being replaced by *S. sonnei* that is currently the main causative agent for dysentery pandemic in many emerging industrialized countries such as Asia and the Middle East with unclear reasons. For a better understanding of *S. sonnei* virulence and antibiotic resistance, we sequenced 12 clinical *S. sonnei* strains with varied antibiotic-resistance profiles collected from four cities in Jiangsu Province, China. Phylogenomic analysis clustered antibiotic sensitive and resistant *S. sonnei* into two distinct groups while pan-genome analysis reveals the presence and absence of featured genes in each group. Screening of 31 classes of virulence factors found out that type 2 secretion system is doubled in resistant strains. Further principle component analysis based on the interactions between virulence and resistance indicated that abundant virulence factors are associated with higher resistant phenotypes. The result present here is based on statistical analysis of a small sample size and serves basically as a guidance for further experimental and theoretical studies.

Keywords *Shigella sonnei*, Virulence factor, Comparative genomics, Antibiotics resistance, Shigellosis

Introduction

Shigellosis is a life-threatening diarrheal infection (dysentery) and is currently a major global burden. Although four species fall into the genus of *Shigella*, *S. sonnei* and *S. flexneri* are now accounting for the most dysentery worldwide while the other two strains *S. dysenteriae* and *S. boydii* experience inexplicable absence[1]. Although *S. flexneri* is the leading cause of endemic shigellosis in developing countries with around 75% episodes, *S. sonnei* is increasing in many rapidly developing countries due to the improvement of socioeconomic conditions [1,2]. The rapid expansion of *S. sonnei* have been attributed to a couple of speculative reasons, such as passive immunization caused by *Plesiomonas shigelloides* and protected niche provided by ubiquitous amoeba species *Acanthamoeba castellaniii*[2]. In addition, *S. sonnei* is more capable of incorporating functional virulence factors (VFs) and/or adept at acquiring antibiotic resistance (AR) from other bacteria[1,3]. Hence, enhanced survival advantages. However, a recent study showed that *S. sonnei* is not able to use amoebae as a protective host to enhance its environmental survival, which makes the explanation of *A. castellaniii* protection for *S. sonnei* expansion compromised[4]. On the other hand, antibiotic resistance is thought to be associated with a fitness cost[5]. Thus, resistance and virulence are historically thought to be negatively correlated[6]. However, recent studies indicated that antibiotic selection pressure and genetic associations could lead to the co-occurrence of resistance and virulence in multiple pathogenic bacteria[7].

In this study, we sequenced 12 clinically isolated strains with diverse antibiotic resistance phenotypes. Sequenced *S. sonnei* genomes were aligned and compared with the reference strain *S. sonnei* 53G. The relationship between the antibiotic resistance and virulence were studied by combining antibiotic resistance profiles with the distribution of putative virulence factors. Here, what we mean by virulence factors is gene products that enable a microorganism to establish itself on or within a host of a particular species, facilitating its abilities to cause diseases, which are divided into 4 categories and 31 functional groups, such as bacterial toxins, cell surface proteins, and hydrolytic enzymes, *etc*[8]. All the virulence factors come from 32 major bacterial pathogens including the *Shigella* genus. By screening and comparing the genomes of sensitive and antibiotic resistant *Shigella sonnies* using this hierarchical set of VF sequence models, we attempted to quantify virulence by the number of virulence factors in specific functional groups. Principal component analysis was then performed to cluster the resistant and sensitive strains, separately, via incorporating both the number of virulence

106 factors and the degree of antibiotic resistance. Genes uniquely associated with sensitive and
107 resistant strains were also studied. Based on these analyses, we could get a better understanding
108 of how virulence and antibiotic resistance are interacted.
109
110 **Results**
111
112 **Genome annotation and comparison**
113
114 General features of 12 *Shigella sonnei* genomes are presented in **Supplementary Table 1**,
115 which include information about resistance profiles to 9 previously mentioned antibiotics,
116 genome size, contigs, coding sequences, and RNAs (tmRNA, tRNA, rRNA). All strains were
117 sensitive to Norfloxacin and most strains were susceptible to Amoxicillin/Clavulanic acid.
118 Genome size ranges from 4.49 Mbps to 4.76Mbps. The number of CDSs ranges from 4328
119 (S13029) to 4646 (S14049). All strains have a single tmRNA coding gene. The number of
120 ribosome RNA (rRNA) and transfer RNA (tRNA) among strains varies slightly with no
121 significant difference. All the genomes were aligned against reference genome S. sonnei G53
122 in circular form via BRIG in terms of distribution of GC content and sequence similarity in
123 **Supplementary Figure 1**[9]. Absence of large blocks in both sensitive and resistant strains were
124 observed, which requires further investigation for their biological meanings.
125
126 **Pan- and phylo-genomic analysis**
127
128 The total pan-genome for the 12 *S. sonnei* strains include 5608 protein CDSs. Of those, 3893
129 (69.42% of total CDSs) are core genes across all 12 species while 1715 (30.58% of total CDSs)
130 constitute the accessory fractions, which are unique to each genome. Strain S13029 has the
131 lowest number of the unique genes (484 CDSs) and S14031 has the highest number of unique
132 genes (803 CDSs) (**Supplementary Figure 2**). Interestingly, both strains are completely
133 sensitive to or only resistant to one of all tested antibiotics. Further comparison of all the 12 *S.*
134 *sonnei* strains give complete map of gene presence and absence in each genome (**Dataset 1**).
135 By comparing sensitive strains with resistant strains, unique genes associated with the two
136 groups were identified and annotated based on sequence homology (**Supplementary Table 2**),
137 respectively. These genes could serve as a guidance for a better understanding of the
138 differences between the two *S. sonnei* groups in terms of their virulence and resistance.
139 Phylogenomic analysis based on the concatenation of 3893 core genes clearly classified the 12

strains into two distinct groups (**Figure 1**). Antibiotic sensitive strains S13029 and S14031 are distantly related with other 10 antibiotic resistant strains, which suggested distinct difference between the two groups in terms of evolutionary pathway.

Interactions between virulence and resistance

In order to investigate the correlations between resistance and virulence, we divided all virulence factors into 4 categories and 31 groups based on VFDB's instructions as specified above[8], which were then used to identify the presence and absence of virulence factors in all studies genomes. Distribution patterns of virulence factors and their abundance in each strain were presented in **Table 1**. Nine groups of virulence factors are completely missing in all *S. sonnei* while three groups of virulence factors are equally distributed in each genome. Among the rest 19 groups of virulence factors, Type 4 pili and T2SS shows apparent differences between sensitive and resistant groups. If only strain S13029 was considered, the number of most groups of virulence factors is reduced except for the three groups Chaperone usher, Flagella, and T3SS. Principal component analysis was also performed by combining antibiotic resistance profiles and distributions of virulence factors, according to which, sensitive strains S13029 and S14031 are separately clustered when compared with other resistant strains (**Supplementary Figure 3**). In addition, S13029 is most isolated due to its tight correlation between reduced virulence and resistance loss.

Further analysis in terms of unique genes associated with sensitive strains or resistant strains were also analysed, which identified 26 genes specifically associated *with Shigella sonnei* sensitive strains and 39 unique genes with resistant strains (**Supplementary Table 2**). Investigation into these genes could shed light on a better understanding of physiological differences between resistant and sensitive *S. sonnei* strains. Genes with no assigned names are generally hypothetical proteins and will not be considered in this study.

Discussion

In this study, 12 newly isolated *S. sonnei* strains were sequenced and assembled into complete genomes, which were then well annotated and thoroughly analysed. It is generally accepted that antibiotic resistance causes fitness costs such as slow growth rate[10] and virulence attenuation [11]. However, recent studies are challenging this view by providing evidence in

which drug resistance leads to increased pathogenicity[11]. This study performed a bioinformatics analysis by focusing on the interplays between virulence and resistance based on 12 newly sequenced *S. sonnei* strains. Through the distribution of functional groups of virulence factors in *S. sonnei* strains, specific patterns were observed. Nine virulence factor groups were completely absent in all S. sonnei strains while another three groups have equal number of virulence factors in all strains (**Table 1**). Thus, their interactions with antibiotic resistance were not considered. As for the rest 19 groups of virulence factors, type 4 pili and T2SS genes are skewedly present in more resistant *S. sonnei* strains when compared with the two sensitive strains S13029 and S14031. Type 4 pili is a bacterial extracellular appendage essential for attachment to host cells while T2SS is responsible for the secretion of numerous degradative enzymes and toxins for bacterial survival[12]. The reasons for their association with high antibiotic resistance are worthy of further exploration. Considering the versatility of virulence mechanisms, PCA was performed to study the interactions between virulence and resistance. It was clearly shown that completely sensitive strain S13029 and single-resistant strain S14031 are distantly separated from other resistant strains that cluster together (**Supplementary Figure 3**). The statistical analysis provided a theoretical support for the view that high virulence is associated with high resistance[11]. However, what the mechanisms are behind this putative relationship is still unclear and requires more efforts to solve the puzzle.

On the other hand, pangenome analysis also revealed vast genome heterogenicity within the same strains of *S. sonnei* according to the identified core and cloud genes (**Supplementary Figure 1**). These cloud genes, more precisely known as strain specific genes, could reveal bacterial characteristics and dynamics, leading to a better understanding of physiological, pathological, and epidemiological features of *S. sonnei*[13]. In specificity, all unique genes that are respectively associated with sensitive and resistant strains were listed in **Supplementary Table 2**. Apparent differences could be observed from the comparisons of these gene functions. However, more theoretical and experimental studies should be performed on a larger scale of datasets in order to get a better understanding of how virulence and resistance are interacted.

Conclusion

In this study, *S. sonnei* sensitive strains was clearly separated from resistant strains based on phylogenomic analysis. By combining antibiotic resistance profiles with the distribution of putative virulence factors, we explored the relationships between virulence and resistance.

Reduced number of putative virulence factors, especially for the group of Type 4 pili and T2SS, was observed to be related with antibiotic sensitivity. Principal component analysis also clustered the resistant and sensitive strains, separately, via incorporating both the abundance of virulence factors and the degree of antibiotic resistance. Finally, genes uniquely associated with sensitive and resistant strains were also reported. In order to better understand how resistance and virulence are interacted, further theoretical and experimental studies should be performed.

Methods and materials

Bacterial isolates and DNA extraction

12 *Shigella sonnei* strains were isolated from different patients with either diarrhea or dysentery in four cities in Jiangsu province, China. Resistance profile to 9 antibiotics (AMC, CFT, CTX, GEN, NAL, NOR, TBT and SMZ) as previously described for each strain was provided by Jiangsu Provincial CDC based on routine screening. Isolates of *Shigella flexneri* were plated on trypticase soya agar (TSA). Picked-up single colony was then inoculated in 5ml trypticase soya broth (TSB) and incubated overnight at 37 ℃ with shaking rate of 200 rpm. DNA isolation was performed using Easy-DNA™ Kit for genomic DNA isolation (Invitrogen Life Technologies, Carlsbad, CA, USA)

Genome sequencing, assembly, and annotation

Sequencing and assembly. Genomes of the 12 Shigella sonnei strains were performed using Illumina Hiseq4000 by generating multiplexed paired-end libraries with an average insert size of 300 bp. In order to obtain more accurate and reliable results in subsequent bioinformatics analysis, the raw data will be treated: (1) Read1 selects 1 bp- 150 bp, read2 selects 1 bp- 150 bp; (2) Remove reads with a certain proportion of low quality (20) bases (40% as default, parameter setting at 60 bp); (3) Remove reads with a certain proportion of Ns (10% as default, parameter setting at 15 bp); (4) Remove adapter contamination (15 bp overlap between adapter and reads as default, parameter setting at 15 bp); (5) Remove duplication contamination. We assemble the short reads into genome sequence using SOAPdenovo[14]. De novo assembly of human genomes with massively parallel short read sequencing[15]. Key parameter K setting at 89 is determined by optimal assembly result for all the 12 strains. Then the assembly result is

local assembled and optimized according to paired-end and overlap relationship via mapping reads to Contig. The detailed description of assembly results were provided in Supplementary Table 4.

Obtained sequences were assessed via FastQC, assembled via SPAdes[16], and reordered via MAUVE[17] based on reference genome *S. sonnei* 53G by following Edwards and Holt's beginner's guide to comparative bacterial genome analysis using next-generation sequence data (Version 2)[18]. For the annotation process, assembled DNA sequences of the drafted genomes from the 12 isolates were run through an automatic annotation pipeline via Prokka (rapid prokaryotic genome annotation)[19].

Pan- and phylo-genomic analysis

Core-/pan-genome analysis was performed by using standalone software Roary[18]. Core genes (99%≤strains≤100%), soft core genes (95%≤strains<99%), shell genes (15%≤strains<95%) and cloud genes (0%≤strains<15%) were calculated. Core and unique genes in the genomes were illustrated in Venn diagram. Presence and absence of all genes in each genome were summarized in **Supplementary Table 1**. *S. sonnei* genomes were visualized in circular form by comparing to the reference genome *S. sonnei* 53G via standalone software BRIG[9]. A Newick tree for 12 *Shigella sonnei* strains, was generated based on 3893 core genes in each genome by the phylogenomic analysis package FastTree[21]. The tree was then visualized through online webserver interactive Tree of Life (iTOL)[22].

Interactions between virulence and resistance

31 functional groups of bacterial virulence factors belonging to four categories were downloaded from the Virulence Factor Database (VFDB)[8] and used to screen translated CDSs of the 12 *Shigella sonnei* strains via phmmer command (full-length alignment with e-value less than 1e-5) in HMMER package[23]. For each group in each proteome, multiple homologous sequences of virulence factors were found, which were then processed to get rid of redundant sequences. MDR (resistance to more than 1 antibiotics) and sensitive *S. sonnei* strains (resistance to 0 or 1 antibiotics) were compared in terms of the abundance of specific groups of virulence factors via Python scripts (available under request). Principal component analysis

was performed by incorporating both antibiotic resistance profiles and distribution of virulence factors.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (81871734, 81471994), Jiangsu Provincial Natural Science Foundation (BK20151154, BK20180997), Jiangsu Provincial Medical Talent (ZDRCA2016053), Six Talent Peaks Project of Jiangsu Province (WSN-135), Advanced Health Talent of Six-one Project of Jiangsu Province (LGY2016042), and Jiangsu Provincial Commission of Health and Family Planning Research Project (H201631), Startup Foundation for Excellent Researchers at Xuzhou Medical University (No. D2016007), The Natural Science Foundation for the Jiangsu Higher Education Institutions of China (No. 16KJB180028), Innovative and Entrepreneurial Talent Scheme of Jiangsu Province (2017), The Natural Science Foundation of the Jiangsu Higher Education Institutions of China (No. 17KJB360014).

References

1 Kotloff, K. L., Riddle, M. S., Platts-Mills, J. A., Pavlinac, P. & Zaidi, A. K. M. Shigellosis. *The Lancet* **391**, 801-812, doi:10.1016/s0140-6736(17)33296-8 (2018).

2 Clements, A. C. A., Thompson, C. N., Duy, P. T. & Baker, S. The Rising Dominance of Shigella sonnei: An Intercontinental Shift in the Etiology of Bacillary Dysentery. *PLOS Neglected Tropical Diseases* **9**, doi:10.1371/journal.pntd.0003708 (2015).

3 Kozyreva, V. K. *et al.* Recent Outbreaks of Shigellosis in California Caused by Two Distinct Populations of Shigella sonnei with either Increased Virulence or Fluoroquinolone Resistance. *mSphere* **1**, doi:10.1128/mSphere.00344-16 (2016).

4 Watson, J., Jenkins, C., Clements, A. & Schottel, J. L. Shigella sonnei Does Not Use Amoebae as Protective Hosts. *Applied and Environmental Microbiology* **84**, doi:10.1128/aem.02679-17 (2018).

5 Guillard, T., Pons, S., Roux, D., Pier, G. B. & Skurnik, D. Antibiotic resistance and virulence: Understanding the link and its consequences for prophylaxis and therapy. *BioEssays* **38**, 682-693, doi:10.1002/bies.201500180 (2016).

6 Cameron, D. R. *et al.* Impact of daptomycin resistance onStaphylococcus aureusvirulence. *Virulence* **6**, 127-131, doi:10.1080/21505594.2015.1011532 (2015).

7 Zhang, L. *et al.* Effects of Selection Pressure and Genetic Association on the Relationship between Antibiotic Resistance and Virulence in Escherichia coli. *Antimicrobial Agents and Chemotherapy* **59**, 6733-6740, doi:10.1128/aac.01094-15 (2015).

8 Chen, L. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Research* **33**, D325-D328, doi:10.1093/nar/gki008 (2004).

9 Alikhan, N.-F., Petty, N. K., Ben Zakour, N. L. & Beatson, S. A. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* **12**, doi:10.1186/1471-2164-12-402 (2011).

10 Andersson, D. I. & Hughes, D. Antibiotic resistance and its cost: is it possible to reverse resistance? *Nature Reviews Microbiology* **8**, 260-271, doi:10.1038/nrmicro2319 (2010).

11 Geisinger, E. & Isberg, R. R. Interplay Between Antibiotic Resistance and Virulence During Disease Promoted by Multidrug-Resistant Bacteria. *The Journal of Infectious Diseases* **215**, S9-S17, doi:10.1093/infdis/jiw402 (2017).

12 Pizarro-Cerdá, J. & Cossart, P. Bacterial Adhesion and Entry into Host Cells. *Cell* **124**, 715-727, doi:10.1016/j.cell.2006.02.012 (2006).

13 Yang, F. Genome dynamics and diversity of Shigella species, the etiologic agents of bacillary dysentery. *Nucleic Acids Research* **33**, 6445-6458, doi:10.1093/nar/gki954 (2005).

14 Li, R. , Zhu, H. , Ruan, J. , Qian, W. , Fang, X. , & Shi, Z. , et al. (2010). De novo assembly of human genomes with massively parallel short read sequencing. *Genome Research*, **20**(2), 265-272, doi:10.1101/gr.097261.109.

15 Li, R. , Li, Y. , Kristiansen, K. , & Wang, J. . (2008). Soap: short oligonucleotide alignment program. *Bioinformatics*, **24**, 713-714, doi:10.1093/bioinformatics/btn025.

16 Bankevich, A. *et al.* SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology* **19**, 455-477, doi:10.1089/cmb.2012.0021 (2012).

17 Darling, A. C. E. Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. *Genome Research* **14**, 1394-1403, doi:10.1101/gr.2289704 (2004).

18 Edwards, D. J. & Holt, K. E. Beginner's guide to comparative bacterial genome analysis using next-generation sequence data. *Microbial Informatics and Experimentation* **3**, doi:10.1186/2042-5783-3-2 (2013).

19 Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068-2069, doi:10.1093/bioinformatics/btu153 (2014).

20 Page, A. J. *et al.* Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691-3693, doi:10.1093/bioinformatics/btv421 (2015).

21 Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. *Molecular Biology and Evolution* **26**, 1641-1650, doi:10.1093/molbev/msp077 (2009).

22 Letunic, I. & Bork, P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research* **44**, W242-W245, doi:10.1093/nar/gkw290 (2016).

23 Johnson, L. S., Eddy, S. R. & Portugaly, E. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics* **11**, doi:10.1186/1471-2105-11-431 (2010).

Author Contribution statement:

B.G. and Z.Z. designed the study; Z.Z. F.G. and L.W. carried out the experimental work; L.W. Y.L and J.S. analyzed the data; H.Z. and Y.C. prepared the figures and the supplementary tables; P.M. and B.G. do the language modification. Z.Z. and L.W. wrote the manuscript. All authors read and approved the final manuscript.

Table and Figures：

Table 1 Distribution patterns of 4 categories of virulence factors that belong to 31 groups among 12 *Shigella sonnei* strains in terms of antibiotic resistance. Sensitive strains have no resistance or only resist to one antibiotics. The four categories of VFs are Adhesion & Invasion, Secretion system & Effectors, Toxin, Iron acquisition. MDR strains have more than one resistance. PFT: Pore-forming toxins. [#]Nine groups of virulence factors are not present in all *S. sonnei* strains. [*]Three groups of virulence factors have equal number of virulence factors in all strains.

	Shigella Sonnei	**S13029**	**S14031**	**S14049**	**S13115**	**S15036**	**S15123**	**S13120**	**S14014**	**S14089**	**S15047**	**S15109**	**S13098**
	MDR	0	1	5	6	6	6	7	7	7	7	7	8
Adhesion & Invasion	**Chaperone usher**	158	151	152	146	146	146	146	145	148	145	142	147
	Extracellular nucleation precipitation	14	14	14	14	14	14	14	14	14	14	13	14
	Type 4 pili	122	127	142	135	135	145	135	135	139	135	133	137
	Sortase assembled pili*	1	1	1	1	1	1	1	1	1	1	1	1
	Flagella	204	203	210	201	200	200	201	200	206	200	200	201
	Autotransporter	14	14	13	12	12	12	12	12	13	12	12	12
	Fibronectin-binding protein	44	47	46	43	44	44	43	44	44	43	44	44
	Fibrinogen-binding protein	4	4	3	3	3	3	3	3	3	3	3	3
	Collagen-binding protein*	5	5	5	5	5	5	5	5	5	5	5	5
	Other adherence invasion related VFs	65	70	70	69	69	69	70	69	69	69	69	70
Secretion	**T2SS**	6	7	14	13	14	18	13	14	15	14	13	13
	T3SS	172	164	221	163	159	163	162	160	214	163	159	164

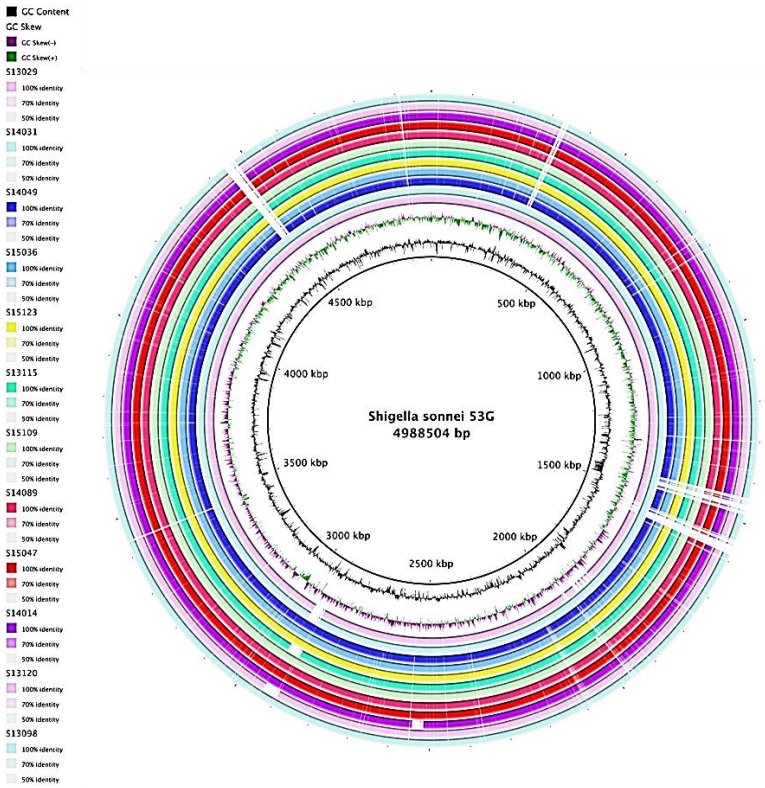
system & Effectors	**T4SS**	38	43	45	44	47	56	45	46	46	45	43	47
	T5SS	15	15	14	13	13	13	13	13	14	13	13	13
	T6SS	172	180	180	181	181	180	182	181	180	180	178	178
	T7SS*	1	1	1	1	1	1	1	1	1	1	1	1
Toxin	**α-PFTs**	1	1	2	1	1	1	2	1	1	1	1	1
	β-PFTs#	0	0	0	0	0	0	0	0	0	0	0	0
	Superantigens and superantigen like protein#	0	0	0	0	0	0	0	0	0	0	0	0
	Surface acting enzymes#	0	0	0	0	0	0	0	0	0	0	0	0
	ADP Ribosyltransferase	2	4	3	3	1	3	4	3	3	3	3	3
	Glucosyltransferase#	0	0	0	0	0	0	0	0	0	0	0	0
	Guanylate adenylate cyclase#	0	0	0	0	0	0	0	0	0	0	0	0
	Deamidase#	0	0	0	0	0	0	0	0	0	0	0	0
	rRNA N-glycosidase#	0	0	0	0	0	0	0	0	0	0	0	0
	Metalloprotease#	0	0	0	0	0	0	0	0	0	0	0	0
	Dnase1 genotoxin	28	31	30	30	30	30	30	29	30	30	30	30
	Intracellular PFTs#	0	0	0	0	0	0	0	0	0	0	0	0
Iron acquisition	**Siderophore mediated iron uptake**	279	285	289	286	287	286	287	283	288	287	285	287
	Heme-mediated iron uptake	116	120	117	117	117	117	117	117	117	117	116	119
	Transferrin and lactoferrin mediated iron uptake	5	5	5	4	5	5	5	5	5	5	5	5


Heilongjiang
Jilin
Liaoning
Beijing
Tianjin
Inner mongolia
Xinjiang
Hebei
Ningxia
Shanxi
Shandong
Qinghai
Gansu
Shaanxi
Henan
Jiangsu
Anhui
Tibet
Sichuan
Chongqing
Hubei
Shanghai
zhejiang
Jiangxi
Hunan
Fujian
Guizhou
Taiwan
Yunnan
Guangxi
Guangdong
Hongkong
Macao
Hainan
Xuzhou
Suqian
Huai'an
Yancheng
Yangzhou
Nantong
Suzhou
S13029
S14031
S14049
S14089
S15024
S15036
S15123
S13120
S15109
S13115
S13098
S15047
Tree scale: 0.1

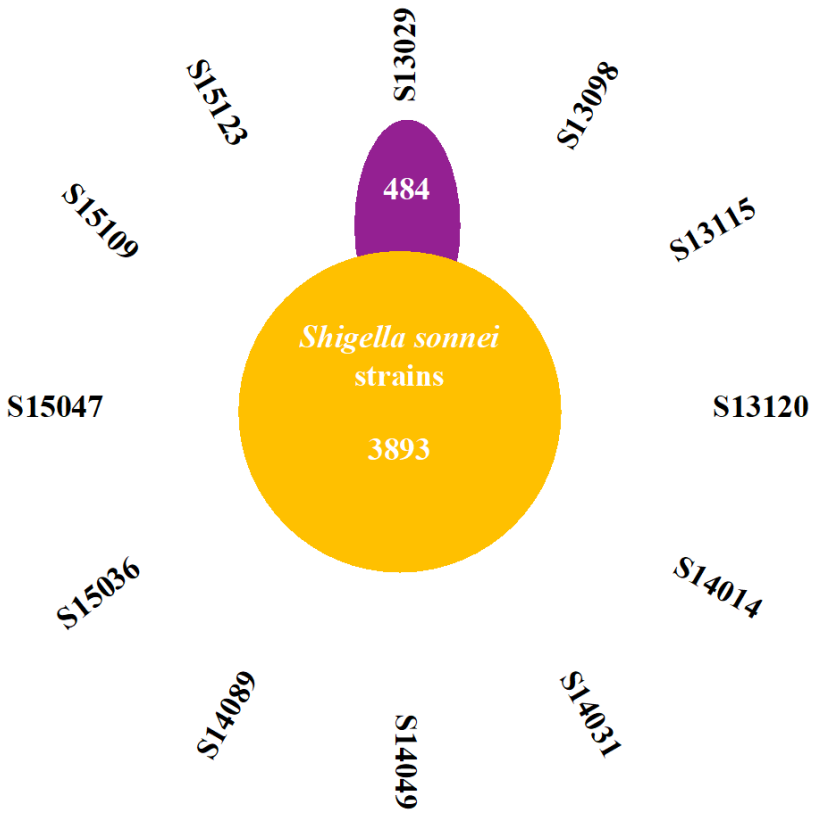
Figure 1 Phylogenomic analysis of 12 clinically isolated *Shigella sonnei* strains with different antibiotic resistant profiles based on 3893 core genes via FastTree. S13029 is completely sensitive to all 9 tested antibiotics, that is, Amoxicillin/Clavulanic acid (AMC), Ceftiophene (CFT), Cefotaxime (CTX), Gentamicin (GEN), Nalidixic acid (NAL), Norfloxacin (NOR), Tetracycline (TBT), and compound Sulfamethoxazole (SMZ) while S14031 is only resistant to SMZ. The other ten strains are resistant to at least five out of the nine antibiotics. All the strains were isolated from 4 municipal cities (Nanjing, Suzhou, Wuxi, Zhenjiang) in Jiangsu Province, China, the geographic distribution of which was depicted in the figure with red numbers. Apparent clusters could be observed for sensitive and resistant strains, respectively. Branch length represents evolutionary distance among strains. Filled red dots indicate bootstrapping value greater than 90%.

Supplementary information：

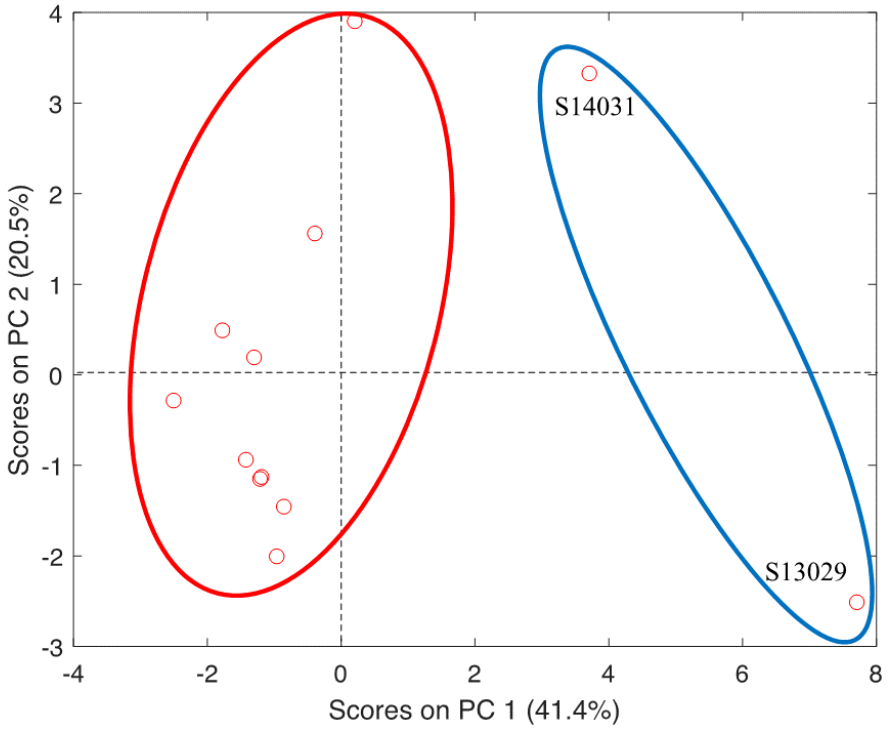
Competing Interests statement: The authors declare no competing interests.



Supplementary Figure 1 Genome comparison of 12 isolated *S. sonnei* strains against reference genome *S. sonnei* 53G generated by BRIG 0.95. The inner cycle (black) represents the complete genome of the reference strain and the shade of each colors denote the similarities between each strain with reference strain. GC content and GC skew (+/-) were illustrated in-between.



Supplementary Figure 2 Pangenome analysis of 12 clinically isolated *Shigella sonnei* strains with different antibiotic resistant profiles. A total of 3893 core genes (99% <= strains <= 100%) were shared by all strains while there are 5608 total genes(0% <= strains <= 100%) in all the strains. S13029 and S14031 are sensitive strains while other ten strains have multiple antibiotic resistance to more than 5 antibiotics.



Supplementary Figure 3 Principal component analysis (PCA) of the relationship between antibiotic resistance profiles and virulence factors in 12 sequenced Shigella sonnei strains. S13029 and S14031 are antibiotics-sensitive strains while the rest strains are multi-drug resistant. (MDR) The two sensitive strains are apparently isolated from other MDR strains.

Supplementary Table 1 Comparison of 12 *S. sonnei* strains based on antibiotic resistance profiles and key genome annotation parameters. The 12 strains were isolated from 4 municipal cities, Nanjing, Suzhou, Wuxi, and Zhenjiang, the economy of which are above the average level of China and are considered the best among Jiangsu province. Antibiotics tested in this study are Amoxicillin/Clavulanic acid (AMC), Ceftiophene (CFT), Cefotaxime (CTX), Gentamicin (GEN), Nalidixic acid (NAL), Norfloxacin (NOR), Tetracycline (TBT), and compound Sulfamethoxazole (SMZ).

Strain ID	Region	AMC	CFT	CTX	GEN	NAL	NOR	TBT	SMZ	MDR	Contig	Bps	CDS	tmRNA	tRNA	rRNA
S13029	Zhenjiang	S	S	S	S	S	S	S	S	0	449	4490117	4328	1	83	5
S13098	Zhenjiang	R	R	R	R	R	S	R	R	8	473	4642881	4505	1	77	4
S13115	Suzhou	S	R	R	R	R	S	S	R	6	409	4530601	4403	1	79	5
S13120	Wuxi	I	R	R	R	R	S	R	R	7	406	4534710	4401	1	80	5
S14014	Zhenjiang	R	R	R	S	R	S	R	R	7	409	4526456	4407	1	80	5
S14031	Nanjing	S	S	S	S	S	S	S	R	1	427	4568178	4414	1	76	5
S14049	Wuxi	S	I	S	R	R	S	R	R	5	477	4756429	4646	1	77	5
S14089	Suzhou	S	R	R	R	R	S	R	R	7	472	4726659	4599	1	80	5
S15036	Nanjing	S	R	R	S	R	S	R	R	6	429	4553996	4425	1	78	5
S15047	Zhenjiang	S	R	R	R	R	S	R	R	7	420	4606954	4484	1	78	5
S15109	Suzhou	S	R	R	R	R	S	R	R	7	587	4541154	4378	1	80	5
S15123	Wuxi	S	R	R	S	R	S	R	R	6	425	4645117	4528	1	80	5

*MDR: the total number of multi-drug resistance; Bps: base pairs; CDS: coding sequences; S: sensitive (marked red); R: resistant; I: intermittent.

Supplementary Table 2: Comparing sensitive strains with resistant strains, unique genes associated with the two groups were identified and annotated based on sequence homology. (in a separate document)

Supplementary Table 3 Unique genes specifically associated with *Shigella sonnei* sensitive strains (26 genes) and resistant strains (39 genes). Genes with no assigned specific names are not included in the list. Functions are assigned based on sequence homology.

Sensitive Strains	Function
bdm	Biofilm-dependent modulation protein
betA	Oxygen-dependent choline dehydrogenase
betB_1	NAD/NADP-dependent betaine aldehyde dehydrogenase
betI	HTH-type transcriptional regulator
betT	Choline transport protein
elaD_1	Deubiquitinase
fimD_1	Outer membrane usher FimD-like protein
fimD_2	Outer membrane usher FimD-like protein
fimD_3	Outer membrane usher FimD-like protein
fimD_4	Outer membrane usher FimD-like protein
fimF	Fimbrial protein
fimG	Fimbrial morphology protein
fimH_1	Adhesin
flhE	Flagellar biosynthesis protein
focC_2	Chaperone protein
gpFI_1	Putative prophage major tail sheath protein
intS_4	Integrase
ompX_2	Outer membrane protein X
pdeL	Cyclic di-GMP phosphodiesterase
sat	Serine protease sat autotransporter
uacT_1	Uric acid transporter
yahB_2	Putative HTH-type transcriptional regulator
ybdO_1	Putative HTH-type transcriptional regulator
ydiO	Putative fimbrial chaperone

ynfE	Putative dimethyl sulfoxide reductase chain
yraI	Putative fimbrial chaperone

Resistant Strains	**Function**
ail	Attachment invasion locus protein
ant1	Streptomycin 3'-adenylyltransferase
citC_2	[Citrate [pro-3S]-lyase] ligase
clpP_2	ATP-dependent Clp protease proteolytic subunit
dhfrI	Dihydrofolate reductase type 1
dmlR_2	HTH-type transcriptional regulator
elfC_4	Putative outer membrane usher protein
erfK_1	Putative L,D-transpeptidase
eutB_1	Ethanolamine ammonia-lyase heavy chain
fhuE_1	FhuE receptor
fucI_2	L-fucose isomerase
gabP	GABA permease
gatA_1	PTS system galactitol-specific EIIA component
gspA_1	Putative general secretion pathway protein A
gspB	Putative general secretion pathway protein B
gspE	Putative type II secretion system protein E
hcaR_2	Hca operon transcriptional activator
hyuA_1	D-phenylhydantoinase
mhpC_2	2-hydroxy-6-oxononadienedioate/2-hydroxy-6-oxononatrienedioate hydrolase
mngB_2	Mannosylglycerate hydrolase
murP_2	PTS system N-acetylmuramic acid-specific EIIBC component
puuC_2	NADP/NAD-dependent aldehyde dehydrogenase
rhaR_2	HTH-type transcriptional activator
tnsA	Transposon Tn7 transposition protein
tnsB	Transposon Tn7 transposition protein

tnsC	Transposon Tn7 transposition protein
tnsE	Transposon Tn7 transposition protein
torZ_2	Trimethylamine-N-oxide reductase 2
umuC_1	Protein UmuC
umuD_1	Protein UmuD
wecD_2	dTDP-fucosamine acetyltransferase
xerC_2	Tyrosine recombinase
xerD_2	Tyrosine recombinase
yadC_3	Putative fimbrial-like protein
ycaM_2	Inner membrane transporter
ycjP_2	Inner membrane ABC transporter permease protein
ygeA	Putative racemase
ygfK_2	Putative oxidoreductase
ypdA_2	Sensor histidine kinase

Supplementary Table 4: The detailed description of assembly results. (in a separate document)