


GENOME REPORT

Open Access



Genome sequence of *Shigella sonnei* 4303

Laura Deutsch-Nagy^{1,2*} , Péter Urbán^{2,3}, Zsuzsanna Tóth^{2,3}, Zoltán Bihari⁴, Béla Kocsis⁵, Csaba Fekete^{2,3} and Ferenc Kilar^{1,2*}

Abstract

Background: *Shigella* spp. are Gram-negative intracellular pathogenic bacteria belonging to the family *Enterobacteriaceae* and can cause bacterial dysentery, a severe diarrheal disease. The pathophysiological impact of the Gram-negative bacteria is highly related to the composition and structural variability of lipopolysaccharides, the major lipid components of the outer membrane. Out of the 114 genes involved in the lipopolysaccharide biosynthesis pathway, 47 genes are specific to *Shigella* spp. Changes in the specific genes can lead to loss of the O polysaccharide side chain, resulting in rough (*R*) type bacteria with increased sensitivity to temperature, or hydrophobic antibiotics. The formation of various different lipopolysaccharides or lipooligosaccharides has been observed previously in a mutant line showing altered biological properties, but the genetic background has not been investigated in detail.

Results: The parental strain of the mutant line, *Shigella sonnei* 4303, was subjected to whole genome sequencing to gain a better insight into the structure and biosynthesis of lipopolysaccharides. The sequencing revealed a 4,546,505 bp long genome including chromosomal and plasmid DNA, and the lipopolysaccharide biosynthesis genes were also identified. A comparison of the genome was performed with the phylogenetically closely related, wild type, well characterized, highly virulent strain, *S. sonnei* 53G.

Conclusion: Analysis of the lipopolysaccharide biosynthetic genes helped us to get more insight into the pathogenicity and virulence of the bacteria. The genome revealed high similarities with *S. sonnei* 53G, which can be used as a standard in characterizing the *S. sonnei* 4303's *R*-type isogenic derivatives.

Keywords: Shigellosis, *Shigella sonnei* 4303, Genome, Lipopolysaccharide biosynthesis

Background

Lipopolysaccharides (LPSs) are of importance in bacterial physiology, and also in host-bacteria crosstalk [1]. The pathogenicity of Gram-negative bacteria is influenced by the molecular variability (structures and lengths) of LPSs, e.g., serum sensitivity and biofilm forming ability of Gram-negative bacteria are correlated with the lengths of O sidechains. Previous studies have described that *R*-type bacteria with truncated LPS molecules (so-called lipooligosaccharides—LOSs) are more sensitive to hydrophobic antibiotics [2].

Recent studies suggested that *Shigella sonnei* have become more dominant in developed countries [3]. The history of *S. sonnei* 4303 dates back more than 60 years,

when the phenomenon of phase variation in *S. sonnei* was examined [4]. This non-pathogenic strain was formed by plasmid loss from a pathogenic *S. sonnei* phase I strain, due to the instable nature of the virulence plasmid [5]. Later, intensive studies were carried out on the strain and its *R*-type isogenic derivatives, and the chemical structures and structural variabilities of their lipopolysaccharides and lipooligosaccharides (LOSs) have been described. Several interesting *R* mutants were characterized, including an absolute *R*-type strain (*S. sonnei* 4350) and a strain having truncated LPSs with a D-glycero-D-mannoheptose component incorporated in the structure (*S. sonnei* 4351) [6–12]. The lack of appropriate genome-scale information of the investigated strains, including structurally different LPSs, however, hinders our ability to answer fundamental biosynthetic questions. In order to gain more insight into the mechanism of the LPS/LOS biosynthesis *S. sonnei* 4303 was subjected to whole genome sequencing.

*Correspondence: laura.nagy@aok.pte.hu; ferenc.kilar@aok.pte.hu

¹ Institute of Bioanalysis, Medical School, University of Pécs, Szigeti út 12, 7624 Pécs, Hungary

Full list of author information is available at the end of the article



Methods

The genomic library was made by enzymatic shearing with the Ion Xpress Plus fragment library kit, followed by size selection on a 2% agarose E-Gel SizeSelect Gel (Thermo Fisher Scientific Inc., Waltham, MA USA). The template was prepared with 100 pM of the library on an Ion One Touch 2 system (Thermo Fisher Scientific Inc., Waltham, MA, USA). Samples were loaded on an Ion 316v2 Chip and sequenced on an Ion Torrent PGM sequencer, with the Ion PGM Sequencing 200 Kit v2 (Thermo Fisher Scientific Inc., Waltham, MA, USA) in compliance with the manufacturer's recommendations. De novo assembly was performed using the SPAdes v3.1 Genome Assembler software [13]. For whole-genome alignment, scaffolds in the draft assemblies were reordered to the *S. sonnei* 53G as reference sequence in Mauve software with default parameters [14]. Sequence annotation was performed using Prokka v. 1.9 [15]. MeDuSa (Multi-Draft based Scaffold) web server was used for genome scaffolding [16]. The genome sequence of *S. sonnei* 4303 has been deposited in the GenBank under the accession number PRJNA361576.

Phylogenetic analysis was performed with the closest relatives selected by 16S rRNA sequences through NCBI (BLASTn). Phylogenetic analysis was performed by Clustal Omega with default settings [17]. Multiple sequence alignment was completed with *adk*, *fumC*, *gyrB*, *mdh*, *purA* housekeeping genes and rRNA genes. The resulted phylogenetic tree represents 12 *S. sonnei* strains including *S. sonnei* 4303 and an outgrouped strain, *Klebsiella oxytocal* FDAARGOS 355.

Nomenclature of the LPS genes were used according to KEGG database [18].

Detailed methodological strategy is described in Additional file 1.

Quality assurance

Morphological and biochemical characterization identified the strain as *S. sonnei*. The genomic DNA used for sequencing was isolated from a single colony of the bacteria. The 16S rDNA gene was extracted from the genome using RNAmmer 1.2 server [19]. The identity of the strain was confirmed through BLAST annotation against NCBI microbial 16S database.

Results and discussion

In total 4,262,518 high quality reads were generated and used to create the genome of *S. sonnei* 4303, which yielded a 100-fold coverage. The genome is 4.5 Mbps in size, and contains 4554 predicted genes, 10 rRNA genes, 60 tRNA genes as well as a CRISPR region. In our comparative studies, the genome of a well characterized, highly pathogenic and phylogenetically highly related strain, *S. sonnei* 53G was used as standard (Fig. 1). *S. sonnei* 53G was isolated in Japan [20] and was used in different serological studies [21].

Since the primary aim of this study was to create a solid and strain specific information about the genetic background with regards to LPS modifications, the genes involved in the LPS biosynthetic pathway have been further analyzed in silico. According to the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways database, 114 genes participate in these complex biological processes. Screening for the presence/absence of these genes in the *S. sonnei* 4303 and in the *S. sonnei* 53G strains revealed 47 genes specific to *S. sonnei*. Comparative DNA analysis on this common subset of *S. sonnei*

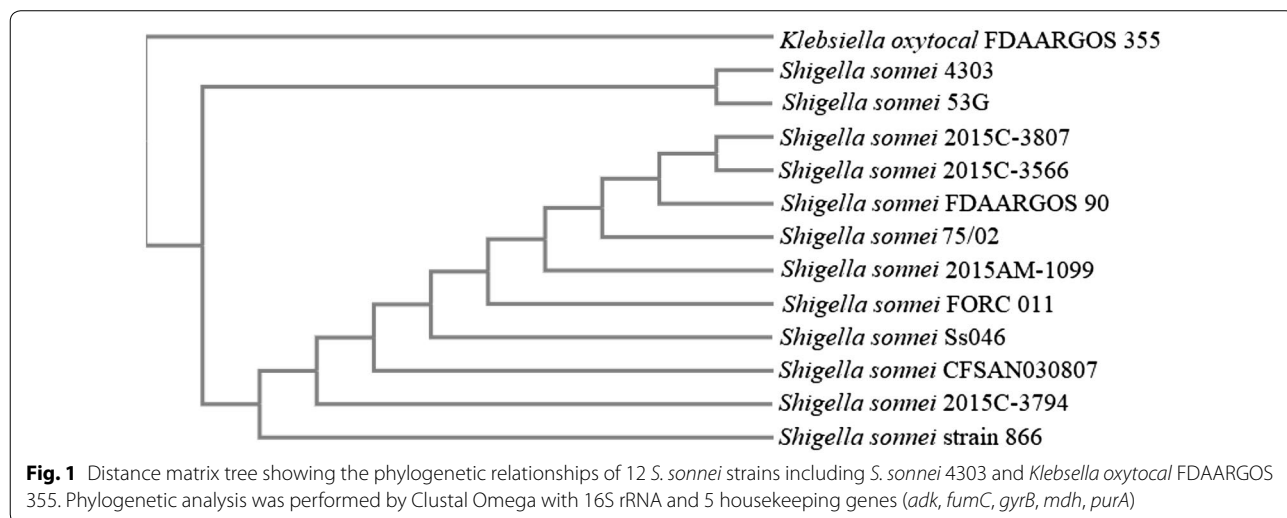


Table 1 Lipopolysaccharide biosynthesis genes according to Kyoto encyclopedia of genes and genomes in *Shigella sonnei* 4303

Gene name/synonym(s)	Similarity to <i>Shigella sonnei</i> 53G (%)
Lipid A	
<i>lpxB</i>	100
<i>lpxA</i>	100
<i>lpxD</i>	100
<i>lpxL/htrB</i>	100
<i>lpxM/msbB</i>	100
<i>pagP</i>	99 ^a
<i>eptA</i>	100
<i>lpxC</i>	100
<i>lpxH</i>	100
<i>arnT</i>	100
<i>lpxK</i>	100
<i>lpxT/yeiU</i>	99 ^b
<i>lpxP/ddg</i>	99 ^a
Core region	
<i>waaA/kdtA</i>	100
<i>rfaC/waaC</i>	100
<i>rfaF/waaF</i>	100
<i>waaQ</i>	100
<i>rfaG/waaG</i>	100
<i>rfaI</i>	100
<i>waaR</i>	100
<i>waaV</i>	99 ^c
<i>waaW</i>	100
<i>rfaP/waaP</i>	100
<i>rfaY/waaY</i>	100
<i>eptB</i>	100
<i>eptC</i>	100
<i>waaH</i>	100
O antigen	
<i>rfaL</i>	100
<i>wecA</i>	100
<i>wzzB</i>	100
Unusual sugar	
<i>kdsD</i>	100
<i>kdsA</i>	99 ^d
<i>kdsC</i>	100
<i>kdsB</i>	100
<i>gmhA</i>	100
<i>gmhC/hldE</i>	100
<i>gmhB</i>	100
<i>gmhD/rfaD/hldD</i>	100
<i>arnA</i>	99 ^d
<i>arnB</i>	100
<i>arnC</i>	100
<i>arnD</i>	100

Table 1 (continued)

Gene name/synonym(s)	Similarity to <i>Shigella sonnei</i> 53G (%)
<i>arnE</i>	100
<i>arnF</i>	100
<i>wecB</i>	100
<i>wbpA/wecC</i>	100

^a Single-nucleotide polymorphism coding nonsense mutation

^b Two gaps and new stop codon

^c The gene has 100% similarity to *Shigella sonnei* Ss046's *waaV* gene

^d Single nucleotide polymorphism coding missense mutation

genes revealed six sequence polymorphisms (summarized in Table 1).

Our former study on the LPS structure of *S. sonnei* 4303 indicated that the lipid A molecules contain only 1 phosphate group at position 1 [6]. Modification of lipid A with an additional phosphate group at position 1, forming a 1-diphosphate species, is mediated by the undecaprenyl phosphotransferase, *LpxT*. The mutation of *lpxT/yeiU* encoding gene may explain the monophosphorylated position 1 in *S. sonnei* 4303.

Taken together, the whole-genome sequencing strategy revealed the mutation of the *lpxT*, and the presence of new variants of the *pagP*, *lpxP*, *kdsA* and *arnA* genes. The sequenced genome can be used as a reference for characterizing *R*-type isogenic derivatives of *S. sonnei* 4303, to reveal the genetic background of mutants with the truncated lipopolysaccharides [6], e.g., having a D-glycero-D-mannoheptose in the core part [7, 8].

Additional file

Additional file 1. Complete methodological strategy to the "Genome sequence of *Shigella sonnei* 4303". Experimental design, Sampling protocol and storage, Nucleic acid isolation, Library preparation and sequencing, Read quality assessment, Comparative genomics.

Abbreviations

LOS: lipooligosaccharide; LPS: lipopolysaccharide.

Authors' contributions

FK and BK contributed to the conception of study. CF contributed to the design of the study and led the project with FK. BK was involved in the creation of lipopolysaccharide rough *R*-type mutant line and description of LPS biosynthesis. LD-N, PU and ZT worked on genome sequencing. ZB assembled and annotated the genome, LD-N and PU analyzed the data. LD-N drafted the manuscript. All authors read and approved the final manuscript.

Author details

¹ Institute of Bioanalysis, Medical School, University of Pécs, Szegedi út 12, 7624 Pécs, Hungary. ² Szentágotthai Research Centre, University of Pécs, Ifjúság

útja 20, 7624 Pécs, Hungary. ³ Department of General and Environmental Microbiology, Faculty of Sciences, University of Pécs, Ifjúság útja 6, 7624 Pécs, Hungary. ⁴ Department of Metagenomics, Institute for Biotechnology, Bay Zoltán Nonprofit Ltd. for Applied Research (BAY-BIO), Széchenyi tér 5, 6720 Szeged, Hungary. ⁵ Department of Medical Microbiology and Immunology, Medical School, University of Pécs, Szigeti út 12, 7624 Pécs, Hungary.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The datasets generated and/or analyzed are available in the GeneBank repository, with Accession Number PRJNA361576, Assembly GCA_002073875.2.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Funding

Sponsorship: This work was supported by the National Research, Development and Innovation Office K-125275. The work was partially supported by the EFOP-3.6.3-VEKOP-16-2017-00009, GINOP-2.3.2-15-2016-00021, and PTE-AOK-KA-2017-19 Grants.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 18 July 2018 Accepted: 10 October 2018

Published online: 24 October 2018

References

- Gnauck A, Lentle RG, Kruger MC. The characteristics and function of bacterial lipopolysaccharides and their endotoxic potential in humans. *Int Rev Immunol*. 2016;35(3):189–218.
- Kim HS, Lee MA, Chun SJ, Park SJ, Lee KH. Role of NtrC in biofilm formation via controlling expression of the gene encoding an ADP-glycero-manno-heptose-6-epimerase in the pathogenic bacterium, *Vibrio vulnificus*. *Mol Microbiol*. 2007;63(2):559–74.
- Thompson CN, Duy PT, Baker S. The rising dominance of *Shigella sonnei*. An intercontinental shift in the etiology of bacillary dysentery. *PLoS Negl Trop Dis*. 2015;9:e0003708.
- Rauss K, Kétyi I, Vertényi A, Vörös S. Studies on the nature of phase variation of *Shigella sonnei*. *Acta microbiol Acad Sci Hung*. 1954;8:53–63.
- Schuch R, Maurelli AT. Virulence plasmid instability in *Shigella flexneri* 2a is induced by virulence gene expression. *Infect Immun*. 1997;65(9):3686–92.
- Kocsis B, Kontrohr T, László V, Milch H. Isolation and characterization of different *Shigella sonnei* rough mutants. *Acta Microbiol Acad Sci Hung*. 1980;27:217.
- Kontrohr T, Kocsis B. Isolation of adenosine 5'-diphosphate-D-glycero-D-mannoheptose. An intermediate in lipopolysaccharide biosynthesis of *Shigella sonnei*. *J Biol Chem*. 1981;256:7715–8.
- Kocsis B, Kontrohr T. Isolation of adenosine 5'-diphosphate-D-glycero-D-mannoheptose, the assumed substrate of heptose transferase(s), from *Salmonella* Minnesota R595 and *Shigella sonnei* Re mutants. *J Biol Chem*. 1984;259:11858–60.
- Bui A, Kilár A, Dörnyei A, Poór V, Kovács K, Kocsis B, Kilár F. Carbohydrate composition of endotoxins from R-type isogenic mutants of *Shigella sonnei* studied by capillary electrophoresis and GC-MS. *Croat Chem Acta*. 2011;84:393–8.
- Kilár A, Dörnyei A, Bui A, Szabo Z, Kocsis B, Kilár F. Structural variability of endotoxins from R-type isogenic mutants of *Shigella sonnei*. *J Mass Spectrom*. 2011;46:61–70.
- Makszin L, Kilár A, Felső P, Péterfi Z, Kocsis B, Kilár F. Quantitative microfluidic analysis of S- and R-type endotoxin components with chip capillary electrophoresis. *Electrophoresis*. 2012;33:3351–60.
- Kilár A, Dörnyei A, Kocsis B. Structural characterization of bacterial lipopolysaccharides with mass spectrometry and on- and off-line separation techniques. *Mass Spectrom Rev*. 2013;32:90–117.
- Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A, Pribelski AD, Pyshkin A, Sirotkin A, Sirotkin Y, et al. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J Comput Biol*. 2013;20(10):714–37.
- Darling AE, Mau B, Perna NT. ProgressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE*. 2010;5(6):e11147.
- Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30(14):2068–9.
- Emanuele B, Beatrice D, Marco G, Sara B, Marie-France S, Pietro L, Pierluigi C, Renato F, Marco F. MeDuSa: a multi-draft based scaffold. *Bioinformatics*. 2015;31(15):2443–51.
- McWilliam H, Li W, Uludag M, Squizzato S, Park YM, Buso N, Peter AC, Lopez R. Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acid Res*. 2013;41(1):597–600.
- Kyoto Encyclopedia of Genes and Genomes Pathway Database. Japan. 2018. https://www.genome.jp/kegg-bin/get_htext?ko01005.keg. Accessed 1 Sept 2018.
- Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res*. 2007;35(9):3100–8.
- Kopecko DJ, Washington O, Formal SB. Genetic and physical evidence for plasmid control of *Shigella sonnei* form I cell surface antigen. *Infect Immun*. 1980;29(1):207–14.
- Formal SB, Baron LS, Kopecko DJ, Washington O, Powell C, Life CA. Construction of a potential bivalent vaccine strain: introduction of *Shigella sonnei* form I antigen genes into the galE *Salmonella typhi* Ty21a typhoid vaccine strain. *Infect Immun*. 1981;34(3):746–50.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

