

Genome Sequence of *Shigella flexneri* 2a: Insights Into Pathogenicity Through Comparison with Genomes of *Escherichia coli* K12 and O157

Jin, Q., Yuan, Z., Xu, J., Wang, Y., Shen, Y., Lu, W., ... Yu, J. (2002). Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity through comparison with genomes of *Escherichia coli* K12 and O157. *Nucleic Acids Research*, 30(20), 4432–4441.

Trixie Roque & Jake Woodlee

**Departments of Biology and Computer Science
Loyola Marymount University**

October 17, 2015

Outline

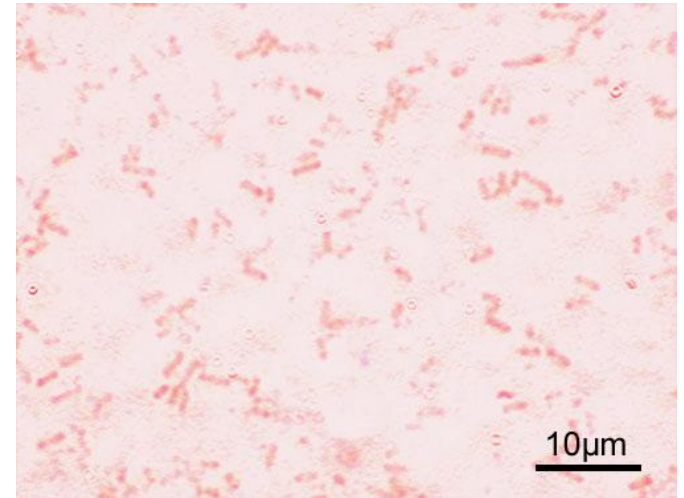
- Shigellosis, caused by *S. flexneri*, is one of the leading causes of death in young children in developing countries.
- Sequencing the genome mainly involved automation to reduce human-induced errors.
- Comparison between *S. flexneri* and its genetic relative, *E. coli.*, revealed distinct and similar characteristics between their chromosomes.
- Viable database for this organism provides a fast and easy way to explore its genome.

Outline

- Shigellosis, caused by *S. flexneri*, is one of the leading causes of death in young children in developing countries.
- Sequencing the genome mainly involved automation to reduce human-induced errors.
- Comparison between *S. flexneri* and its genetic relative, *E. coli.*, revealed distinct and similar characteristics between their chromosomes.
- Viable database for this organism provides a fast and easy way to explore its genome.

***S. flexneri* is a species of Bacteria that Causes a Worldwide Health Concern to Developing Countries**

- Gram-negative, non-sporulating, facultative anaerobes
- Causes bacillary dysentery and shigellosis
- Estimated deaths of 161.1 Million annually
- Invasive in the colon and rectum
- Due to the lack of adequate treatment strategies, the World Health Organization has made an anti-*Shigella* vaccine a priority



The Strain Observed was Sf301 serotype 2a

- *S. flexneri* 2a is the most prevalent species and serotype.
- Determining the connection between the chromosome and virulence plasmid required the discovery of the entire genome sequence of *Shigella flexneri*.
- The reference strain was isolated in 1984 from a man in China who carried the disease
- The strain was cultured at 37°C overnight on tryptic soy agar containing 0.01% Congo red. Colonies were inoculated into tryptic soy broth and grown to stationary phase at 37°C for isolating plasmid and chromosomal DNAs

Outline

- Shigellosis, caused by *S. flexneri*, is one of the leading causes of death in young children in developing countries.
- Sequencing the genome mainly involved automation to reduce human-induced errors.
- Comparison between *S. flexneri* and its genetic relative, *E. coli.*, revealed distinct and similar characteristics between their chromosomes.
- Viable database for this organism provides a fast and easy way to explore its genome.

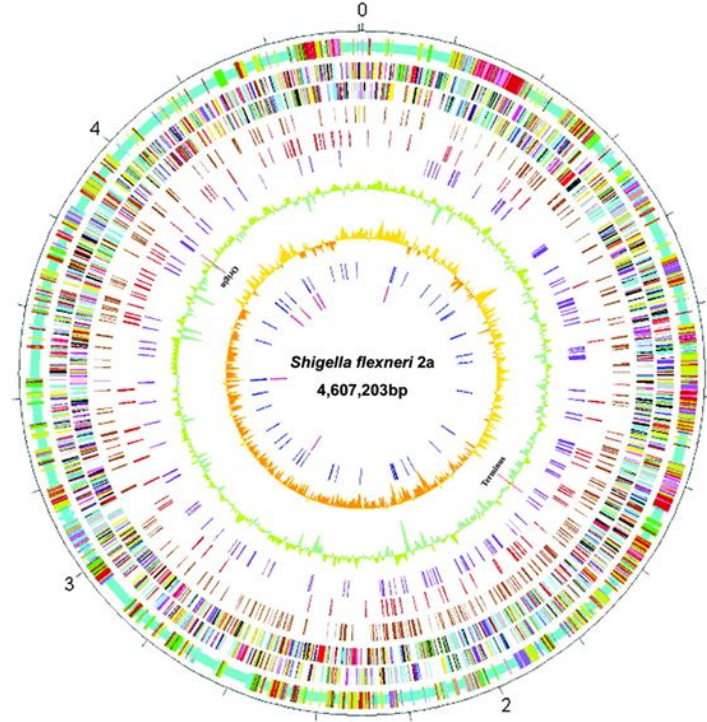
Automated Shotgun Sequencing was Used to Create the Genome

- Initially involved the employment of a highly accurate base-calling software, called *phred*
- Significantly reduced human interaction with the DNA sequences, thus also reducing the errors that would have resulted from human involvement
- After reaching 318 overlapping regions in the species' genome, a program called *consed* was then used for sequence finishing.
- Identifying open reading frames involved the Glimmer 2.0 software, but some manual inspection was still employed for overlapping ORFs.
- The databases BLASTP and COGs were used to identify families of related proteins. Genomic comparison with *E. coli* K12 was then executed using the GenomeComp software.

Outline

- Shigellosis caused by *S. flexneri* is one of the leading causes of death in young children in developing countries.
- Sequencing the genome mainly involved automation to reduce human-induced errors.
- Comparison between *S. flexneri* and its genetic relative, *E. coli.*, revealed distinct and similar characteristics between their chromosomes.
- Viable database for this organism provides a fast and easy way to explore its genome.

Circular Genome Map of Sf301 Chromosome Compared with those of *E.coli* K12 MG1655 and 0157 EDL933



Qi Jin et al. Nucl. Acids Res. 2002;30:4432-4441

Nucleic Acids Research

Comparing the Chromosomes of *S. flexneri* and *E. coli* Shows their Similarities

Table 1. General features of the Sf301 genome compared with genomes of *E.coli* K12 and 0157, and the virulence plasmid, pWR501, from *S.flexneri* M90T 5a

Chromosome	Sf301	MG1655 ^a	EDL933 ^b
Total length (bp)	4 607 203	4 639 221	5 528 445
No. of total ORFs	4434	4289	5349
Average length of ORFs (bp)	891	954	905
Percentage of coding sequence (%)	80.4	87.8	87.1
G + C content			
Total genome (%)	50.89	50.79	50.40
Protein coding regions (%)	51.95	51.85	51.51
RNA genes (%)	54.79	54.84	54.88
Intergenic regions (%)	46.07	42.28	42.76
Ribosomal RNA			
No. of 16S	7	7	7
No. of 23S	7	7	7
No. of 5S	8	8	8
No. of transfer RNA	97	92	93
No. of tmRNA	1	1	1
No. of non-classical RNA	9	5	5
Translocations and inversions ^c	13	–	1
IS elements	314	39	40
Of which partial copies	67	7	19
Plasmid	pCP301	pWR501 ^d	
Total length (bp)	221 618	221 851	
No. of total ORFs	267	293	
Average length of ORFs (bp)	658	636	
Percentage of coding sequence	76.24	82.09	
G + C content			
Total (%)	45.77	46.36	
Coding regions (%)	46.13	46.95	
Intergenic regions (%)	44.59	43.69	
IS elements	88	92	
Of which partial copies	62	69	

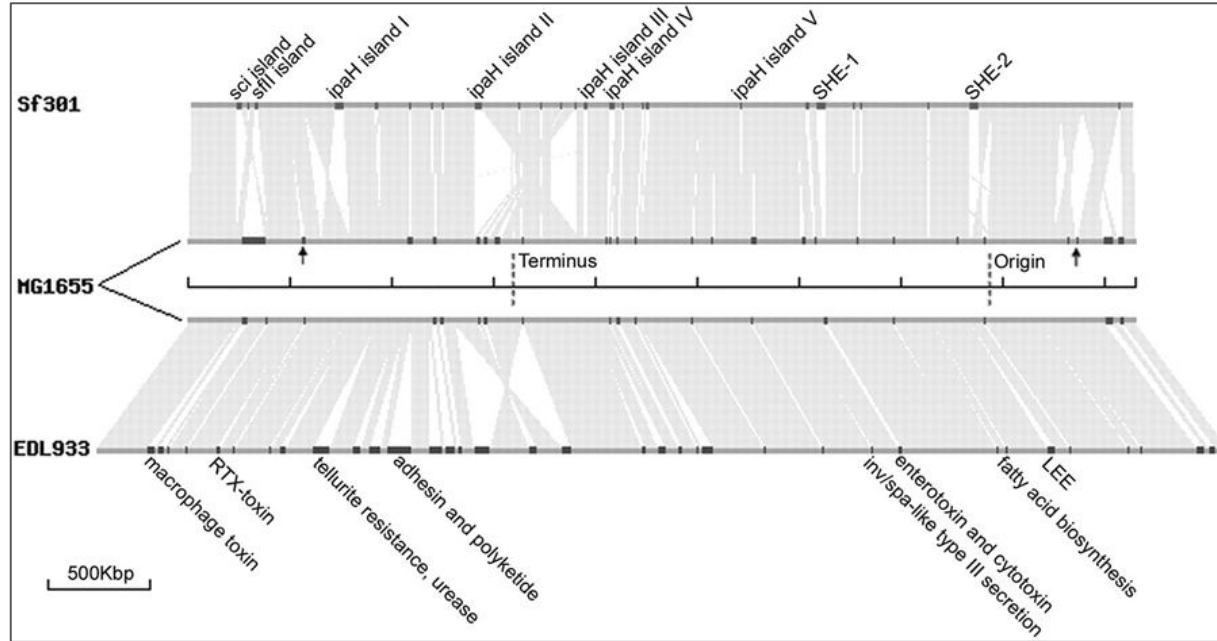
^aData are from Blattner *et al.* (10).

^bData are from Perna *et al.* (11).

^cOnly those with DNA segments >5 kb are listed.

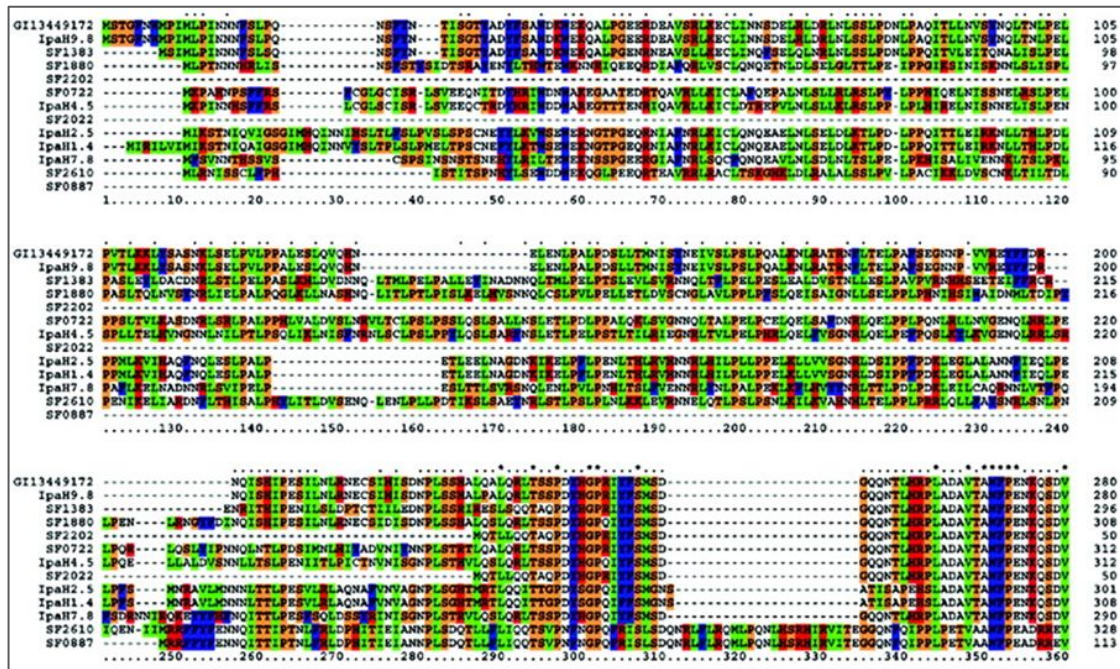
^dData are from Venkatesan *et al.* (8).

Schematic Representation of Translocations and Inversions, and Strain-Specific Islands



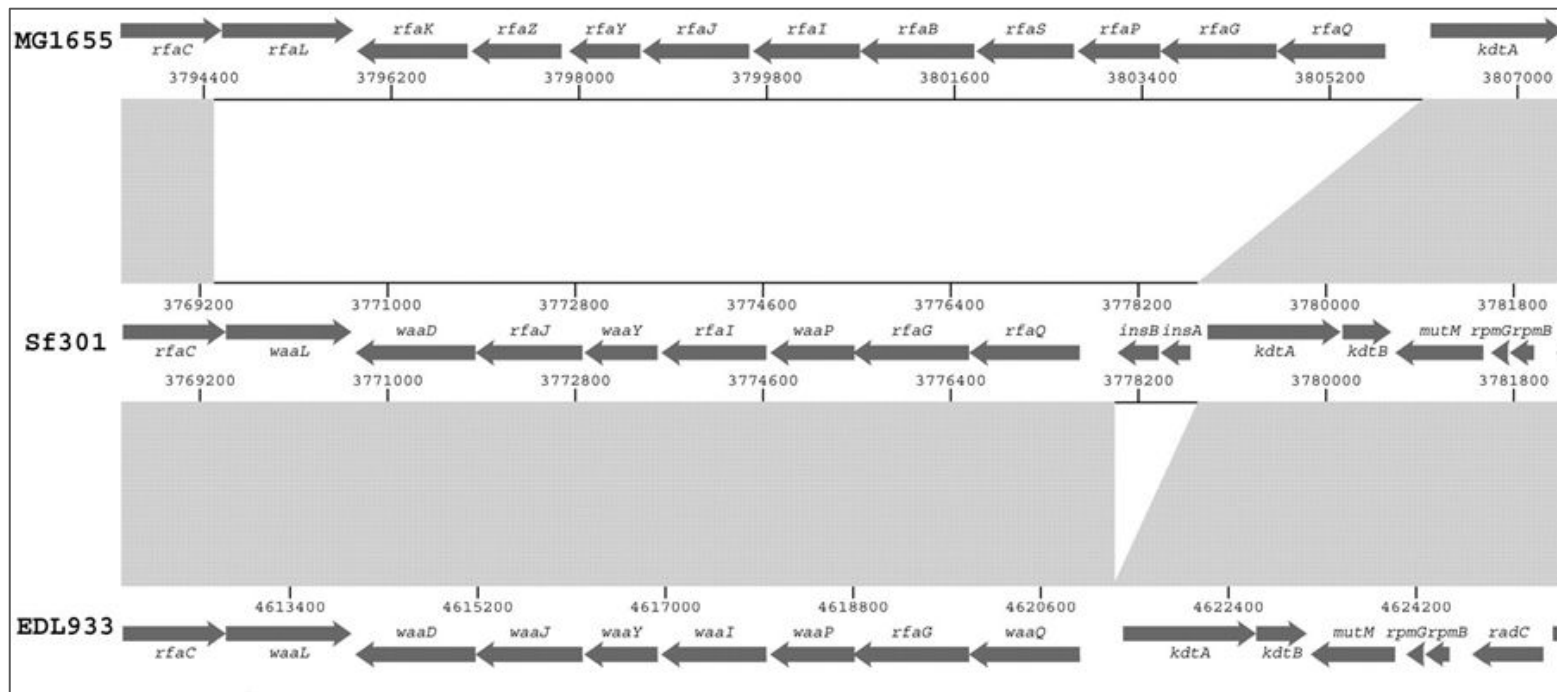
Qi Jin et al. Nucl. Acids Res. 2002;30:4432-4441

Amino Acid Sequence Alignment of N-terminal Halves of IpaH Proteins were Identified in Sf301



Qi Jin et al. Nucl. Acids Res. 2002;30:4432-4441

Comparison of the rfa/waa Region



Qi Jin et al. Nucl. Acids Res. 2002;30:4432-4441

Insertion sequence elements found in the three strains were also compared with those of 5a

Table 2. IS elements identified in genomes of Sf301, MG1655 and EDL933, the virulence plasmid, and pWR501, from *S.flexneri* 5a

Name	Length (bp)	No. of ORFs	No. of intact elements					No. of partial elements				
			Sf301	K12	0157	pCP301	pW501	Sf301	K12	0157	pCP301	pWR501
IS1	768	2	108	6	2	2	3	9	0	0	1	1
iso-IS1	803	2	0	0	0	0	0	1	0	3	5	5
IS2	1331	2	30	6	1	1	2	5	1	0	2	2
IS3	1258	2	5	5	0	0	0	3	0	2	7	8
IS4	1428	2	18	1	0	1	1	3	0	0	1	2
IS5	1198	1	0	10	0	0	0	0	1	0	0	0
iso-IS10R*	1329	1	13	0	0	2	0	0	0	0	0	0
IS21	2131	2	0	0	0	0	0	0	0	0	3	3
IS91	1830	1	3	0	0	0	0	2	0	0	6	6
IS100	1963	2	0	0	0	0	0	0	0	0	7	6
IS150	1443	3	0	1	0	0	0	5	0	0	2	2
IS186	1372	1	0	3	0	0	0	0	0	0	0	0
IS600	1264	2	35	0	0	3	2	17	1	6	10	13
IS629	1310	2	10	0	18	8	5	11	0	3	3	9
IS630	1164	1	0	0	0	1	1	0	0	4	2	2
IS911	1250	2	16	0	0	1	1	0	4	0	0	0
IS1294	1714	1	0	0	0	1	2	3	0	0	7	4
ISS#1	929	1	0	0	0	1	1	0	0	0	2	3
ISS#2	1374	1	6	0	0	2	2	0	0	0	1	0
ISS#3	1302	1	0	0	0	1	1	1	0	0	1	1
ISS#4	2754	3	3	0	0	2	2	7	0	1	2	2
Total			247	32	21	26	23	67	7	19	62	69

*iso-IS10R is a homolog of IS10R identified in Sf301 in this study.

Pseudogenes were identified

Table 3. Pseudogenes with known functions identified in Sf301 genome

Pathway	Mutation	Description
Carbohydrate metabolism		
<i>araA</i>	Stop codon	L-Arabinose isomerase; arabinose catabolism
<i>ugd</i>	Stop codon	UDP-glucose 6-dehydrogenase; colanic acid synthesis
<i>fucK</i>	Stop codon	1-Funulokinase, fucose catabolism
<i>glcD</i>	Stop codon	Glycolate oxidase subunit D
<i>xylA</i>	Stop codon	D-Xylose isomerase; D-xylose catabolism and D-glucose conversion
<i>aceB</i>	Stop codon	Malate synthetase A; glyoxylate bypass
<i>dgoA</i>	Stop codon	D-Galactonate hydro-lyase; galactonate catabolism
<i>fdhF*</i>	Stop codon	Formate dehydrogenase-H; anaerobic respiration
<i>zwf</i>	Stop codon	G6PD; oxidative branch of pentose phosphate pathway
Energy metabolism		
<i>cyoB</i>	Stop codon	Cytochrome o ubiquinol oxidase subunit I; active under high oxygen growth conditions
<i>cyoA</i>	Truncation	Cytochrome o ubiquinol oxidase subunit II; as <i>cyoB</i>
<i>acs</i>	Stop codon	Acetyl-CoA synthetase; scavenging acetate
<i>hyfB</i>	Stop codon	Hydrogenase 4 subunit; anaerobic respiration
<i>narZ</i>	Stop codon	NRZ; anaerobic terminal electron acceptor
<i>torA</i>	Stop codon	Trimethylamine N-oxide reductase subunit; electron acceptor (anaerobic respiration)
<i>torD</i>	Insertion	Chaperone of TorA; preventing TorA degradation
Lipid metabolism		
<i>hcaD</i>	Stop codon	Ferredoxin reductase; utilization of aromatic acids
Amino acid metabolism		
<i>speF</i>	Stop codon	Ornithine decarboxylase isozyme; putrescine synthesis
<i>speG</i>	Frame shift	Spermidine acetyltransferase; polyamine synthesis
<i>nadB</i>	Stop codon	Quinolinate thymethylase B; pyridine synthesis
<i>gabD</i>	Stop codon	Succinate-semialdehyde dehydrogenase; aminobutyrate catabolism
<i>mga</i>	Frame shift	Peptidoglycan enzyme; cell wall formation
<i>mtaA</i>	Truncation	Homoserine transsuccinylase; methionine synthesis
<i>actC</i>	Stop codon	Acetylornithine transaminase; arginine catabolism
Cofactors and vitamins		
<i>nfHb</i>	Insertion	Dihydropteridine reductase; recycling the quinoid dihydrobiopterin cofactor by reducing it
<i>lhr</i>	Stop codon	ATP-dependent helicase, dispensable
<i>lplA</i>	Frame shift	Lipoate-protein ligase A; ligation of lipoyl to apoprotein
Complex lipids		
<i>gldA</i>	Stop codon	Glycerol dehydrogenase; glycerol dissimilation
Complex carbohydrates		
<i>ycjM</i>	Insertion	Putative polysaccharide hydrolase
<i>otsA</i>	Truncation	Trehalose-6-phosphate synthase; response to high osmolarity
<i>aceK</i>	Stop codon	Isocitrate dehydrogenase kinase/phosphatase; control flux between the TCA cycle and the glyoxylate bypass
Translation		
<i>prfB</i>	Stop codon	Peptide chain release factor RF-2
Transport		
<i>araF</i>	Stop codon	L-Arabinose-binding periplasmic protein
<i>cysW</i>	Stop codon	Sulfate transport system permease W protein
<i>yhdX</i>	Truncation	Permease; putative amino acid ABC transporter
<i>ugpC</i>	Insertion	ATP-transporter; glycerol-3-phosphate uptake
<i>hbaA</i>	Insertion	ATP-binding component; D-ribose transport
<i>hbtB</i>	Stop codon	ABC transporter; D-ribose periplasmic binding protein
<i>hbtA</i>	Frame shift	6-Phospho- β -glucosidase; arbutin fermentation
<i>ptsA</i>	Stop codon	PEP-protein phosphotransferase system enzyme I
<i>ypfH</i>	Stop codon	ABC transporter; periplasmic binding

Signal transduction		
<i>ctbB</i>	Truncation	Regulator (paired with <i>ctbR</i>); citrate fermentation
<i>kdpE</i>	Stop codon	Regulator of the <i>kdp</i> operon; potassium transport
<i>kdpD</i>	Stop codon	Sensor of the <i>kdpDE</i> system; potassium transport
<i>narQ</i>	Stop codon	Nitrate/nitrite sensor protein; acts on NarL/NarP
<i>arp</i>	Stop codon	Regulator of acetyl CoA synthetase
<i>malT</i>	Stop codon	Positive regulator of <i>mal</i> operon
Cell motility		
<i>flaA</i>	Frame shift	σ^{28} for flagellar operons
<i>flgF</i>	Stop codon	Cell-proximal portion of basal-body rod
<i>flgK</i>	Stop codon	Hook-filament junction protein 1
<i>flgL</i>	Stop codon	Hook-filament junction protein
<i>fljF</i>	Stop codon	Basal-body MS-ring and collar protein
<i>fljI</i>	Truncation	FljJ protein
<i>flhA</i>	Stop codon	Export of flagellar proteins
Unassigned enzymes		
<i>tesA</i>	Stop codon	Acyl-CoA thioesterase I; hydrolyzes long chain acyl thioesters
<i>pphA</i>	Stop codon	Protein phosphatase 1; modulates phosphoproteins signaling protein misfolding

Table 3. Continued

Pathway	Mutation	Description
pphB		
<i>pphB</i>	Stop codon	Removal of a phosphate group attached to serine or threonine residue; signaling protein misfolding through cpxRA system
Unassigned non-enzymes		
<i>yuaJ</i>	Stop codon	Transport protein; sodium/alanine symporter
<i>nfrA</i>	Stop codon	Omp; bacteriophage N4 receptor
<i>csyG</i>	Stop codon	Transporter; curli assembly
<i>csyA</i>	Insertion	Curlin major subunit; coiled surface structures
<i>fehE</i>	Stop codon	Transporter; ferric enterobactin (enterochelin)
<i>fhuE</i>	Stop codon	Omp; receptor for ferric iron uptake
<i>entC</i>	Stop codon	Isochorismate synthase; enterobactin biosynthesis
<i>hlyE</i>	Stop codon	Hemolysin E; hemolytic to sheep blood
<i>hslJ</i>	Truncation	Heat shock protein HslJ
<i>uidB</i>	Truncation	Transporter; specific to α - and β -glucuronides
<i>celD</i>	Insertion	Negative regulator of <i>cel</i> operon (cryptic); ferment cellobiose, arbutin and salicin
<i>molR</i>	Insertion	Molybdate metabolism regulator, first fragment
<i>molR_2</i>	Stop codon	Molybdate metabolism regulator, fragment 2
<i>cirA</i>	Stop codon	Porin and receptor; colicin I uptake
<i>focB</i>	Frame shift	Formate transporter (formate channel 2)
<i>emrA</i>	Stop codon	Multidrug resistance secretion protein
<i>ppdA</i>	Frame shift	Prepilin peptidase dependent protein A
<i>glcF</i>	Frame shift	Glycolate oxidase iron-sulfur subunit; ferridoxin related
<i>aer</i>	Stop codon	Aerotaxis sensor receptor; transducing signals for aerotaxis
<i>ompG</i>	Truncation	Outer membrane protein; forms large channels
<i>yaeG</i>	Stop codon	Regulator of D-galactarate, D-glucarate and D-glycerate metabolism
<i>nagD</i>	Stop codon	N-Acetylglucosamine metabolism
<i>fimD</i>	Insertion	Export and assembly of type I fimbriae

**fdhF* has a stop codon (UAA) in addition to the stop codon UGA used for introducing selenocysteine.

Chinese-operated Database Presents a Viable Source of Genome Information on *Shigella*

SHIBASE Quick Guide User Help

Genome maps

- *S. dysenteriae* Sd197
- *S. flexneri* SF301
- *S. boydii* Sb227
- *S. sonnei* Ss046

Comparison

- General features
- Linear alignment
- Orthologs order
- Metabolic pathway
- Virulence factors
- CGH results

Analysis tools

- Text query
- BLAST search
- Shi-align service

Links

- GenomeComp
- VFDB database

About Shigella

Shigella is a group of Gram-negative, facultative intracellular pathogens. Recognised as the etiologic agents of bacillary dysentery or shigellosis in the 1890s, *Shigella* was adopted as a genus in the 1950s and subgrouped into four species: *S. dysenteriae*, *S. flexneri*, *S. boydii* and *S. sonnei* (also designated as serogroups A to D).

Shigella grows only in the intestinal tract of humans. It's transmitted by the fecal-oral route. Fliers, fingers, and food are the usual vehicles. But because *Shigella* cells survive for a long time in contaminated water or on fomites, they transmit it too. People who live in crowded conditions where cleanliness is difficult are particularly likely to contract shigellosis. Children are far more likely than adults to get shigellosis. Those under 5 year old account for about half the reported cases, because they are too young to follow good hygiene habits and they are more susceptible to *Shigella* infection.

Many fecal-oral infections, including cholera and typhoid fever, have been nearly eradicated from industrialized countries, but not shigellosis. Shigellosis is difficult to eradicate partly because it is so infectious. A person must ingest thousands to millions of bacterial cells to contract typhoid fever or cholera, but only 200 cells are sufficient to cause shigellosis. It is well-established that the virulence plasmid (VP) carries the primary virulence genes that enable the invasiveness of the bacteria in the colon and the rectum and the induction of apoptosis to resident macrophages and dendritic cells, leading to inflammatory infection.

For more detailed information about the pathogenesis, epidemiology, diagnosis and clinical aspects of *Shigella*, please refer the recent comprehensive reviews:

- † Jennison AV, Verma NK, 2004. *Shigella flexneri* infection: pathogenesis and vaccine development. *FEMS Microbiol Rev* 28(1):43-58.
- † Phalipon A, Sansonetti P, 2003. Shigellosis: innate mechanisms of inflammatory destruction of the intestinal epithelium, adaptive immune response, and vaccine development. *Crit Rev Immunol* 23(5-6):371-401.
- † Fernandez MI, Sansonetti PJ, 2003. *Shigella* interaction with intestinal epithelial cells determines the innate immune response in shigellosis. *Int J Med Microbiol* 293(1):55-67.
- † Sansonetti P, 2002. Host-pathogen interactions: the seduction of molecular cross talk. *Cell Suppl* 3:1112-8.
- † Nhieue GT, Enninga J, Sansonetti P, Grompone G, 2005. Tyrosine Kinase signaling and type III effectors orchestrating *Shigella* invasion. *Curr Opin Microbiol* 8(1):16-20.

About SHIBASE

As shigellosis remains to be the top one diarrhoeal disease in China, the Chinese government has given strong support to the genome sequence project of *Shigella*. During the next five years, we determined the genomes of

SHIBASE Quick Guide User Help

Genome maps

- *S. dysenteriae* Sd197
- *S. flexneri* SF301
- *S. boydii* Sb227
- *S. sonnei* Ss046

Comparison

- General features
- Linear alignment
- Orthologs order
- Metabolic pathway
- Virulence factors
- CGH results

Analysis tools

- Text query
- BLAST search
- Shi-align service

Links

- GenomeComp
- VFDB database

Quick navigations

Locate genome element by ID

Go to page of individual element (ORF, RNA and IS) by its genome synonym or *Shi*BASE ID. For example, 'SF1006' or 'HJS06391'.

Input its genome synonym or *Shi*BASE ID:

Go

Locate linear map by position

Go to detailed linear map of the specified region in *Shigella* genomes.

60 Kb around bp of *S. flexneri* 2a strain 301 show GC curve

Go

Locate linear sequence comparison by position

Go to detailed linear sequence comparison of the specified region in *Shigella* genomes.

20 Kb around bp of *S. dysenteriae* 1 strain 197

Go

Retrieve sub-sequence from genome by position

Get the specified segment from *Shigella* genomes in FASTA format.

10 Kb around bp of *S. dysenteriae* 1 strain 197

Go

List genomic elements other than ORFs

Get the list of special genomic elements (pseudogene, IS and RNA) in *Shigella* genomes.

List all from *S. dysenteriae* 1 strain 197

Go

Copyright © 2005 State Key Laboratory for Molecular Virology and Genetic Engineering, Beijing, China

ShiBASE Provides More Query Options for Finding Specific Genes

SHIBASE

Quick Guide User Help

Genomic maps

- S. dysenteriae* Sd197
- S. flexneri* SF301
- S. boydii* Sb227
- S. sonnei* Ss046

Comparison

- General features
- Linear alignment
- Orthologs order
- Metabolic pathway
- Virulence factors
- CGH results

Analysis tools

- Text query
- BLAST search
- Sh-align service

Links

- GenomeComp
- VFDB database

Search ShiBASE

Quick Search

Gene/Synonym Search

Refined Search

Synonym contains

AND Gene contains

AND Product contains

AND EC code contains

AND COG code contains

Search

Search in

[All - None - Invert]

Chromosomes

Sd197

SF301

Sb227

Ss046

Plasmids

pSD1_197

pCP301

pSB4_227

pSS_046

Reset all

Copyright © 2005 State Key Laboratory for Molecular Virology and Genetic Engineering, Beijing, China

SHIBASE

Quick Guide User Help

Genomic maps

- S. dysenteriae* Sd197
- S. flexneri* SF301
- S. boydii* Sb227
- S. sonnei* Ss046

Comparison

- General features
- Linear alignment
- Orthologs order
- Metabolic pathway
- Virulence factors
- CGH results

Analysis tools

- Text query
- BLAST search
- Sh-align service

Links

- GenomeComp
- VFDB database

Search results

Total 1 hits found in 1 pages

Begin Previous Current: 1 Next End

Synonym	Genome	Gene	Start	End	Product
<input checked="" type="checkbox"/> SF3011	SF301	insA	3105199	3105474	IS1 ORF1

[Check All - Clear All - Invert Selection]

Linear maps

Download sequences of selected elements in FASTA format

Copyright © 2005 State Key Laboratory for Molecular Virology and Genetic Engineering, Beijing, China

SHIBASE

Quick Guide User Help

Genome maps

- S. dysenteriae* Sd197
- S. flexneri* SF301
- S. boydii* Sb227
- S. sonnei* Ss046

Comparison

- General features
- Linear alignment
- Orthologs order
- Metabolic pathway
- Virulence factors
- CGH results

Analysis tools

- Text query
- BLAST search
- Sh-align service

Links

- GenomeComp
- VFDB database

Shigella flexneri serotype 2a strain 301

SF3011

Gene insA

Strand Minus

Start 3105199

End 3105474

G+C 52.90% (Genome average: 50.89%) [Show G+C content curve]

COG SC06387Z

Code 1

Product IS1 ORF1

DNA

```
GTGGTTCACATTCACAGAAATGCTTCTTCCTCCCTACTAAAGCGCGTGGCTAAAGCGCAAGCAGCTCCGGG  
ACACTGAGGTTAATCTCTCTCTCATTCGCTAAACAGGCGAGCTACTAGCAGGCTTCTAGCAGGCTTCTAGCAGGCT  
CGCACAGAAAATCATGATATGGCATGAAATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGATGGAT  
AACCAGGTTTATGCTCACTTAAAGACTCGAGCGCGGCTCGGTAA
```

Download [10 kb] segment around Get

Send to BLAST with NR

Copyright © 2005 State Key Laboratory for Molecular Virology and Genetic Engineering, Beijing, China

Summary

- Combatting Shigellosis is of great concern in developing countries like China.
- Discovering the pathogenicity of *S. flexneri* required the sequencing of its entire genome.
- The research group used softwares in order to efficiently conduct the shotgun sequencing process of *S. flexneri*.
- The results revealed the extreme similarities between 5a and 2a serotypes of *S. flexneri* and between Sf301 and *E. coli* K12.
- A database, called *Shi*BASE, developed by Chinese researchers presents these compiled information to other potential scientists.

Acknowledgments

- Dr. Dahlquist
- Dr. Dionisio
- Biological Database students

Questions?