Genomics Data 14 (2017) 63–65

Contents lists available at ScienceDirect

Genomics Data

journal homepage: www.elsevier.com/locate/gdata

Data in Brief

Whole genome shotgun sequencing of Indian strains of *Streptococcus agalactiae*

Balaji Veeraraghavan[a,*], Naveen Kumar Devanga Ragupathi[a], Sridhar Santhanam[b], Valsan Philip Verghese[c], Francis Yesurajan Inbanathan[a], Charles Livingston[a]

[a] Department of Clinical Microbiology, Christian Medical College, Vellore 632004, India
[b] Department of Neonatology, Christian Medical College, Vellore 632004, India
[c] Department of Child Health, Christian Medical College, Vellore 632004, India

ARTICLE INFO

Keywords:
Streptococcus agalactiae
Whole genome shotgun sequencing
Antimicrobial resistance
MLST
Group B Streptococcus

ABSTRACT

Group B streptococcus is known as a leading cause of neonatal infections in developing countries. The present study describes the whole genome shotgun sequences of four Group B Streptococcus (GBS) isolates. Molecular data on clonality is lacking for GBS in India. The present genome report will add important information on the scarce genome data of GBS and will help in deriving comparative genome studies of GBS isolates at global level. This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession numbers NHPL00000000 – NHPO00000000.

Specifications	
Organism/cell line/tissue	*Streptococcus agalactiae*
Sex	NA
Sequencer or array type	Ion Torrent PGM
Data format	Processed
Experimental factors	*S. agalactiae* strains were cultured on blood agar medium. Genomic DNA from cultures was isolated using QIAamp DNA mini kit (Qiagen, Germany).
Experimental features	*Sequencing was performed according to Illumina specific protocols for library preparation and DNA-seq.*
Consent	NA
Sample source location	Clinical *S. agalactiae* strains were isolated from blood specimen received at the Department of Clinical Microbiology, Christian Medical College, Vellore, India (12.9248° N 79.1354° E).

1. Direct link to deposited data

Data have been deposited in repository https://www.ncbi.nlm.nih.gov/bioproject/PRJNA387519 and the sequence files are accessible under the accession numbers NHPL00000000 – NHPO00000000.

2. Introduction

Group B *Streptococcus* (GBS) is one of the leading causes of neonatal infections in developing countries [1]. Between 1998 and 2010, we have reported an early onset infection incidence among newborn of 0.68/1000 live births in our hospital [2]. There is lack of data on molecular epidemiology and genomic content of GBS in India. The present study reports on draft genome sequences of clinical *S. agalactiae* strains.

2.1. Experimental design, materials and methods

Four clinical *S. agalactiae* strains (VB11227, VB12497, VBP4522 and VBP3124) from blood stream infections were isolated at the Department of Clinical Microbiology, Christian Medical College,

* Corresponding author.
E-mail address: vbalaji@cmcvellore.ac.in (B. Veeraraghavan).

http://dx.doi.org/10.1016/j.gdata.2017.10.001
Received 1 July 2017; Received in revised form 26 September 2017; Accepted 1 October 2017
Available online 02 October 2017

Table 1
Genome data of Group B Streptococcus (n = 4) from blood stream infections.

Isolate ID	Sequence types	Serotypes	Year of isolation	Total size (bp)	Coverage	CDS	Contigs	CRISPR finder	AMR genes	Toxin and virulence genes	Streptococcal and Enterobacteriaceae plasmids	Accession
VB11227	103	Ia	2017	2,060,254	85 ×	2107	21	–	tet(M), msr(D), mef(A)	cylE beta hemolysin, hemolysin III, cfb CAMP factor, phagelysin	–	NHPO00000000
VB12497	249	Ia	2017	1,982,686	90 ×	2024	46	6	–	cylE beta hemolysin, hemolysin III, cfb CAMP factor, phagelysin	–	NHPN00000000
VBP3124	23	Ia	2017	2,070,970	81 ×	2163	124	14	tet(M)	cylE beta hemolysin, hemolysin III, cfb CAMP factor	–	NHPL00000000
VBP4522	1	V	2017	2,067,355	80 ×	2092	44	5	tet(M)	cylE beta hemolysin, hemolysin III, cfb CAMP factor	–	NHPM00000000

Vellore, India. Of these, VBP4522 and VBP3124 were from pediatric patients. All four isolates were identified to be group B using latex agglutination.

The isolates were sequenced further to investigate their whole genome for understanding of their genetic arrangements. DNA was isolated using QiAamp DNA mini Kit (Qiagen, Germany). Whole genome shotgun sequencing was done in the Ion Torrent PGM platform (Life Technologies) with 400 bp chemistry. *De novo* assembly of the raw reads was performed using AssemblerSPAdes v.5.0.0.0 embedded in Torrent suite server v.5.0.5. Genome annotation of the assembled sequences were performed in the PATRIC database (the bacterial bioinformatics database and analysis resource) (http://www.patricbrc.org) [3], and the NCBI Prokaryotic Genome Automatic Annotation Pipeline (PGAAP) (http://www.ncbi.nlm.nih.gov/genomes/static/Pipeline.html). Downstream analysis was completed using the Center for Genomic Epidemiology (CGE) server (http://www.cbs.dtu.dk/services), and PATRIC database. Specific components like antimicrobial resistance (AMR) genes and plasmids were made using ResFinder 2.1 and PlasmidFinder 1.3 tools from the CGE server [4–5]. Further analysis to identify the clusters of regularly interspaced short palindromic repeats (CRISPR) and spacer sequences in the genome were performed using CRISPR finder (http://crispr.u-psud.fr/Server/) [6]. To identify the clonality, the sequence types (STs) of *S. agalactiae* isolates were investigated using the following house-keeping genes *adhp*, *atr*, *glck*, *glna*, *phes*, *sdha* and *tkt* by comparing with the standard references available at the MLST 1.8 database (https://cge.cbs.dtu.dk//services/MLST/). To visualize the possible evolutionary relationships between isolates, STs of the study isolates and the globally reported strains were computed using PHYLOViZ software v2.0 based on goeBURST algorithm [7].

3. Data description

The size of the *S. agalactiae* genomes ranged from ~1.9 to ~2.1 Mbp with coverage of 80 × to 90 × (Table 1). The number of coding DNA sequences (CDS) per genome ranged between 2024 and 2163. Serotype of the GBS isolates were identified using the whole genome data, by performing *in silico* PCR for capsular polysaccharide genes [8]. Primers and interpretations were used as described by Imperi et al. [9] for *in silico* PCR. VBP4522 was identified to be serotype V, and VB11227, VB12497 and VBP3124 were identified as serotype Ia.

Three out of four *S. agalactiae* isolates carried AMR genes. VB11227 had *tet*(M) gene responsible for tetracycline resistance and *msr*(D), *mef*(A) genes for macrolide resistance respectively. All three genes were mostly reported in *Streptococcus* spp. All isolates were positive for *cyl*E beta hemolysin, hemolysin III and *cfb* gene responsible for CAMP factor. All four isolates were negative for plasmids. However, three isolates carried CRISPR regions in their genomes (Table 1).

The MLST data reveals that the study isolates are of different phylogeny and is suggestive of different clones circulating in India. goeBURST reveals 68 clonal complexes from the available datasets globally. The goeBURST diagram shows that the sequence types ST-1 and ST-103 from the study isolates belong to clonal complex (CC) 1 with ST1 as founder ST, whereas, ST-23 and ST249 belong to a different clonal complex (CC2) with ST23 as founder ST (Fig. 1). Further studies on molecular epidemiology will provide a baseline data of the GBS clones available in India.

Conflict of interest

The authors declare that there is no conflict of interest.

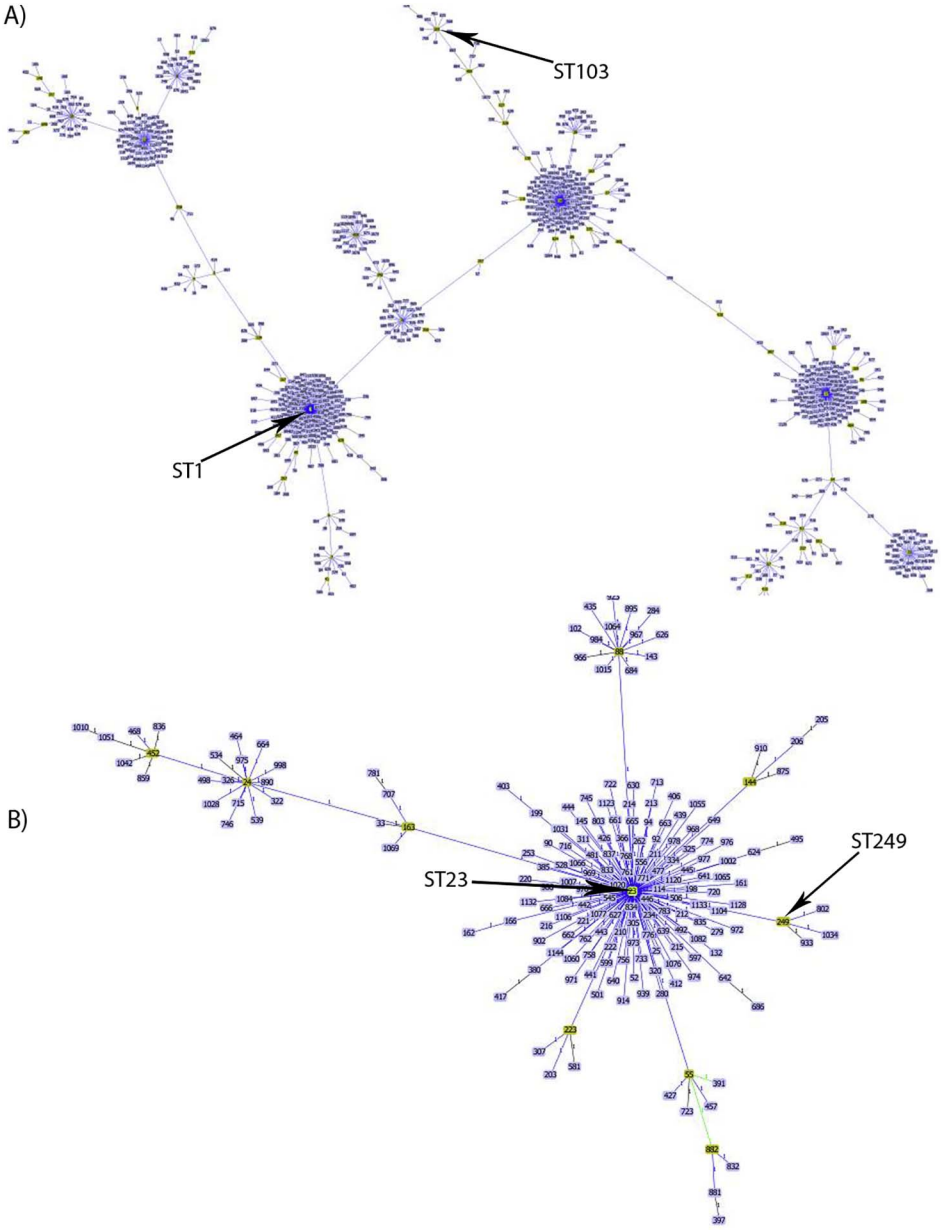

Fig. 1. goeBURST algorithm for Group B Streptococcus was calculated using PHYLOViZ software v2.0. The figure depicts, A) ST1 and ST103 belonging to clonal complex 1 with ST1 as founder ST, and B) ST23 and ST249 belonging to clonal complex 2 with ST23 as founder ST.

References

[1] P.Y. Iroh Tam, S.F. Delair, S.K. Obaro, Neonatal group B streptococcus disease in developing countries: are we ready to deploy a vaccine? Expert Rev. Vaccines 14 (2015) 1401–1403.

[2] S. Sridhar, R. Grace, P.J. Nithya, V. Balaji, T. Niranjan, K. Manish, K.K. Anil, J.K. Atanu, Group B, Streptococcal infection in a tertiary hospital in India—1998–2010, Pediatr. Infect. Dis. J. 33 (2014) 1091–1092.

[3] A.R. Wattam, D. Abraham, O. Dalay, T.L. Disz, T. Driscoll, J.L. Gabbard, J.J. Gillespie, R. Gough, D. Hix, R. Kenyon, D. Machi, C. Mao, E.K. Nordberg, R. Olson, R. Overbeek, G.D. Pusch, M. Shukla, J. Schulman, R.L. Stevens, D.E. Sullivan, V. Vonstein, A. Warren, R. Will, M.J.C. Wilson, H.S. Yoo, C. Zhang, Y. Zhang, B.W. Sobral, PATRIC, the bacterial bioinformatics database and analysis resource, Nucleic Acids Res. 42 (2014) D581–D591.

[4] E. Zankari, H. Hasman, S. Cosentino, M. Vestergaard, S. Rasmussen, O. Lund, F.M. Aarestrup, M.V. Larsen, Identification of acquired antimicrobial resistance genes, J. Antimicrob. Chemother. (2012), http://dx.doi.org/10.1093/jac/dks261.

[5] A. Carattoli, E. Zankari, A. Garcia-Fernandez, M. Voldby Larsen, O. Lund, L. Villa, F.M. Aarestrup, H. Hasman, PlasmidFinder and pMLST: in silico detection and typing of plasmids, Antimicrob. Agents Chemother. 58 (2014) 3895–3903.

[6] I. Grissa, G. Vergnaud, C. Pourcel, CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats, Nucleic Acids Res. 35 (Web Server issue) (2007) W52–W57.

[7] A.P. Francisco, C. Vaz, P.T. Monteiro, J. Melo-Cristino, M. Ramirez, J.A. Carriço, PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods, BMC Bioinf. 13 (2012) 87.

[8] R.M. San Millán, I. Martínez-Ballesteros, A. Rementeria, J. Garaizar, J. Bikandi, Online exercise for the design and simulation of PCR and PCR-RFLP experiments, BMC Res. Notes 6 (2013) 513.

[9] M. Imperi, M. Pataracchia, G. Alfarone, L. Baldassarri, G. Orefici, R. Creti, A multiplex PCR assay for the direct identification of the capsular type (Ia to IX) of *Streptococcus agalactiae*, J. Microbiol. Methods 80 (2010) 212–214.