GENOME ANNOUNCEMENT – Physiology & Biochemistry

Complete genome sequence of BS49 and draft genome sequence of BS34A, *Bacillus subtilis* strains carrying Tn916

Hilary P. Browne[1,#], Seyed Yahya Anvar[2,#], Jeroen Frank[2], Trevor D. Lawley[1], Adam P. Roberts[3] and Wiep Klaas Smits[4,*]

[1]Host-Microbiota Interactions Laboratory, Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1SA, UK, [2]Leiden Genome Technology Center, Human and Clinical Genetics, Leiden University Medical Center, 2300 RC, Leiden, The Netherlands, [3]Department of Microbial Diseases, UCL Eastman Dental Institute, WC1X 8 LD London, UK and [4]Department of Medical Microbiology, Leiden University Medical Center, 2300 RC, Leiden, The Netherlands

*Corresponding author: Department of Medical Microbiology, Leiden University Medical Center, PO Box 9600, 2300 RC, Leiden, The Netherlands. Tel: +31-71-526 1229; Fax: +31-71-526 6761; E-mail: w.k.smits@lumc.nl
[#]These authors contributed equally.
One Sentence Summary: The genomes of two *Bacillus subtilis* strains with a common mobile element (Tn916) have been sequenced.
Editor: David Studholme

ABSTRACT

Bacillus subtilis strains BS49 and BS34A, both derived from a common ancestor, carry one or more copies of Tn*916*, an extremely common mobile genetic element capable of transfer to and from a broad range of microorganisms. Here, we report the complete genome sequence of BS49 and the draft genome sequence of BS34A, which have repeatedly been used as donors to transfer Tn*916*, Tn*916* derivatives or *oriT*$_{Tn916}$-containing plasmids to clinically important pathogens.

Key words: mobile element; donor; conjugation; conjugative transposon

The BS49 and BS34A strains were originally obtained through introduction of Tn*916* into *Bacillus subtilis* CU2189 (Christie *et al.,* 1987) either by direct transformation of pAM120 plasmid (GenBank U49939.1) (Gawron-Burke and Clewell 1984) (Peter Mullany, personal communication) or by conjugation from *Clostridium difficile* strain FM12A (Roberts *et al.,* 2003), respectively. Tn*916* from strain FM12A was originally transferred from BS49 to *C. difficile* CD37 to give FM12A (Roberts *et al.,* 2003; Brouwer *et al.,* 2012). The BS49 strain contains multiple copies (our unpublished observations), whereas BS34A carries a single copy of Tn*916* (Roberts *et al.,* 2003).

Tn*916* (reviewed in references Roberts and Mullany (2009, 2011)) has been used as a vector to perform (random) mutagenesis of recipient bacteria (Kathariou *et al.,* 1990; Lin and Johnson 1991; Whetzel *et al.,* 2003; Cookson *et al.,* 2011; Mullany *et al.,* 2012), to deliver cargo DNA (Rubens and Heggen 1988; Haraldsen and Sonenshein 2003; Roberts *et al.,* 2003; McBride and Sonenshein 2011) or to mobilize plasmids containing the Tn*916* origin of transfer, *oriT*$_{Tn916}$ (Francis *et al.,* 2013; Mackin *et al.,* 2013). BS49 and BS34A like their close relative *B. subtilis* subsp. *subtilis* demonstrate high levels of competence for genetic transformation. The genome sequences of strains BS49

Table 1. Variants in BS34A compared to BS49. Illumina HiSeq reads of strain BS34A were assembled against the complete genome sequence of strain BS49. Variants were called by Geneious R7 (http://www.geneious.com) using a minimal coverage of 25 and a frequency greater than 0.75. Variants at the ends of the Tn*916* elements and immediately flanking the insertion sites were ignored.

bp start in BS49	bp end in BS49	Nucleotide change	Codon change	Amino acid change	Poly-morphism	Location in BS49	Description
786781	786782	+TA			Insertion	Intergenic region of BS49_08160 and BS49_08170	Insertion restores pseudogene BS49_08161 (BS34A_08161) corresponding to *yetI* of strain 168.
1152310	1152311	(A)7 -> (A)8			Insertion	BS49_11890	Frameshift results in a fusion of BS49_11890 with BS49_11900/*hcpE* (restores pseudogene BS49_11891) (BS34A_11890/*hcpE*), corresponding to *yisK* of strain 168.
1706699	1706699	(A)8 -> (A)7			Deletion	HslU (BS49_18020)	Truncated after 5 amino acids in BS34A (BS34A_17841 pseudogene). Possible alternative start codon annotated (BS34A_17840/*hslU*)
4188826	4188826	G -> A	GCC -> GTC	A -> V	Transition	WalK (BS49_43870)	In PAS domain. Strain 168 has an alanine at this position, like BS49.
4189142	4189142	A -> G	TTC -> CTC	F -> L	Transition	WalK (BS49_43870)	Between TM and HAMP domain. Strain 168 has a leucine at this position, like BS34A.

and BS34A will therefore contribute to the study of Tn*916* conjugation requirements, including the possible role of host factors, thus facilitating the genetic manipulation of various bacterial strains.

Genomic DNA was isolated as described before (He *et al.*, 2013) or using the Qiagen GenomicTip 500/G according to the manufacturer's instructions. Libraries of BS49 and BS34A chromosomal DNA were prepared and sequenced using the Illumina Hi-Seq platform as described (Pettit *et al.*, 2014) (EMBL ERS370048 and EMBL ERS370049). For single molecule real-time sequencing of strain BS49, a SMRTbell DNA template library with an insert size of ∼20 kb was prepared according to the manufacturer's specification. SMRT sequencing was carried out on the Pacific Biosciences RSII machine according to standard protocols (EMBL ERS550338). Sequencing reads were corrected using the HGAP pipeline (Chin *et al.*, 2013) and *de novo* assembly was performed using Celera Assembler 8.1 (http://wgs-assembler.sourceforge.net). After manually curating the sequence of the Tn*916* elements to ensure their accuracy, the chromosome was circularized using strain *B. subtilis* subsp.

subtilis strain 168 (Genbank NC˙000964.3) (Kunst *et al.*, 1997) as a reference and the start of the genome was set at 409 bp upstream of the *dnaA* gene. The single contig was verified by aligning the Illumina reads using Geneious R7 (http://www.geneious.com) to correct for errors in homopolymeric stretches as a result of the consensus calling algorithm.

To generate the BS34A genome sequence, a consensus was generated in Geneious R7 from a reference alignment against the complete BS49 genome. A *de novo* assembly of the BS34A Illumina sequence data was carried out using Velvet (Zerbino and Birney 2008) to reconstruct the two empty target sites of Tn*916* from strain BS49 and to determine the location and context of the single Tn*916* element in strain BS34A. The element was subsequently manually inserted into the consensus BS34A sequence. This genome sequence should be considered an improved high-quality draft (Chain *et al.*, 2009).

Following the assembly, Prokka (Seemann 2014) was used to generate annotations based on a database of trusted protein sequences derived from the genome sequence of strain 168. To identify modified bases, kinetic signals were processed for all

genomic positions after aligning sequencing reads to the final single chromosome sequence of strain BS49. The identification of sequence motifs was performed using the SMRT Portal. The GFF output of SMRT Portal (available as Supporting Information) was converted into an Artemis-compatible TAB file using an in-house script. Raw genome sequence and annotations were merged and the Tn*916* elements were manually annotated in Artemis (Rutherford *et al.*, 2000).

Our analysis of the 4.251.652 bp BS49 genome (EMBL PRJEB7327) yielded 4248 predicted coding sequences, 30 rRNAs, 87 tRNAs, 1 tmRNA and 94 miscellaneous RNAs. We failed to detect an over-represented motif associated with ^{m4}C modifications ($n = 2035$), but found a GACGAG motif associated with ^{m6}A modifications (1255/2446; modified residue underlined).

The sequence of the 18 032 bp Tn*916* elements (bps 1475100-1493131 and 3254988-3273019) (Genbank KM516885) showed that the elements in BS49 are identical to each other and highly similar to the element described for *Enterococcus faecalis* DS16 (GenBank U09422.1). The three single nucleotide variations in the Tn*916* element of BS49 compared to the *Enterococcus* element were [nomenclature from Genbank U09422.1]: (1) the deletion of a guanine residue between ORF21 and ORF20 in the *oriT* region but not affecting the nick site (TGGTGTGG), (2) the insertion of a cytosine in ORF15, resulting in a frameshift and a mutant protein of 725 amino acids and (3) a G > T conversion in ORF9 resulting in a Q > K amino acid substitution.

For BS34A (EMBL PRJEB7328), we identified 4233 predicted coding sequences within the 4.233.615 bp genome sequence (the numbers of RNAs was the same as for BS49). The variations of the Tn*916* element of BS34A (bps 1886552-1904 583) compared to the *E. faecalis* element were the same as for BS49, suggesting that they arose either within the pAM120 plasmid or during the transformation that generated BS49.

We additionally verified the locations of Tn*916* in the genome assembly using conventional PCRs for both BS34A and BS49 (text and Fig. S1, Supporting Information). The element was found to be inserted between BS34A_19270 and BS34A_19450 (*xynP*), BS49_15600/BS49_15780 (*ykuC*) and BS49_34490 (*yufK*)/BS49_34670 (*yufL*) in regions that are conserved in closely related *B. subtilis* strains. As described for *C. difficile* (Mullany *et al.*, 2012), the insertion sites are characterized by stretches of A and T (Fig. S2A, Supporting Information). Interestingly, a sequence flanking the Tn*916* element in BS34A shows homology to the *ykyB-ykuC* insertion site from BS49, and a sequence flanking Tn*916* at that locus resembles the *yufK-yufL* insertion site (Fig. S2B, Supporting Information), suggesting that additional (chromosomal) sequence was carried over when Tn*916* mobilized.

We compared the BS34A and BS49 genome sequences. Besides the Tn*916* insertion sites, we identified five variants, including two non-conservative substitutions in the histidine kinase WalK (Table 1).

A progressiveMAUVE alignment (Darling *et al.*, 2010) of the BS49 and 168 genome sequences showed a single large collinear block with >98% pairwise identity, demonstrating that strains BS49 and strain 168 are highly related. We then also aligned the Illumina sequence reads from BS34A and BS49 to the genome of strain 168 using Geneious R7. The resulting list of variants confirms the variants identified above that distinguish the two strains, and in addition shows other mutations that are in common and are likely present in the ancestral strain CU2189 (Christie *et al.*, 1987) (Table S1, Supporting Information). These include frameshifts in *ydcH*, *yesY*, *pksN* and *ywbD* [168 nomenclature]. Additionally, there are point mutations in for instance

comP, *sigI*, *sepF*, *ytpS/sftA* and a G > A mutation in *epsC* that was shown to result in partial restoration of biofilm formation in strain 168 (McLoon *et al.*, 2011). One or more of the mutations may underlie the difference between the auxotrophy of CU2189 (*metB5 hisA1 thr*-5) and strain 168 (*trpC2*).

SUPPLEMENTARY DATA

Supplementary data is available at FEMSLE online.

REFERENCES

Brouwer MS, Allan E, Mullany P, *et al*. Draft genome sequence of the nontoxigenic *Clostridium difficile* strain CD37. *J Bacteriol* 2012;**194**:2125–6.

Chain PS, Grafham DV, Fulton RS, *et al*. Genomics. Genome project standards in a new era of sequencing. *Science* 2009;**326**:236–7.

Chin CS, Alexander DH, Marks P, *et al*. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 2013;**10**:563–9.

Christie PJ, Korman RZ, Zahler SA, *et al*. Two conjugation systems associated with *Streptococcus faecalis* plasmid pCF10: identification of a conjugative transposon that transfers between *S. faecalis* and *Bacillus subtilis*. *J Bacteriol* 1987;**169**:2529–36.

Cookson AL, Noel S, Hussein H, *et al*. Transposition of Tn*916* in the four replicons of the *Butyrivibrio proteoclasticus* B316(T) genome. *FEMS Microbiol Lett* 2011;**316**:144–51.

Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 2010;**5**:e11147.

Francis MB, Allen CA, Shrestha R, *et al*. Bile acid recognition by the *Clostridium difficile* germinant receptor, CspC, is important for establishing infection. *PLoS Pathog* 2013;**9**:e1003356.

Gawron-Burke C, Clewell DB. Regeneration of insertionally inactivated streptococcal DNA fragments after excision of transposon Tn*916* in *Escherichia coli*: strategy for targeting and cloning of genes from gram-positive bacteria. *J Bacteriol* 1984;**159**:214–21.

Haraldsen JD, Sonenshein AL. Efficient sporulation in *Clostridium difficile* requires disruption of the *sigmaK* gene. *Mol Microbiol* 2003;**48**:811–21.

He M, Miyajima F, Roberts P, *et al*. Emergence and global spread of epidemic healthcare-associated *Clostridium difficile*. *Nat Genet* 2013;**45**:109–13.

Kathariou S, Stephens DS, Spellman P, *et al*. Transposition of Tn916 to different sites in the chromosome of *Neisseria meningitidis*: a genetic tool for meningococcal mutagenesis. *Mol Microbiol* 1990;**4**:729–35.

Kunst F, Ogasawara N, Moszer I, *et al*. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* 1997;**390**:249–56.

Lin WJ, Johnson EA. Transposon Tn*916* mutagenesis in *Clostridium botulinum*. *Appl Environ Microb* 1991;**57**:2946–50.

Mackin KE, Carter GP, Howarth P, *et al*. Spo0A differentially regulates toxin production in evolutionarily diverse strains of *Clostridium difficile*. *PLoS One* 2013;**8**:e79666.

McBride SM, Sonenshein AL. Identification of a genetic locus responsible for antimicrobial peptide resistance in *Clostridium difficile*. *Infect Immun* 2011;**79**:167–76.

McLoon AL, Guttenplan SB, Kearns DB, *et al*. Tracing the domestication of a biofilm-forming bacterium. *J Bacteriol* 2011;**193**:2027–34.

Mullany P, Williams R, Langridge GC, *et al*. Behavior and target site selection of conjugative transposon Tn*916* in two different strains of toxigenic *Clostridium difficile*. *Appl Environ Microb* 2012;**78**:2147–53.

Pettit LJ, Browne HP, Yu L, *et al*. Functional genomics reveals that *Clostridium difficile* Spo0A coordinates sporulation, virulence and metabolism. *BMC Genomics* 2014;**15**:160.

Roberts AP, Hennequin C, Elmore M, *et al*. Development of an integrative vector for the expression of antisense RNA in *Clostridium difficile*. *J Microbiol Meth* 2003;**55**:617–24.

Roberts AP, Mullany P. A modular master on the move: the Tn*916* family of mobile genetic elements. *Trends Microbiol* 2009;**17**:251–8.

Roberts AP, Mullany P. Tn*916*-like genetic elements: a diverse group of modular mobile elements conferring antibiotic resistance. *FEMS Microbiol Rev* 2011;**35**:856–71.

Rubens CE, Heggen LM. Tn*916* delta E: a Tn*916* transposon derivative expressing erythromycin resistance. *Plasmid* 1988;**20**:137–42.

Rutherford K, Parkhill J, Crook J, *et al*. Artemis: sequence visualization and annotation. *Bioinformatics* 2000;**16**:944–5.

Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;**30**:2068–9.

Whetzel PL, Hnatow LL, Keeler CL, Jr, *et al*. Transposon mutagenesis of *Mycoplasma gallisepticum*. *Plasmid* 2003;**49**:34–43.

Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 2008;**18**:821–9.