Infection, Genetics and Evolution 54 (2017) 287–298

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Research paper

First report of two complete *Clostridium chauvoei* genome sequences and detailed in silico genome analysis

Prasad Thomas [a], Torsten Semmler [c], Inga Eichhorn [b], Antina Lübke-Becker [b], Christiane Werckenthin [d], Mostafa Y. Abdel-Glil [a], Lothar H. Wieler [c], Heinrich Neubauer [a], Christian Seyboldt [a,*]

[a] Institute of Bacterial Infections and Zoonoses, Friedrich-Loeffler-Institut, Naumburger Str. 96A, 07743 Jena, Germany
[b] Institute of Microbiology and Epizootics, Department of Veterinary Medicine, Freie Universität, Robert-von-Ostertag-Str. 7-13, Building 35, 14163, Berlin, Germany
[c] Robert Koch Institute, Nordufer 20, 13353 Berlin, Germany
[d] LAVES, Lebensmittel- und Veterinärinstitut Oldenburg, Martin-Niemöller-Straße 2, 26133 Oldenburg, Germany

ARTICLE INFO

Article history:
Received 22 March 2017
Received in revised form 12 July 2017
Accepted 13 July 2017
Available online 15 July 2017

Keywords:
Clostridium chauvoei
Complete circular genomes
Spo0A
CRISPR
fliC

ABSTRACT

Clostridium (*C.*) *chauvoei* is a Gram-positive, spore forming, anaerobic bacterium. It causes black leg in ruminants, a typically fatal histotoxic myonecrosis. High quality circular genome sequences were generated for the *C. chauvoei* type strain DSM 7528^T (ATCC 10092^T) and a field strain 12S0467 isolated in Germany. The origin of replication (*oriC*) was comparable to that of *Bacillus subtilis* in structure with two regions containing DnaA boxes. Similar prophages were identified in the genomes of both *C. chauvoei* strains which also harbored hemolysin and bacterial spore formation genes. A CRISPR type I-B system with limited variations in the repeat number was identified. Sporulation and germination process related genes were homologous to that of the *Clostridia* cluster I group but novel variations for regulatory genes were identified indicative for strain specific control of regulatory events. Phylogenomics showed a higher relatedness to *C. septicum* than to other so far sequenced genomes of species belonging to the genus *Clostridium*. Comparative genome analysis of three *C. chauvoei* circular genome sequences revealed the presence of few inversions and translocations in locally collinear blocks (LCBs). The species genome also shows a large number of genes involved in proteolysis, genes for glycosyl hydrolases and metal iron transportation genes which are presumably involved in virulence and survival in the host. Three conserved flagellar genes (*fliC*) were identified in each of the circular genomes. In conclusion this is the first comparative analysis of circular genomes for the species *C. chauvoei*, enabling insights into genome composition and virulence factor variation.

1. Introduction

Blackleg (black quarter, quarter evil, Rauschbrand, charbon symptomatique) is an economically important disease with high mortality that affects cattle, sheep and other animals. The disease is caused by *Clostridium* (*C.*) *chauvoei*, a Gram-positive, motile, spore producing anaerobic bacterium. Blackleg occurs worldwide, although mainly observed in ruminants, sporadic human cases have been reported (Nagano et al., 2008; Weatherhead and Tweardy, 2012). The typical infection in ruminants is characterized by myonecrosis of striated and cardiac muscle and often peracute death. A unique feature of the disease in cattle is the appearance as a non-traumatic endogenous infection (Hatheway, 1990). The presumed oral way of infection is by uptake of spores with feed from contaminated pastures. However, only little is known about the molecular mechanisms of pathogenicity and the role of various toxins and other virulence factors in the pathogenesis of the disease. The better characterized virulence factors include flagella, NanA sialidase (NanA) and the *Clostridium chauvoei* toxin A (CctA) (Frey and Falquet, 2015; Frey et al., 2012; Tamura et al., 1995; Vilei et al., 2011). A draft genome sequence for the pathogen was available for a virulent Swiss isolate (JF4335) consisting of 12 contigs with a 2.8-Mb genome and 5-kb plasmid (Falquet et al., 2013) and was recently (March 2017) resequenced and submitted as a 2.8 Mb complete genome. The genome sequence analysis of JF4335 strain identified prophage elements, several primary virulence factors, proteases and antibiotic resistance genes (Frey and Falquet, 2015). 16S rDNA-based phylogenetic analysis of the genus *Clostridium* has revealed 19

Abbreviations: PacBio, Pacific Biosciences; CRISPR, Clustered Regularly Interspaced Short Palindromic Repeats; LCBs, locally collinear blocks; *fliC*, flagellin type C.

* Corresponding author at: Friedrich-Loeffler-Institut, Federal Research Institute for Animal Health, Institute of Bacterial Infections and Zoonoses, Naumburger Str. 96a, 07743 Jena, Germany.
E-mail addresses: prasadthomas99@gmail.com (P. Thomas), semmlert@rki.de (T. Semmler), inga.eichhorn@fu-berlin.de (I. Eichhorn), antina.luebke-becker@fu-berlin.de (A. Lübke-Becker), christiane.werckenthin@laves.niedersachsen.de (C. Werckenthin), mostafa.abdel-glil@fli.de (M.Y. Abdel-Glil), wielerlh@rki.de (L.H. Wieler), heinrich.neubauer@fli.de (H. Neubauer), christian.seyboldt@fli.de (C. Seyboldt).

http://dx.doi.org/10.1016/j.meegid.2017.07.018

phylogenetic clusters with cluster I forming the core group of the genus (Collins et al., 1994). More than half of the pathogenic species are members of this cluster including those considered the major pathogenic agents, i.e. *C. botulinum*, *C. haemolyticum*, *C. novyi*, *C. perfringens*, *C. tetani*, *C. septicum* and *C. chauvoei* (Stackebrandt et al., 1999). The phylogenetic positioning of *C. chauvoei* based on 16 s rDNA sequence shows this pathogen close to *C. septicum* (Kuhnert et al., 1996).

Genome sequencing studies have identified remarkable genetic variability among isolates within the same species from the genus *Clostridia* cluster I, as reported for *C. botulinum* and *C. perfringens* (Carter and Peck, 2015; Fang et al., 2010; Myers et al., 2006). So far, there is only limited information regarding the genetic diversity of *C. chauvoei*. Presence of tandem copies of flagellin type C genes (*fliC*) with a central variable region (Sasaki et al., 2002) and nucleotide polymorphisms for *cctA* and *nanA* (Frey et al., 2012; Vilei et al., 2011) have been reported.

Genome sequencing is a powerful tool for understanding a bacterial species in its diversity which is often caused by various types of mobile genetic elements (Binnewies et al., 2006). The Pacific Biosciences (PacBio) developed Single Molecule Real Time (SMRT) sequencing technology which generates reads with an average read length of above 10-kb and thus can also resolve long repeat regions (Eid et al., 2009; Liao et al., 2015). Hierarchical Genome Assembly Process (HGAP), a recently developed non-hybrid assembly process when applied to SMRT DNA sequencing was found suitable for finishing bacterial genomes with more than 99.999% accuracy (Chin et al., 2013; Liao et al., 2015).

Considering the disease significance and the importance of genome sequence data to understand the pathogens life cycle, the complete genome sequences of two *C. chauvoei* strains, 12S0467 (a virulent field isolate from Germany) and DSM 7528^T (ATCC 10092^T) (Type strain of the species) were unraveled. Comparative genomics and genome analysis tools were applied to provide information regarding the genome assembly and annotation, origin of replication, collinear blocks, phylogenetic relatedness, orthologous genes, subsystem category distribution features, CRISPR and phage elements, sporulation and germination, virulence factors and antibiotic resistance.

2. Material and methods

2.1. Bacterial strains and DNA extraction

Bacterial strains used in the present study are a virulent strain obtained from a blackleg case in cattle from northern Germany in 2011 (12S0467) and the type strain (DSM 7528^T (ATCC 10092^T). For DNA extraction, isolates were cultured in 20 ml Selzer broth (Selzer et al., 1996) (tryptone - 30 g, beef extract - 20 g, glucose - 4 g, L-cysteine hydrochloride - 1 g in 1000 ml H_2O, pH 7.2) at 37 °C for 18 to 24 h under anaerobic conditions. Genomic DNA was extracted using Qiagen Genomic-tip 100/Q and genomic DNA buffer set (Qiagen, Germany) with minor modification as 40 U of achromopeptidase (Frey et al., 2012) was included in the enzymatic lysis buffer and the DNA was incubated at 37 °C until dissolved. DNA quality was examined by using Qubit 2.0 fluorometer (Life Technologies, Germany) and by agarose gel electrophoresis and species confirmation was made with PCR (Sasaki et al., 2000).

2.2. Genome sequencing and assembly

Genome sequencing was carried out by SMRT DNA sequencing (Eid et al., 2009) using PacBio RSII sequencer at GATC Biotech (Germany). Genome assembly was carried out using HGAP algorithm version 3 (Chin et al., 2013) implemented in PacBio SMRT portal version 2.3.0 at GATC Biotech (Germany). The overlapping regions of circular sequences were determined using Gepard software (Krumsiek et al., 2007). Circularization of genome and plasmid sequence represented by a single contig and merging of contigs were carried out using Geneious 9.0.5 (Kearse et al., 2012), Minimus2 (Sommer et al., 2007) and Circlator (Hunt et al., 2015), visualization was carried out using ACT (Carver et al., 2005). The circular contigs were polished with RS_Resequencing.1 protocol in SMRT portal v2.3.0.

2.3. Genome annotation

Origin of replication (*oriC*) was identified using Ori-Finder (Gao and Zhang, 2008). Protein-coding sequences were predicted by Glimmer software version 3.0 (Delcher et al., 2007), Ribosomal RNA genes and transfer RNA genes were detected using RNAmmer software version 1.2 and using tRNAscan-SE respectively (Lagesen et al., 2007; Lowe and Eddy, 1997). CRISPR loci were searched using the CRISPR Recognition tool (Bland et al., 2007). Prophage elements were identified using PHAST (Zhou et al., 2011). Genome Annotation was performed using the Rapid Annotation using Subsystem Technology (RAST) server (Aziz et al., 2008; Overbeek et al., 2014), Prokka 1.11 (Seemann, 2014) and NCBI's Prokaryotic Genome Annotation Pipeline (PGAP). Genome annotation based on RAST was also carried out for the recently published genome sequence of the *C. septicum* strain CSUR P1044 isolated from human gut (Benamar et al., 2016). Genome annotations were additionally done with Gene Ontology (GO) terms using Blast2GO PRO software in CLC genomics workbench 8.5.1 and the results were summarized in General GO slim functional categories (Conesa and Gotz, 2008; Conesa et al., 2005; Gotz et al., 2011; Götz et al., 2008). BLASTP search was carried out against the non-redundant database and the GO terms associated with each BLAST hit were used for annotation. InterPro annotations were also used to retrieve domain/motif information for the proteins. Antibiotic resistance genes were identified using

Table 1
Details of published sequence data used for this study showing species, strain designation, type of data used and accession number.

Assembly of genomes retrieved from NCBI[b]			
Species	Strain designation	Type of data	NCBI accession number
C.[a] *chauvoei*	JF4335	Assembly	NZ_LT799839.1
C. septicum	CSUR P1044	Assembly	PRJEB146921
C. botulinum C/D	BKT2873	Assembly	PRJNA233460
C. botulinum C/D	BKT12695	Assembly	PRJNA233471
C. botulinum C	str.Eklund	Assembly	PRJNA20017
C. botulinum A	Hall	Assembly	NC_009698
C. botulinum E3	str. Alaska E43	Assembly	NC_010723
C. botulinum E1	str. BoNT E Beluga	Assembly	PRJNA29861
C. botulinum B	str. Eklund 17B	Assembly	NC_010674/NC_010680
C. botulinum A	ATCC 3502	Assembly	NC_009495/NC_009496
C. acetobutylicum	DSM 1731	Assembly	NC_015687
C. acetobutylicum	ATCC 824	Assembly	NC_003030
C. acetobutylicum	EA 2018	Assembly	NC_017295
C. tetani	E88	Assembly	NC_004557/NC_004565
C. butyricum	KNU-L09	Assembly	NZ_CP013252/NZ_CP013489
C. butyricum	5521	Assembly	NZ_ABDT00000000
C. butyricum	E4 BoNT E BL5262	Assembly	NZ_ACOM00000000
C. haemolyticum	NCTC 8350	Assembly	NZ_JDSA00000000
C. novyi NT	NT	Assembly	NC_8593
C. novyi B	str. ATCC	Assembly	PRJNA233467
C. sporogenes	DSM 795	Assembly	NZ_CP011663
C. beijerinckii	ATCC 35702	Assembly	NZ_CP006777
C. beijerinckii	NCIMB 8052	Assembly	NC_009617
C. perfringens	ATCC 13124	Assembly	NC_008261
C. perfringens	str. 13	Assembly	NC_003366/NC_003042
C. baratii	str. Sullivan	Assembly	NZ_CP006905/ NZ_CP006906

[a] *Clostridium*.
[b] National Centre for Biotechnology Information.

Resistance Gene Identifier 3.0.9 (RGI) with the Comprehensive Antibiotic Research Database (CARD) (McArthur et al., 2013).

2.4. Comparative genome analysis and visualization

The circular plot of genomes was made using Artemis and DNAPlotter (Carver et al., 2009; Rutherford et al., 2000) and locally collinear blocks (LCBs) were plotted using progressiveMauve (Darling et al., 2010).

Gegenees software_v2.2.1 was utilized for carrying out phylogenomic analysis of the genus *Clostridia* (Ågren et al., 2012). TBLASTX with settings of 500/500 were used and heat maps were generated with a cut-off threshold of 35% for non-conserved genetic material. Phylogenetic tree (Neighbour Joining method) was created in Splits Tree 4 (Huson and Bryant, 2006) using the PHYLIP file exported from Gegenees software. The phylogenomic study involved various species within the genus such as *C. chauvoei*, *C. septicum, C. botulinum* type C/D, *C. botulinum* type C, *C. botulinum* type A, *C. botulinum* type E3, *C. botulinum* type E1, *C. botulinum type* B, *C. acetobutylicum*, *C. tetani*, *C. butyricum*, *C. haemolyticum, C. novyi* NT, *C. novyi* B, *C. sporogenes, C. beijerinckii*, *C. perfringens* and *C. baratii*. Species, strain designation and accession number of the genomes used for comparative genomic studies are shown in Table 1.

The orthologous gene identification (PGAP annotation) for the three completed *C. chauvoei* strains and also the *C. septicum* CSURP1044 strain was carried out using Pan-genome ortholog clustering tool 3.23 (panOCT) with default parameters (Fouts et al., 2012). The venn diagram depicting the core, shared and unique genes identified using panOCT were generated using jvenn (Bardou et al., 2014).

3. Results

3.1. Genome assembly and annotation

HGAP 3 assembly generated one contig representing the genome for the DSM 7528^T strain. The final circular genome of 2872,664 bp was created after circularization and final polishing with the RS_Resequencing protocol. Two contigs representing the chromosome and one contig for the plasmid was obtained for strain 12S0467. The merging of the 2 contigs representing the genome was carried out using Minimus2, but no circularization was achieved and hence was carried out from the overlapping ends using Geneious 9.0.5 (Kearse et al., 2012). The contig representing the plasmid generated by HGAP was showing multiple copies of the plasmid sequence based on an assembly generated in Circlator and was visualized in Artemis Comparison Tool (ACT) (Fig. 1). The final genome of strain 12S0467 was represented by a 2,885,628 bp chromosome and a 3941 bp plasmid respectively after polishing with RS_Resequencing protocol. Genome assembly and circularization summary for the complete genomes are depicted in Table 2.

Glimmer 3 predicted 2678 and 2636 open reading frames for the DSM 7528^T and strain 12S0467 respectively. RNAmmer1.2 and tRNAscan-SE predicted 27 ribosomal RNAs and 87 transfer RNAs respectively for both genomes. Annotation of the Glimmer 3 predicted proteomes were carried out using Blast2GO which revealed various functional elements for the species. The genome summary table (Table 3) is showing chromosome, plasmid and genome sizes, GC content; the number of rRNA operons, tRNA and coding sequences of the three complete *C. chauvoei* strains (Table 3).

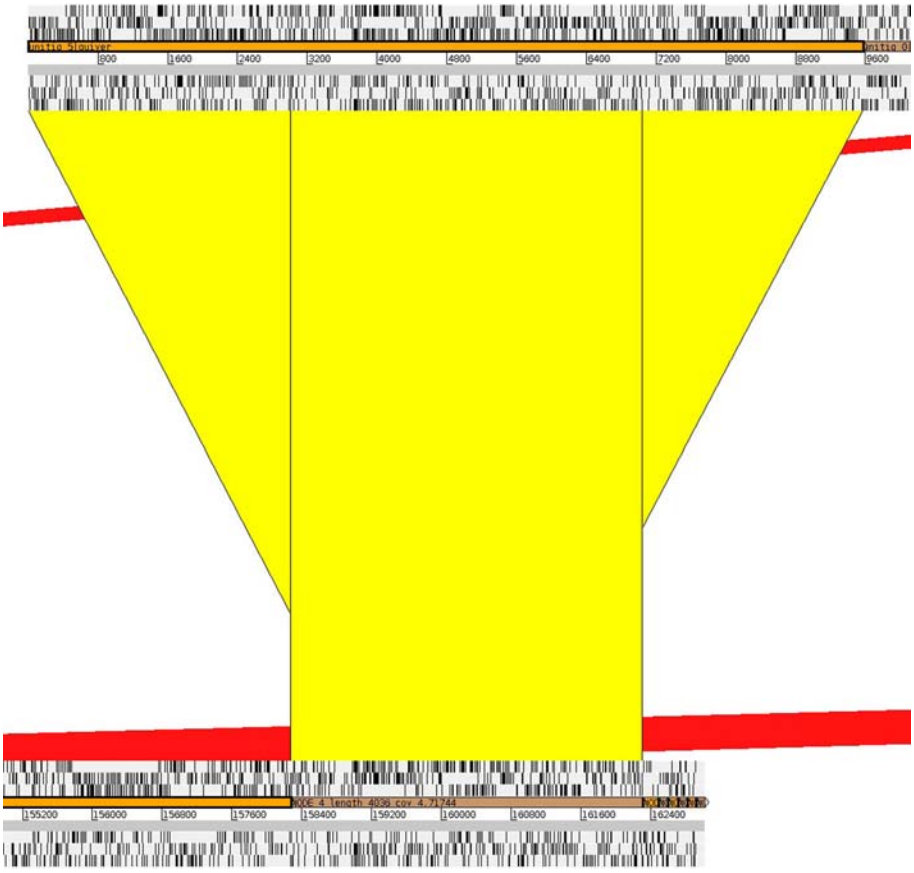


Fig. 1. ACT view showing multiple copies of a plasmid sequence Contig unitig_5 (shown above) generated by HGAP represents multiple copies of the plasmid sequence. Circlator reduced the multiplicity to generate contig NODE 4 (shown below) which was used to circularize and determine the 4 kb plasmid sequence.

Table 2
Genome assembly and circularization summary (*Clostridium chauvoei* strains DSM 7528^T and 12S0467).
Contig/genome size, coverage and consensus accuracy of the genome assembly (HGAP3) and circularization summary of the strains DSM 7528^T and 12S0467. Increased coverage and consensus accuracy in RS_resequencing (error correction) indicate correct circularization of the genomes.

Clostridium chauvoei strain DSM 7528^T				
	Contig	Length (bp)	Coverage	Consensus accuracy
HGAP summary				
1	unitig_0	2,888,463	250.2×	99.98%
Circularization summary				
1	Genome	2,872,664	270×	100%

Clostridium chauvoei strain 12S0467				
	Contigs	Length (bp)	Coverage	
HGAP summary				
1	unitig_0	2,735,737	377×	99.98%
2	unitig_1	169,089	394×	
3	unitig_5	9570	596×	
Circularization summary				
1	Genome	2885,630	417×	100%
2	Plasmid	3941	1218×	

3.2. Origin of replication

Orifinder predicted a short and long *oriC* region of 257 bp and 734 bp at either side of *dna*A displaying five and eight DnaA box sequence motives (ttatccaca) with not more than one mismatch to *Escherichia coli* DnaA box, respectively. Marker genes commonly observed near the bacterial origin of replication were observed near the *oriC* region. Two probable DNA-unwinding elements (DUE) sites were identified within the shorter *oriC* region based on its higher A/T composition (Fig. 2).

3.3. Comparative genome analysis and visualization

The circular genome plot with RAST predicted proteins and other elements for the strain DSM 7528^T was created using DNAPlotter software (Fig. 3). Collinear blocks with inversion and insertions of blocks were identified among the *C. chauvoei* genomes (DSM 7528^T, 12S0467 and JF4335) and also among *C. chauvoei* and *C. septicum* genomes (DSM 7528^T and CSUR P1044) (Fig. 4).

The analysis of genome sequence data to depict phylogenetic relatedness in the genus Clostridium revealed groups of species sharing homology. *C. chauvoei* was showing high relatedness with *C. septicum* (74%) in phylogenomic analysis within the genus *Clostridium* (Fig. 5).

Comparative genome analysis carried out to detect orthologous genes (Fouts et al., 2012) among the three *C. chauvoei* genomes identified the number of 2548 core genes and 38 accessory genes (Fig. 6A). The number of orthologous genes (Fouts et al., 2012) shared by the three *C. chauvoei* genomes (DSM 7528^T, 12S0467 and JF4335) and the *C. septicum* (CSUR P1044) genome was 1520 genes (Fig. 6B). The number of species-specific unique genes among the 4123 gene pan-genome identified in both species together was 1029 genes for *C. chauvoei* and 1538 genes for *C. septicum*, respectively. Orthologous protein clusters shared by all three *C. chauvoei* genomes (DSM 7528^T, 12S0467 and JF4335) and with *C. septicum* strain CSUR P1044 is represented in venn diagrams (Fig. 6).

The genome features of *C. chauvoei* (DSM 7528^T) and *C. septicum* (CSUR P1044) were compared based on subsystem category distribution features in RAST (Aziz et al., 2008; Overbeek et al., 2014). Both genomes share similar subsystem category distributions whereas the subsystem feature counts were higher for *C. septicum* for genes related to prophages, phosphorus metabolism and carbohydrates (Fig. 7).

3.4. CRISPR and phage elements

Two CRISPR elements were identified in the genomes with site lengths of around 0.6-kb (site 1) and 2-kb (site 2) separated by a distance of around 1.5-kb. There were 11 repeats for site 1 in both genomes, whereas 32 and 34 repeats were observed for site 2 in 12S0467 and DSM 7528^T, respectively. Direct repeat (DR) sequences were of the same sequence for both isolates at both sites (GATTAACATTAACATGAGATGTATTTAAAT). The spacers between the CRISPR repeat elements at site 2 were variable with 31 and 33 numbers for 12S0467 and DSM 7528^T, respectively, with a length ranging from 35 to 37 bp. Both CRISPR elements were accompanied by the same genes for CRISPR associated proteins (Cas) flanking the site. The direct repeat number for the Swiss strain (JF4335) was 11 and 34, similar to that of the DSM 7528^T strain. Phage analysis carried out using PHAST server identified intact prophages of 53-kb and 46.3-kb for DSM 7528^T and 12S0467, respectively, and an incomplete prophage of 10.2-kb for both strains with similar genetic structure. The incomplete phage element showed the presence of a small number of bacterial specific protein genes.

3.5. Spore resistance, sporulation and germination

The characterized genome has around 70 predicted genes for spore resistance, sporulation and germination which include 5 small, acid-soluble spore proteins (SASPs) involved in spore resistance, set of genes involved in dipicolinic acid (DPA) synthesis was similar to that identified in the *Clostridia* cluster I group. Proteins involved in the various stages of

Table 3
Genome summary (*Clostridium chauvoei* strains DSM 7528^T, 12S0467 and JF4335)
Genome summary table showing chromosome, plasmid, genome sizes and their respective GC content obtained. The number of rRNA operons, tRNA and coding sequences for all the three strains is also compared.

Genome (kb)		DSM 7528^T		12S0467		JF4335 (NZ_LT799839.1)	
		Chromosome	2872	Chromosome	2885	Chromosome	2887
		Plasmid	–	Plasmid	3,0.9	Plasmid	–
		Genome	2872	Genome	2889	Genome	2887
GC content		Genome	28.30%	Genome	28.30%	Genome	28.30%
		Plasmid	–	Plasmid	27%	Plasmid	–
rRNA		9 operons		9 operons		9 operons	
tRNA		87		87		87	
CDS	RAST	2676		2632		–	
	Prokka	2662		2626			
	PGAP	2582		2577		2569	

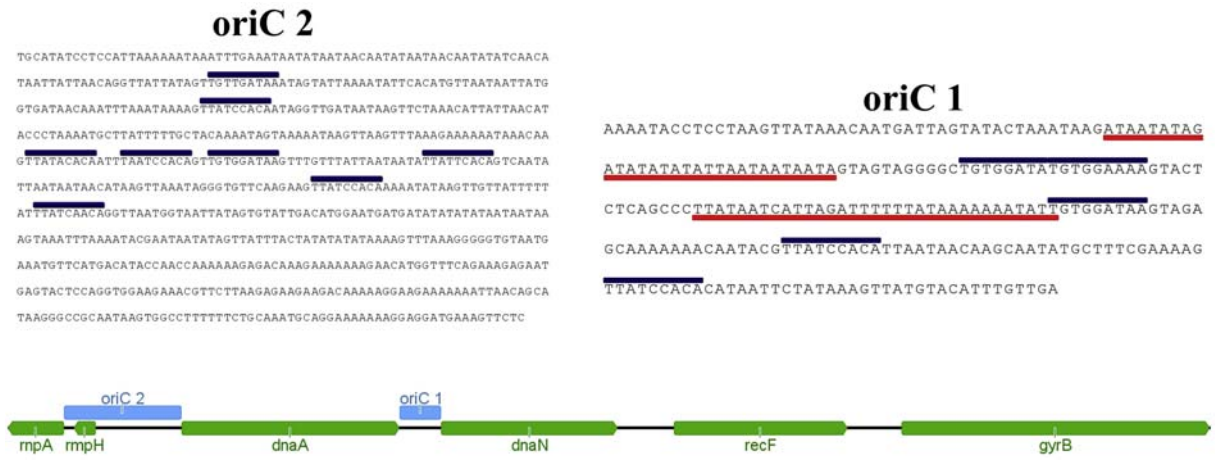


Fig. 2. Schematic diagram of the origin of replication (*oriC*). The two *oriC* regions identified by orifinder are depicted in the lower part of the figure; sky blue boxes represent *oriC1* and *oriC2* with the presence of marker genes (*rnpA–rmpH–dnaA–dnaN–recF–gyrB*) near to the origin of replication. The sequence of two *oriC* regions (*oriC1* and *oriC2*) with DnaA boxes (dark blue) and two probable DNA-unwinding elements (DUE) (red) is shown. The schematic representation was created using Geneious 9.0.5 (http://www.geneious.com/).

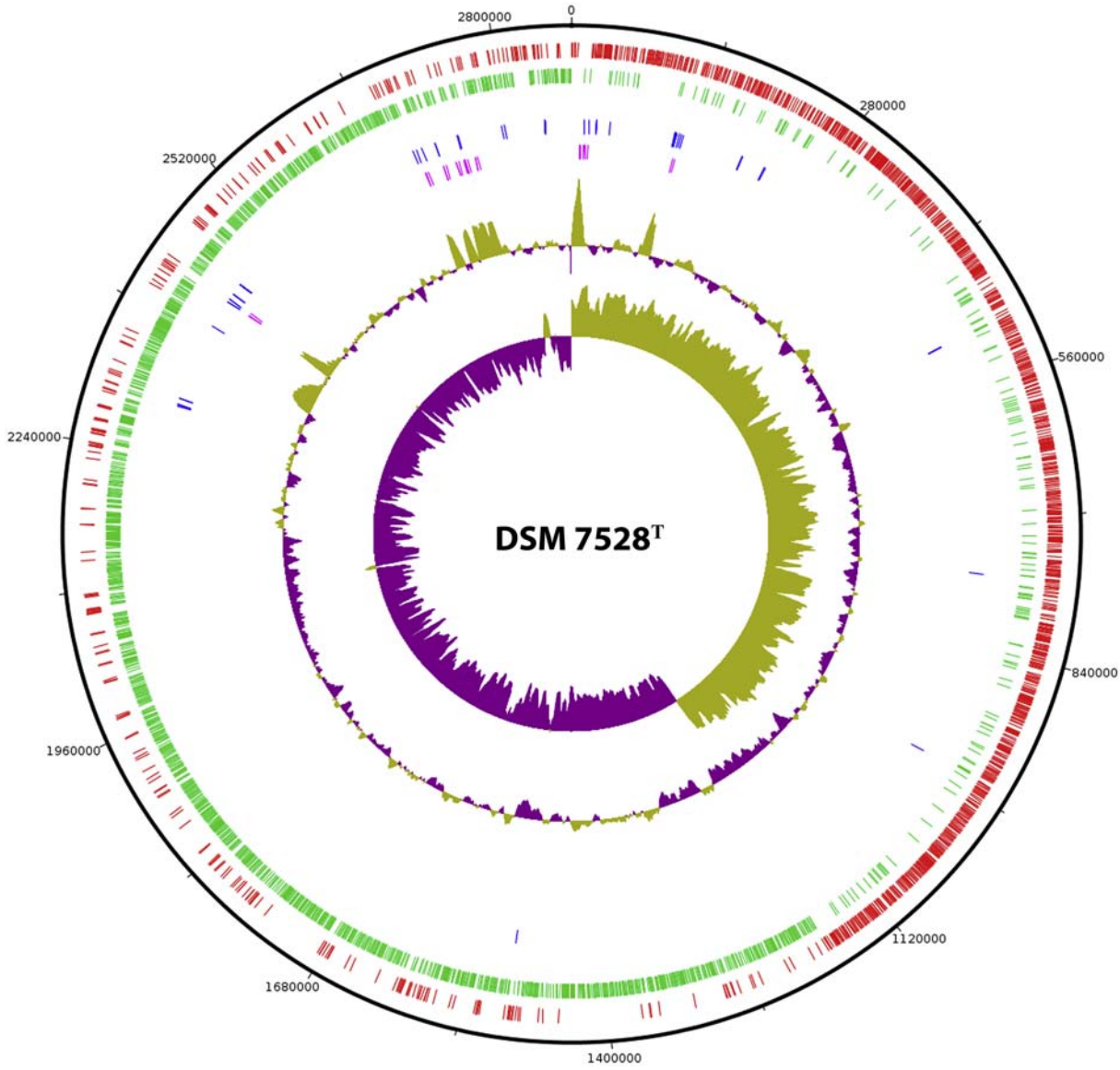
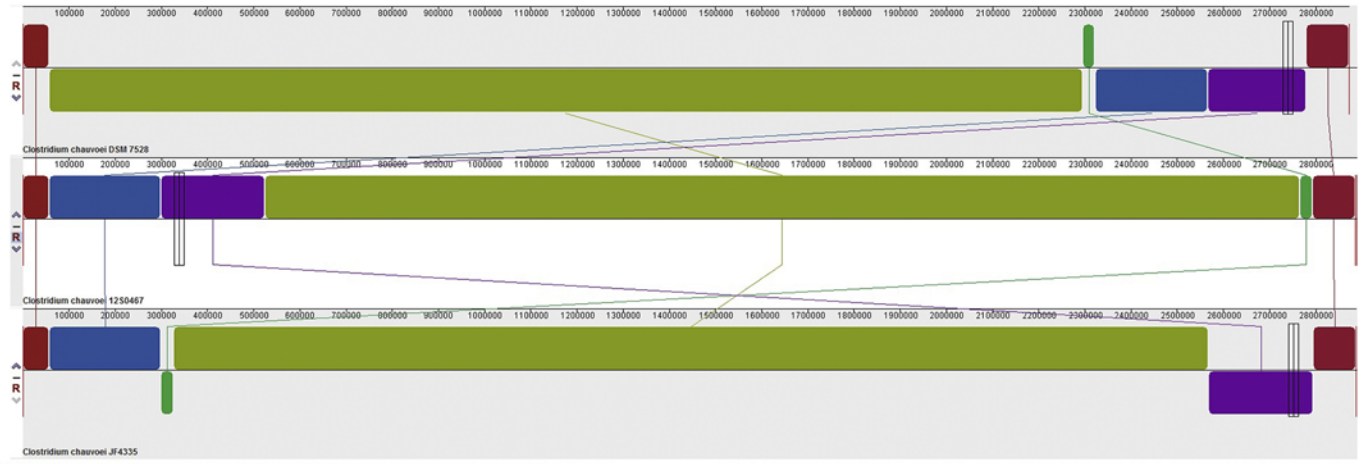
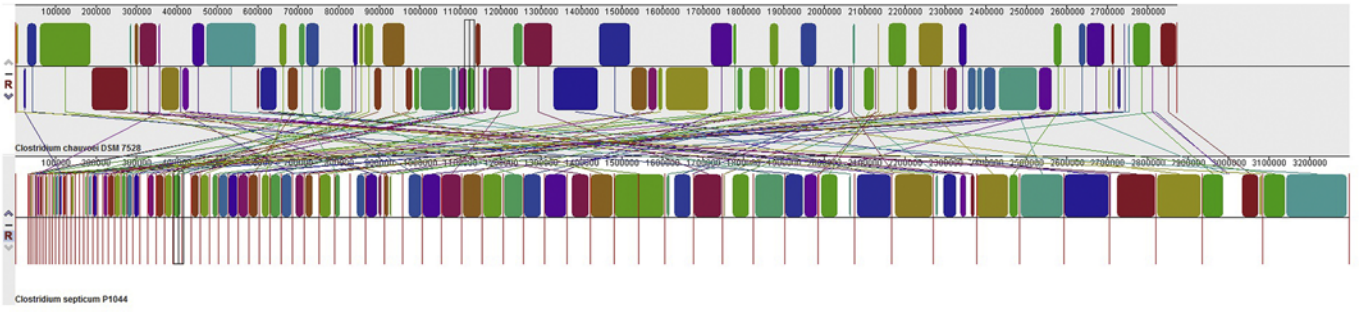


Fig. 3. Circular plot of the genome of strain DSM 7528^T. The plot was made using DNAPlotter. Red and green indicate forward and reverse genes predicted by RAST. Blue and pink lines correspond to tRNA and rRNA genes respectively. The inner circle displays the GC skew and the second circle from the center displays the G + C content.



A: DSM 7528^T (upper), 12S0467 (middle) and JF4335 (lower)



B: DSM 7528^T (above), CSUR P1044 (below)

Fig. 4. A: DSM 7528^T (upper), 12S0467 (middle) and JF4335 (lower). B: DSM 7528^T (above), CSUR P1044 (below.)

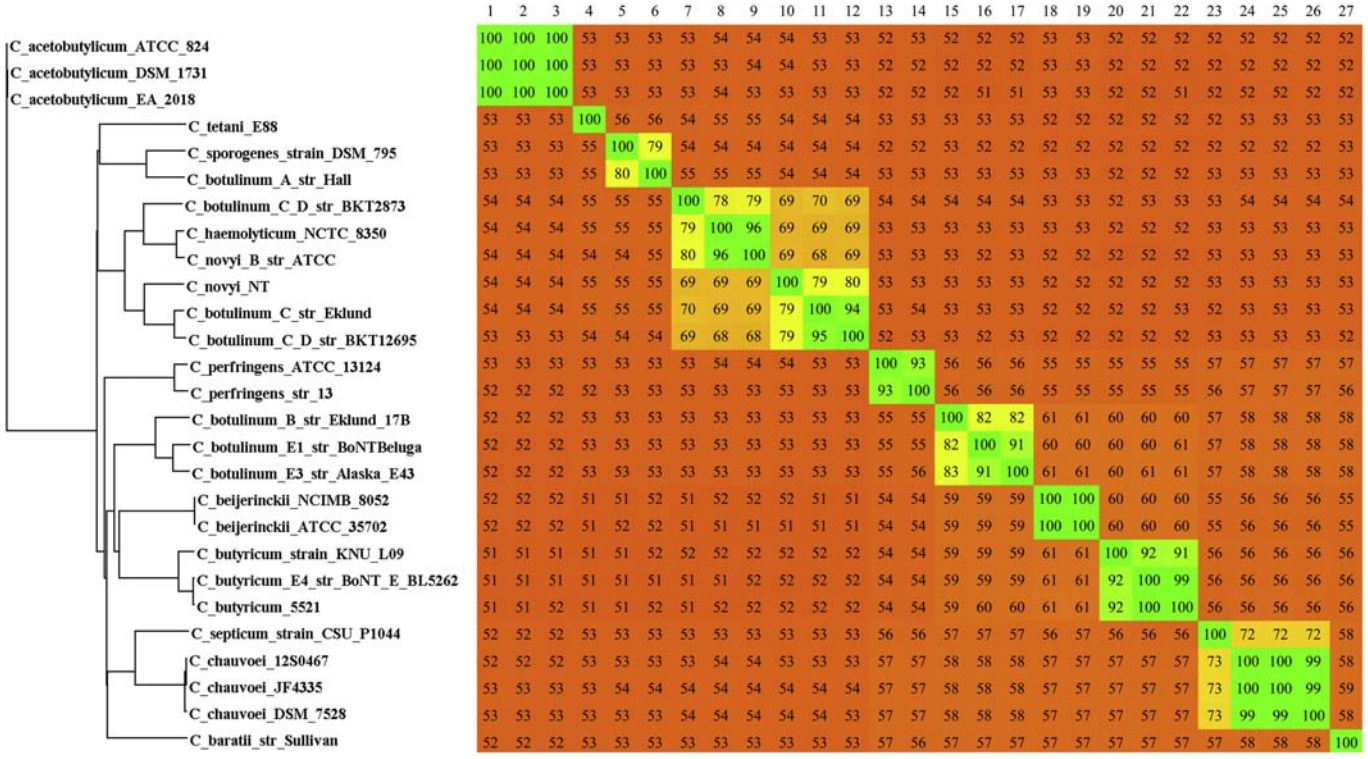


Fig. 5. Phylogenomic relatedness within genus *Clostridium* TBLASTX with fragment alignment settings of 500/500 were used for comparative genomics and heat maps were generated with a cut-off threshold of 35% for non-conserved genetic material. A neighbour joining method based phylogenetic tree was created in Splits Tree 4 using the PHYLIP file exported from Gegenees software.

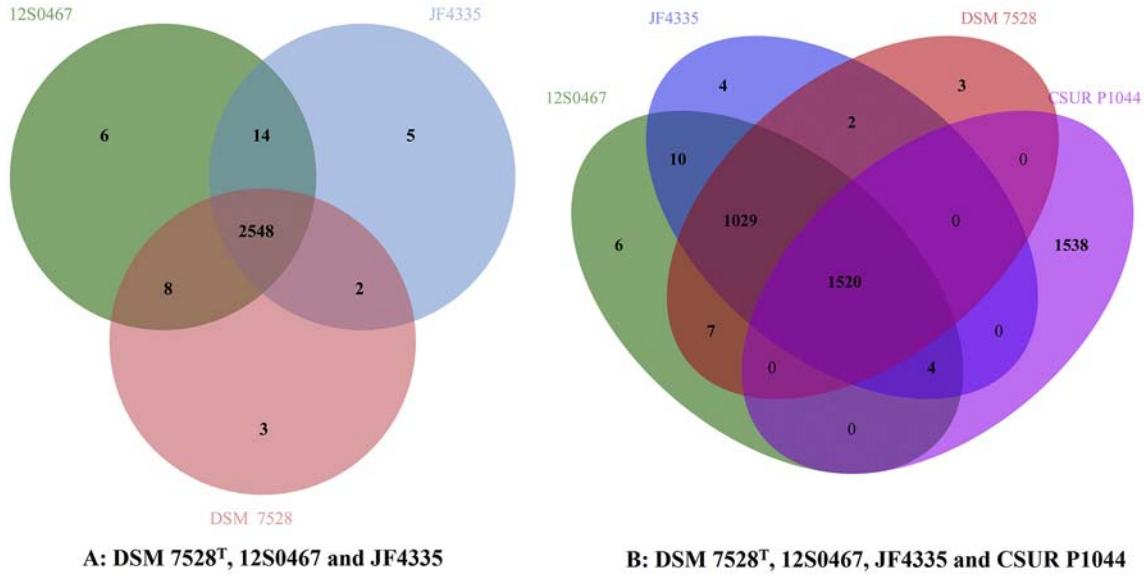


Fig. 6. A: DSM 7528^T, 12S0467 and JF4335. B: DSM 7528^T, 12S0467, JF4335 and CSUR P1044.

sporulation could also be identified including the major regulator Spo0A and key sigma factors regulating spore formation. The genome also harbors the *gerK* operon, identified as the Germinant Receptor (GR) and the lytic enzymes involved in the germination processes. The genetic relatedness of important genes regulating sporulation and germination of *C. chauvoei* was compared to that of *C. septicum* (CSUR P1044), *C. perfringens* (ATCC 13124) and *C. botulinum* type A (ATCC 3502) based on BLASTP (Altschul et al., 1997) searches with the respective proteome. The gene list used for comparison, length of coding sequences, pairwise similarity and coverage of query genes are given in Table 4.

3.6. Virulence factors and antibiotic resistance

Antibiotic resistance genes were predicted using the CARD database which identified eight genes conferring resistance to antibiotics including aminocoumarin (two genes), beta-lactam antibiotics, rifampicin, tetracycline, macrolide, peptide antibiotic and a gene conferring antibiotic resistance via molecular bypass for both *C. chauvoei* (DSM 7528^T and 12S0467) genomes. The genomes also showed the presence of three *fliC* genes with a coding length of 414 amino acids. Both genomes contain a gene coding for CctA, NanA sialidase, haemolysin III, haemolysin XhlA, internalin A2, hyaluronidase J (NagJ) and hyaluronidase H (NagH). The Blast2GO annotation identified the major functional gene classes found in the genome which includes several glycosidases and various proteins involved in metal ion binding and transportation. The genome also harbors several classes of proteolytic enzymes.

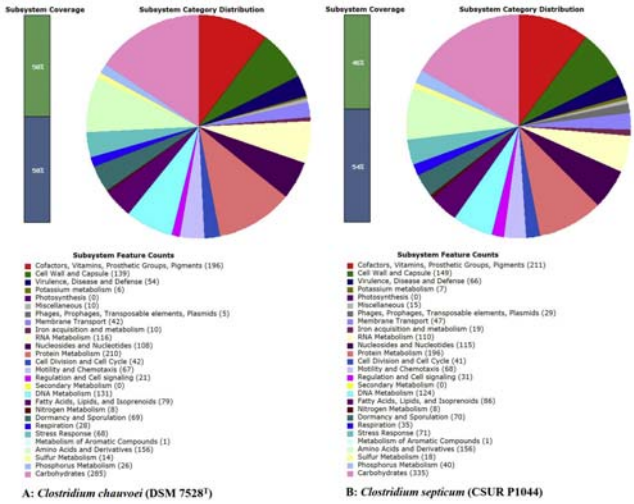


Fig. 7. A: *Clostridium chauvoei* (DSM 7528^T). *Clostridium septicum* (CSUR P1044)

4. Discussion

4.1. Genome assembly and annotation

Complete genome sequence of two *Clostridium chauvoei* strains was achieved for the type strain and an isolate of German origin which had been isolated from a diseased animal (Table 2). The strategy of microbial genome sequencing with PacBio long reads and HGAP successfully finished both genomes with high coverage and quality values without the need of manual completion or additional short read data. Earlier studies have shown the advantages of using PacBio generated long reads based on single-library and non-hybrid assemblies for completing bacterial genomes (Brown et al., 2014; Liao et al., 2015). The genomic sizes of both, the type strain and the field isolate corroborated the published genome sequence of the Swiss *C. chauvoei* field strain JF4335 (Table 3). No plasmid representing contig was obtained for DSM 7528^T but a 3.9-kb plasmid was identified in strain 12S0467. Like in other bacterial species the rRNA gene cluster is used as an important genetic marker to differentiate and identify *C. chauvoei* and *C. septicum* by PCR (Sasaki et al., 2000). Commonly, most of the rRNA genes are positioned close to each other, a constellation which is considered as the major hindrance for genome finishing (Koren et al., 2013). The characterization of the region seems to be problematic without long read based sequencing techniques such as the PacBio RS II system capable of generating reads with N50 close to 20 kb (Rhoads and Au, 2015). Using this technique we were able to resolve these difficulties, our genome analysis showed the presence of 87 tRNA genes and 9 rRNA gene clusters (Table 3).

4.2. Origin of replication

Orifinder identified two possible *oriC* regions in close vicinity on either side of *dnaA*, as well as near to other marker genes (Fig. 2) in the order *rnpA–rmpH–dnaA–dnaN–recF–gyrB*, the conserved gene order for a bacterial origin of replication (Mackiewicz et al., 2004). The predicted regions were similar to the origin of replication known for

Table 4
Genetic relatedness of important genes involved in spore resistance, sporulation and germination: The genetic relatedness of important genes involved in spore resistance, sporulation and germination of *C. chauvoei* (DSM 7528^T) was compared to *C. septicum* (CSUR P1044), *C. perfringens* (ATCC 13124) and *C. botulinum* type A (ATCC 3502) proteomes based on BLASTP searches. The percentage pair wise identity and percentage query coverage of the coding sequence was computed for assessing the genetic relatedness. Most of the *C. chauvoei* genes for sporulation and germination were showing greater homology with that of *C. septicum*.

Gene description	Product	Length (CDS)	% Pair wise identity	Query coverage	% Pair wise identity	Query coverage	% Pair wise identity	Query coverage
			C. septicum		*C. perfringens*		*C. botulinum*	
4-Hydroxy-tetrahydrodipicolinate synthase	DapA	293	81.20%	100%	65.20%	98.29%	72.90%	100.00%
Electron transfer flavoprotein, alpha subunit	EtfA CDS 1	397	92.20%	100%	76.10%	99.24%	72.50%	99.75%
Electron transfer flavoprotein, alpha subunit	EtfA CDS 2	336	96.50%	100%	75.70%	99.40%	64.20%	98.81%
Spore germination protein GerKA	GerKA	478	89.70%	100%	62.40%	99.58%	31.80%	95.60%
Spore germination protein GerKC	GerKC	373	85.80%	100%	48.40%	100.00%	23.40%	85.75%
Serine/threonine protein kinase PrkC	PrKC	661	77.30%	99.24%	50.60%	96.22%	45.70%	86.54%
RNA polymerase sporulation specific sigma factor SigE	SigE	236	97.40%	100%	83.40%	100.00%	83.80%	100.00%
RNA polymerase sporulation specific sigma factor SigF	SigF	252	98%	100%	72.00%	98.01%	70.50%	96.02%
RNA polymerase sporulation specific sigma factor SigG	SigG	258	99.20%	100%	89.50%	100.00%	88.30%	100.00%
RNA polymerase sporulation specific sigma factor SigH	SigH	215	93.00%	100%	76.00%	93.46%	81.10%	88.79%
RNA polymerase sporulation specific sigma factor SigK	SigK	232	97.80%	100%	70.10%	98.70%	68.10%	99.13%
Spore cortex-lytic enzyme precursor	SleB	195	79.10%	100%	61.20%	69.07%	59.30%	69.07%
Spore cortex-lytic enzyme SleC	SleC	445	91.00%	100%	60.50%	98.65%	40.80%	16.67%
Stage 0 sporulation two-component response regulator (Spo0A)	Spo0A	268	95.50%	100%	80.90%	100.00%	77.30%	100.00%

Bacillus (*B.*) *subtilis oriC* displaying DnaA box clusters in the intergenic regions both upstream and downstream of *dnaA* and with probably two DNA-unwinding elements (DUE) in the *dnaA-dnaN* region in comparison to one predicted for *B. subtilis* (Briggs et al., 2012). The number of DnaA boxes with the consensus sequence (ttatccaca) was two for both *oriC* regions (Fig. 2).

4.3. Comparative genome analysis and visualization

Comparison of the three *C. chauvoei* chromosomes showed inversion and translocation of LCBs (Fig. 4). Symmetric inversion around the bacterial origin of replication is a common feature for closely related species (Eisen et al., 2000). Presence of blocks switched to different positions among isolates has also been observed in strains of *C. botulinum* (Fang et al., 2012). The *C. chauvoei* genome also shares some similar LCBs with *C. septicum* confirming the close relatedness of the two species (Fig. 4).

The analysis of genome sequence data to depict phylogenetic relatedness in the genus *Clostridium* revealed groups of species sharing homology. Significant homology was observed for *C. chauvoei* and *C. septicum*, *C. botulinum* type C and D, *C. novyi* and *C. haemolyticum*, for C. *botulinum* type A and *C. sporogenes* and for *C. botulinum* type B and E, *C. beijerinckii* and *C. butyricum* (Fig. 5). These findings are similar to previous phylogenetic studies based on 16S rDNA (Kuhnert et al., 1996; Stackebrandt et al., 1999). Based on *rrs* sequence analysis for nucleotide signatures and unique restriction enzyme digestion patterns, *C. chauvoei* was found to be phylogenetically related to *C. septicum* and *C. carnis* (Kalia et al., 2011; Kuhnert et al., 1996).

Based on panOCT (Fouts et al., 2012) analysis, the core and accessory genome of the three *C. chauvoei* isolates, 12S0467, DSM 7528^T (this study) and JF4335 were investigated. The genomes are highly similar in respect to gene content and differ only in a limited number of genes (Fig. 6). 12S0467 and the JF4335 harbored a segment of genes including a sialidase, alpha-L-fucosidase and NPTQN specific sortase B which are absent in the type strain and could be associated with bacterial pathogenicity. Alpha-L-fucosidases are involved in the degradation of intestinal fucosyl glycans by *C. perfringens* (Fan et al., 2016) and sortase B (SrtB), a cysteine transpeptidase has been identified in *Staphylococcus aureus* that attaches a polypeptide involved in heme iron transport and thus help in iron acquisition during infection (Zong et al., 2004). The majority of the unique genetic elements among the three genomes are genes for transposition, mobile element proteins and hypothetical proteins.

The genome features of *C. chauvoei* (DSM 7528^T) and *C. septicum* (CSUR P1044) were compared based on subsystem category distribution features in RAST (Aziz et al., 2008; Overbeek et al., 2014). Both genomes share similar subsystem category distributions whereas the subsystem feature counts were higher for *C. septicum* for genes related to prophages, phosphorus metabolism and carbohydrates (Fig. 7). Also the published genome for *C. septicum* has a 32-kb plasmid whereas similar large plasmids were not observed so far in characterized strains of *C. chauvoei*.

4.4. CRISPR and phage elements

CRISPR elements in bacteria confer protection against bacteriophages (Barrangou et al., 2007). The *C. chauvoei* genomes have two CRISPR sites which are similar to the CRISPR type I-B system based on the presence of *cas* genes in the order (*cas6-cas8b1-cas7-cas5-cas3-cas4-cas1-cas2*) according to the updated CRISPR type classification (Makarova et al., 2015). CRISPR elements have also been reported from a pathogenic *C. perfringens* type A strain isolated from a calf (Nowell et al., 2012). Interestingly no CRISPR elements were identified on the chromosome of a *C. botulinum* type III strain (BKT015925) of poultry origin but CRISPR elements were present in the prophage (p1) and on a plasmid (p2), respectively (Skarin et al., 2011). In a detailed analysis of the same strain it was found that the CRISPR type I—B system was incomplete for all *cas* genes and was predicted to be inactive (Woudstra et al., 2016). Both genomes in the current study show the presence of a complete CRISPR/Cas system with variations in respect to two spacer numbers between the two isolates indicating an active CRISPR/Cas system even though the genomes harbored prophages. The genomes have both intact and incomplete prophages. The intact prophage of both strains carried similar genes but was predicted with different attP/attL sequences for strain DSM 7528^T (TAAAAATAATGT)

and strain 12S0467 (TTTTTTTATTTT). One of the hemolysin genes previously described from the *C. chauvoei* genome belonging to the haemolysin XhlA superfamily was also found to be present in the prophage of both genomes (Frey and Falquet, 2015). The incomplete prophage carried various coding sequences (CDS) of bacterial origin which include few spore coat proteins and a sporulation-specific protease (YabG) which was recently characterized as a major regulator of germination in *C. difficile* (Kevorkian et al., 2016).

4.5. Spore resistance, sporulation and germination

Genes for small, acid-soluble spore proteins (SASPs) of the α/β types were identified in the genome. SASPs have a significant contribution towards spore resistance against heat, UV radiation, and chemical agents (Meaney et al., 2016; Setlow, 1988, 1994) and *C. perfringens* strains lacking a significant number of α/β-type SASPs developed more sensitive spores (Paredes-Sabja et al., 2008a; Raju et al., 2007). Dipicolinic acid (DPA), a major component of bacterial endospores, is made from the precursor dihydro-dipicolinic acid (DHDPA) by DHDPA synthase and finally DHDPA is oxidized to DPA by the products of the *spoVF* operon in the case of *Bacillus* and many *Clostridium* species. However many pathogens belonging to *Clostridium* cluster I such as *C. perfringens*, *C. botulinum* and *C. tetani* have no *spoVF* orthologues in their genomes (Durre, 2014). In the case of *C. perfringens*, it has been discovered that electron transfer flavoprotein (EtfA) carries out the production of dipicolinic acid from DHDPA (Orsburn et al., 2010). The *C. chauvoei* genome encodes a DHDPA synthase designated as DapA coding for 292 amino acids but a *spoVF* operon is absent as in other organisms of *Clostridia* cluster I. Two genes encoding EtfA proteins of 335 and 396 amino acids, respectively, were identified and were of similar length to that predicted for *C. perfringens* (Orsburn et al., 2010) suggesting an analogous function.

Spo0A, a transcriptional factor plays a central role in the sporulation process of bacteria of the genera *Clostridium* and *Bacillus*. There is also evidence showing Spo0A is involved in regulating various metabolic and virulence factors such as toxins in the genus *Clostridium* (Paredes-Sabja et al., 2011; Pettit et al., 2014). The *C. chauvoei* genome data also revealed the presence of the *spo0A* and other genes involved in various stages of sporulation. Initiation of sporulation in *Clostridium* species is carried out by kinases known as orphan kinases as they lack the cognate response regulator and the phosphorelay mechanism found in *Bacillus* (Talukdar et al., 2015). Most of the *C. chauvoei* genes for sporulation and germination showed greater homology with those of *C. septicum* followed by *C. perfringens* and *C. botulinum* which suggests a genetic relatedness and similar mechanisms for sporulation and germination. The Spo0A of all four species was compared and the result shows a central region with amino acid deletions for *C. chauvoei* and *C. septicum* (Fig. 8). The limited sequence variation of *Spo0A* has already been employed for specific detection of *C. chauvoei* and *C. septicum* based on real-time PCR (Lange et al., 2010). The exact role of Spo0A and its interaction with orphan kinases in *C. chauvoei* waits to be explored; furthermore, also the possible role of Spo0A in regulating the metabolic pathways and virulence factors in *C. chauvoei* is unclear.

Organization of germination receptors in clostridiae are found to be different from *B. subtillis* and also are strain dependent (Durre, 2014). In the case of *C. perfringens* the *gerK* operons are organized as a monocistronic *gerKB* and a bicistronic *gerKA-gerKC* transcriptional unit (Paredes-Sabja et al., 2009). The characterized *C. chauvoei* genome also shows the same arrangement. It can be assumed that a similar germination pattern with the essentiality for L-asparagine and the possibility of KCl-induced germination as observed for *C. perfingens* (Paredes-Sabja et al., 2008b) is present. Also the genome harbors a *prkC*, encoding a Ser/Thr kinase observed in *C. perfringens* suggestive for an alternative germination pathway triggered by environmental peptidoglycan fragments as described for *B. subtilis* (Xiao et al., 2015). The translated proteins of *C. chauvoei* DapA, GerKC, PrKC and germination-specific spore cortex-lytic enzymes SleB showed significantly lower homology with that of *C. septicum*, *C. perfringens* and *C. botulinum* suggestive for the involvement of species specific factors required for germination (Table 4).

4.6. Virulence factors and antibiotic resistance

Flagellin is the principle component of the bacterial flagellum, an important structure in the mobility of bacterial pathogens and also involved in pathogenicity. Earlier studies carried out at the nucleotide sequence level have shown the existence of two *fliCs* which occur in tandem, designated as *fliA*(*C*) and *fliB*(*C*) having highly conserved N and C terminal regions. Also flagella and *fliC* genes are considered as one of the important vaccine and diagnostic candidate antigens of the pathogen (Tamura and Tanaka, 1984; Tanaka et al., 1987; Usharani et al., 2015). The sequence analysis of the two *C. chauvoei* isolates described in the current study revealed the presence of three *fliC* genes instead of two. This is the first report of triplicate *fliC* genes for the species, a finding of possible importance for pathogen detection and immunization. The spacer regions between the three *fliC* genes are conserved and all three genes have a central variable region (Fig. 9), a structure not easily resolved by sequencing techniques employing short reads. Other genes involved in motility and chemotaxis were found to be homologous between the three sequenced strains. Flagella, toxins and lytic enzymes are primary virulence factors in *C. chauvoei* (Frey and Falquet, 2015). Recently "*Clostridium chauvoei* toxin A" (CctA) belonging to the leucocidin superfamily of bacterial toxins was identified as a major virulence factor in *C. chauvoei,* and not found in other clostridiae (Frey et al., 2012). No genetic variability for the CctA protein at the amino acid level was

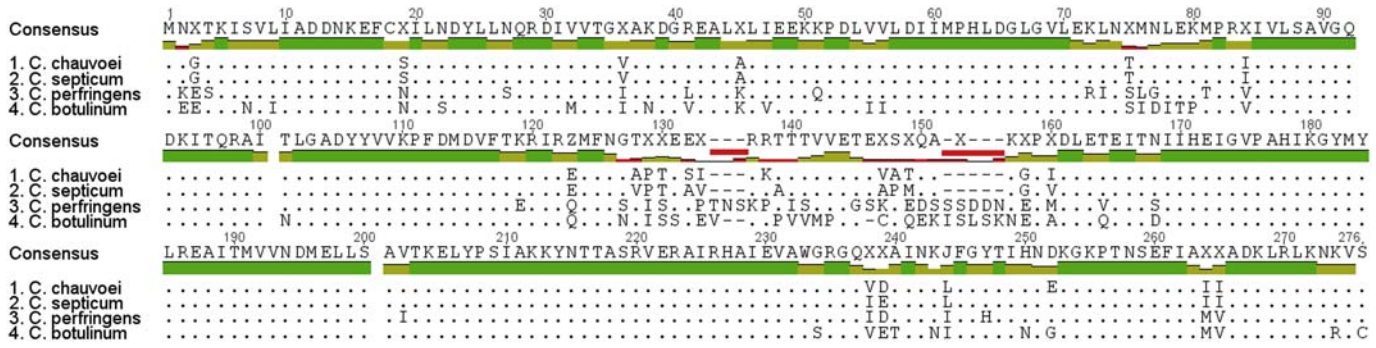


Fig. 8. Alignment of the Spo0A protein sequence from four species. Alignment of the SpoA protein sequence of *C. chauvoei*, *C. septicum* (CSUR P1044), *C. perfringens* (ATCC 13124) and *C. botulinum* (ATCC 3502). Conserved amino acids are depicted as dots. The red line indicates the amino acid deletions for *C. chauvoei* and *C. septicum* SpoA proteins. Alignment and schematic representation was created using Geneious 9.0.5 (http://www.geneious.com/).

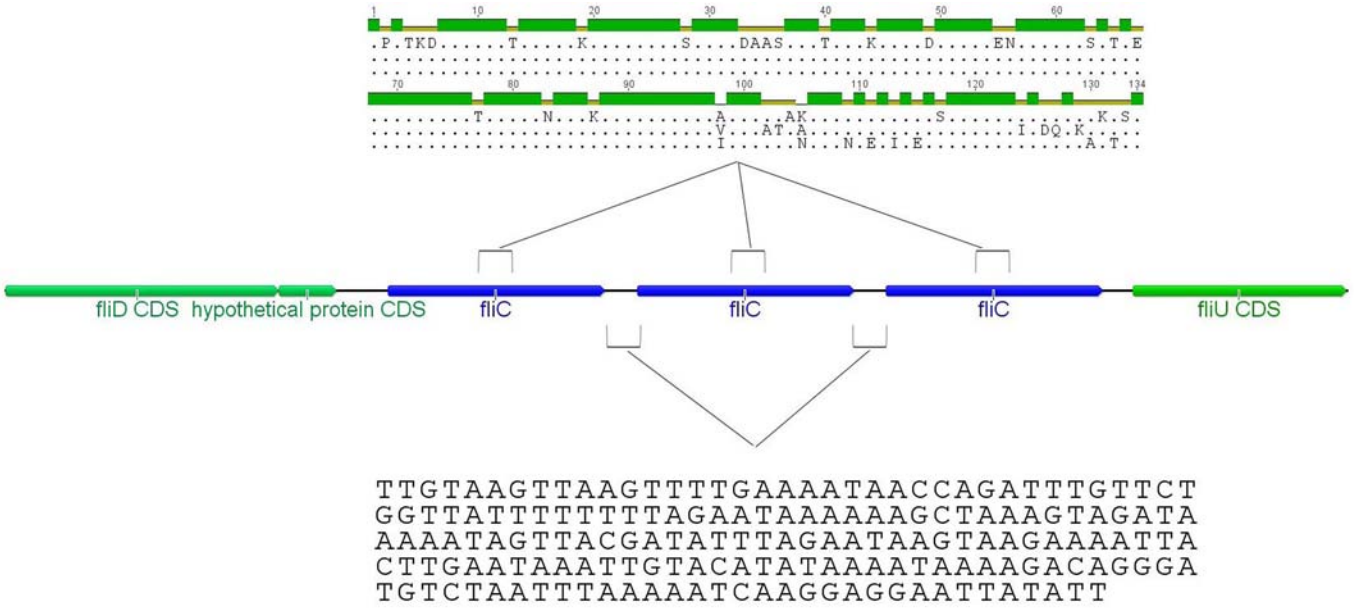


Fig. 9. Schematic representation of multiple *fliC* genes. Arrangement of the three *fliC* genes in the flagellar operon. The two 181 bp spacer regions between the *fliC* genes were identical (shown below). Alignment of central variable region of *fliC* (amino acid position 177 to 310) is shown above with conserved amino acids depicted as dots. Alignment and schematic representation was created using Geneious 9.0.5 (http://www.geneious.com/).

found for the three strains, even though an earlier study reported three nucleotide polymorphisms for one strain isolated 1956 in New Zealand (Frey et al., 2012).

Sialidases or neuraminidases are among the few previously characterized virulence factors *of C. chauvoei* (Useh et al., 2006; Vilei et al., 2011). The sialidase gene *nanA* showed 100% similarity at amino acid level in the three genomes. All genomes harbored a NagH homologue gene encoding 1887 amino acids with a variation in the coding for the amino acid at position 596 where methionine is substituted with an isoleucine for 12S0467. The *C. chauvoei* genome also harbors many genes involved in hydrolysis of *O*-linked glycans based on GO annotation. Glycoside hydrolases have been identified as virulence factors of bacterial pathogens during infections (Frederiksen et al., 2013). Pneumococcal *O*-glycosidase in conjunction with the neuraminidase NanA, sequentially deglycosylates *O*-linked glycan structures in the mucous layer of epithelia promoting bacterial colonization (Marion et al., 2009). A recent study on the membrane proteome of *C. chauvoei* has identified a putative glycosyl hydrolase to be cell-wall associated and immunogenic (Jayaramaiah et al., 2016). Proteolytic enzymes may help in the survival of *C. chauvoei* in muscle tissue and thus contribute to pathogenesis. There are also many genes involved in the transportation and acquisition of iron and other metal ions which are essential for the survival of pathogens in the host (Porcheron et al., 2013). Further genes potentially involved in the pathogenesis of *C. chauvoei* include phospholipases, collagen binding proteins and genes encoding homologues to internalin A (Frey and Falquet, 2015). The genomes harbor three coding sequences for internalin A which showed no variation among the strains.

Antibiotic resistance genes encoded in the genome confer resistance to bacterial pathogens. An earlier study based on genome sequence of strain *C. chauvoei* JF4335 reported a number of genes for antibiotic resistance (a gene for penicillin resistance (a beta-lactamase gene), an elongation factor G (EF-G) type tetracycline resistance gene (*tetM* and *tetO* analogues), a vancomycin B-type resistance gene (*vanW*)). However, that same study described the isolate (and others used in the study) sensitive to all antibiotics tested which are usually used against clostridial infections and hence postulated the possibility for non-expression or expression of non-functional proteins (Frey and Falquet, 2015). Hence, additional genes potentially conferring resistance for peptide antibiotics and macrolides were found in the present study in both genomes. Three genes which are protein variants in the CARD database were identified of which two encoded a *gyr*B related gene possibly conferring resistance to aminocoumarin and one *rpo*B potentially conferring resistance to rifampicin. Even though specific SNPs can confer resistance to antibiotics in bacteria (McArthur et al., 2013), whether the protein variants identified in *C. chauvoei* genomes play any role in antibiotic resistance is doubtful. However, antimicrobial resistance was not tested in our study.

5. Conclusion

High quality circular genome sequences were generated for the *C. chauvoei* type strain DSM 7528^T (ATCC 10092^T) and a field strain 12S0467 isolated in Germany. Similar prophages were identified in both genomes which harbored hemolysin and bacterial spore formation genes. A CRISPR type I-B system was observed in both genomes. Sporulation and germination process related genes were homologous to those of the *Clostridia* cluster I group, but novel variations in regulatory genes were also identified. Phylogenomics showed a higher relatedness to *C. septicum* than to other so far sequenced genomes of species belonging to the genus *Clostridium*. Comparative genome analysis of three *C. chauvoei* circular genome sequences revealed the presence of few inversions and translocations in locally collinear blocks (LCBs). The species genome also shows a large number of genes involved in proteolysis, glycosyl hydrolase genes and metal iron transportation genes which are presumably involved in virulence and survival in the host. Three conserved flagellar genes (*fliC*) were identified in each of the circular genomes. The data presented here will facilitate future studies to investigate *C. chauvoei* evolution and pathobiology.

Acknowledgements

The ICAR-International Fellowship from Indian Council of Agricultural Research (ICAR), New Delhi, India, for Prasad Thomas is gratefully acknowledged.

References

Ågren, J., Sundström, A., Håfström, T., Segerman, B., 2012. Gegenees: fragmented alignment of multiple genomes for determining Phylogenomic distances and genetic signatures unique for specified target groups. PLoS One 7, e39107.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., Zagnitko, O., 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics 9, 75.

Bardou, P., Mariette, J., Escudié, F., Djemiel, C., Klopp, C., 2014. jvenn: an interactive Venn diagram viewer. BMC Bioinformatics 15, 293.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., Horvath, P., 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science (New York, N.Y.) 315, 1709–1712.

Benamar, S., Cassir, N., Caputo, A., Cadoret, F., La Scola, B., 2016. Complete genome sequence of *Clostridium septicum* strain CSUR P1044, isolated from the human gut microbiota. Genome Announc. 4.

Binnewies, T.T., Motro, Y., Hallin, P.F., Lund, O., Dunn, D., La, T., Hampson, D.J., Bellgard, M., Wassenaar, T.M., Ussery, D.W., 2006. Ten years of bacterial genome sequencing: comparative-genomics-based discoveries. Funct. Integr. Genomics 6, 165–185.

Bland, C., Ramsey, T.L., Sabree, F., Lowe, M., Brown, K., Kyrpides, N.C., Hugenholtz, P., 2007. CRISPR Recognition Tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. BMC Bioinformatics 8, 1–8.

Briggs, G.S., Smits, W.K., Soultanas, P., 2012. Chromosomal replication initiation machinery of low-G + C-content Firmicutes. J. Bacteriol. 194, 5162–5170.

Brown, S.D., Nagaraju, S., Utturkar, S., De Tissera, S., Segovia, S., Mitchell, W., Land, M.L., Dassanayake, A., Köpke, M., 2014. Comparison of single-molecule sequencing and hybrid approaches for finishing the genome of *Clostridium autoethanogenum* and analysis of CRISPR systems in industrial relevant clostridia. Biotechnology for Biofuels 7, 1–18.

Carter, A.T., Peck, M.W., 2015. Genomes, neurotoxins and biology of *Clostridium botulinum* Group I and Group II. Res. Microbiol. 166, 303–317.

Carver, T.J., Rutherford, K.M., Berriman, M., Rajandream, M.A., Barrell, B.G., Parkhill, J., 2005. ACT: the Artemis Comparison Tool. Bioinformatics (Oxford, England) 21, 3422–3423.

Carver, T., Thomson, N., Bleasby, A., Berriman, M., Parkhill, J., 2009. DNAPlotter: circular and linear interactive genome visualization. Bioinformatics (Oxford, England) 25, 119–120.

Chin, C.S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat. Methods 10.

Collins, M.D., Lawson, P.A., Willems, A., Cordoba, J.J., Fernandez-Garayzabal, J., Garcia, P., Cai, J., Hippe, H., Farrow, J.A., 1994. The phylogeny of the genus *Clostridium*: proposal of five new genera and eleven new species combinations. Int. J. Syst. Bacteriol. 44, 812–826.

Conesa, A., Gotz, S., 2008. Blast2GO: a comprehensive suite for functional analysis in plant genomics. International Journal of Plant Genomics 2008, 619832.

Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., Robles, M., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics (Oxford, England) 21, 3674–3676.

Darling, A.E., Mau, B., Perna, N.T., 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5, e11147.

Delcher, A.L., Bratke, K.A., Powers, E.C., Salzberg, S.L., 2007. Identifying bacterial genes and endosymbiont DNA with glimmer. Bioinformatics (Oxford, England) 23, 673–679.

Durre, P., 2014. Physiology and sporulation in *Clostridium*. Microbiology Spectrum 2, pp. Tbs-0010–2012.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R., Clark, S., Dalal, R., Dewinter, A., Dixon, J., Foquet, M., Gaertner, A., Hardenbol, P., Heiner, C., Hester, K., Holden, D., Kearns, G., Kong, X., Kuse, R., Lacroix, Y., Lin, S., Lundquist, P., Ma, C., Marks, P., Maxham, M., Murphy, D., Park, I., Pham, T., Phillips, M., Roy, J., Sebra, R., Shen, G., Sorenson, J., Tomaney, A., Travers, K., Trulson, M., Vieceli, J., Wegener, J., Wu, D., Yang, A., Zaccarin, D., Zhao, P., Zhong, F., Korlach, J., Turner, S., 2009. Real-time DNA sequencing from single polymerase molecules. Science (New York, N.Y.) 323, 133–138.

Eisen, J.A., Heidelberg, J.F., White, O., Salzberg, S.L., 2000. Evidence for symmetric chromosomal inversions around the replication origin in bacteria. Genome Biol. 1.

Falquet, L., Calderon-Copete, S.P., Frey, J., 2013. Draft genome sequence of the virulent *Clostridium chauvoei* reference strain JF4335. Genome Announc. 1.

Fan, S., Zhang, H., Chen, X., Lu, L., Xu, L., Xiao, M., 2016. Cloning, characterization, and production of three alpha-l-fucosidases from *Clostridium perfringens* ATCC 13124. J. Basic Microbiol. 56, 347–357.

Fang, P.K., Raphael, B.H., Maslanka, S.E., Cai, S., Singh, B.R., 2010. Analysis of genomic differences among *Clostridium botulinum* type A1 strains. BMC Genomics 11, 725.

Fang, G., Munera, D., Friedman, D.I., Mandlik, A., Chao, M.C., Banerjee, O., Feng, Z., Losic, B., Mahajan, M.C., Jabado, O.J., Deikus, G., Clark, T.A., Luong, K., Murray, I.A., Davis, B.M., Keren-Paz, A., Chess, A., Roberts, R.J., Korlach, J., Turner, S.W., Kumar, V., Waldor, M.K., Schadt, E.E., 2012. Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing. Nat. Biotechnol. 30.

Fouts, D.E., Brinkac, L., Beck, E., Inman, J., Sutton, G., 2012. PanOCT: automated clustering of orthologs using conserved gene neighborhood for pan-genomic analysis of bacterial strains and closely related species. Nucleic Acids Res. 40, e172.

Frederiksen, R.F., Paspaliari, D.K., Larsen, T., Storgaard, B.G., Larsen, M.H., Ingmer, H., Palcic, M.M., Leisner, J.J., 2013. Bacterial chitinases and chitin-binding proteins as virulence factors. Microbiology 159, 833–847.

Frey, J., Falquet, L., 2015. Patho-genetics of *Clostridium chauvoei*. Res. Microbiol. 166, 384–392.

Frey, J., Johansson, A., Burki, S., Vilei, E.M., Redhead, K., 2012. Cytotoxin CctA, a major virulence factor of *Clostridium chauvoei* conferring protective immunity against myonecrosis. Vaccine 30, 5500–5505.

Gao, F., Zhang, C.T., 2008. Ori-Finder: a web-based system for finding oriCs in unannotated bacterial genomes. BMC Bioinformatics 9, 79.

Götz, S., García-Gómez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talón, M., Dopazo, J., Conesa, A., 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36, 3420–3435.

Gotz, S., Arnold, R., Sebastian-Leon, P., Martin-Rodriguez, S., Tischler, P., Jehl, M.A., Dopazo, J., Rattei, T., Conesa, A., 2011. B2G-FAR, a species-centered GO annotation repository. Bioinformatics (Oxford, England) 27, 919–924.

Hatheway, C.L., 1990. Toxigenic clostridia. Clin. Microbiol. Rev. 3.

Hunt, M., Silva, N.D., Otto, T.D., Parkhill, J., Keane, J.A., Harris, S.R., 2015. Circlator: automated circularization of genome assemblies using long sequencing reads. Genome Biol. 16, 1–10.

Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254–267.

Jayaramaiah, U., Singh, N., Thankappan, S., Mohanty, A.K., Chaudhuri, P., Singh, V.P., Nagaleekar, V.K., 2016. Proteomic analysis and identification of cell surface-associated proteins of *Clostridium chauvoei*. Anaerobe 39, 77–83.

Kalia, V.C., Mukherjee, T., Bhushan, A., Joshi, J., Shankar, P., Huma, N., 2011. Analysis of the unexplored features of rrs (16S rDNA) of the genus *Clostridium*. BMC Genomics 12, 18.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., Drummond, A., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics (Oxford, England) 28, 1647–1649.

Kevorkian, Y., Shirley, D.J., Shen, A., 2016. Regulation of *Clostridium difficile* spore germination by the CspA pseudoprotease domain. Biochimie 122, 243–254.

Koren, S., Harhay, G.P., Smith, T.P., Bono, J.L., Harhay, D.M., Mcvey, S.D., Radune, D., Bergman, N.H., Phillippy, A.M., 2013. Reducing assembly complexity of microbial genomes with single-molecule sequencing. Genome Biol. 14, 1–16.

Krumsiek, J., Arnold, R., Rattei, T., 2007. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. Bioinformatics (Oxford, England) 23, 1026–1028.

Kuhnert, P., Capaul, S.E., Nicolet, J., Frey, J., 1996. Phylogenetic positions of *Clostridium chauvoei* and *Clostridium septicum* based on 16S rRNA gene sequences. Int. J. Syst. Bacteriol. 46, 1174–1176.

Lagesen, K., Hallin, P., Rodland, E.A., Staerfeldt, H.H., Rognes, T., Ussery, D.W., 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 35, 3100–3108.

Lange, M., Neubauer, H., Seyboldt, C., 2010. Development and validation of a multiplex real-time PCR for detection of *Clostridium chauvoei* and *Clostridium septicum*. Mol. Cell. Probes 24, 204–210.

Liao, Y.-C., Lin, S.-H., Lin, H.-H., 2015. Completing bacterial genome assemblies: strategy and performance comparisons. Sci Rep 5, 8747.

Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25, 0955–0964.

Mackiewicz, P., Zakrzewska-Czerwinska, J., Zawilak, A., Dudek, M.R., Cebrat, S., 2004. Where does bacterial replication start? Rules for predicting the oriC region. Nucleic Acids Res. 32, 3781–3791.

Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J.J., Charpentier, E., Haft, D.H., Horvath, P., Moineau, S., Mojica, F.J.M., Terns, R.M., Terns, M.P., White, M.F., Yakunin, A.F., Garrett, R.A., van der Oost, J., Backofen, R., Koonin, E.V., 2015. An updated evolutionary classification of CRISPR-Cas systems. Nat. Rev. Microbiol. 13, 722–736.

Marion, C., Limoli, D.H., Bobulsky, G.S., Abraham, J.L., Burnaugh, A.M., King, S.J., 2009. Identification of a pneumococcal glycosidase that modifies *O*-linked glycans. Infect. Immun. 77, 1389–1396.

McArthur, A.G., Waglechner, N., Nizam, F., Yan, A., Azad, M.A., Baylay, A.J., Bhullar, K., Canova, M.J., De Pascale, G., Ejim, L., Kalan, L., King, A.M., Koteva, K., Morar, M., Mulvey, M.R., O'Brien, J.S., Pawlowski, A.C., Piddock, L.J., Spanogiannopoulos, P., Sutherland, A.D., Tang, I., Taylor, P.L., Thaker, M., Wang, W., Yan, M., Yu, T., Wright, G.D., 2013. The comprehensive antibiotic resistance database. Antimicrob. Agents Chemother. 57, 3348–3357.

Meaney, C.A., Cartman, S.T., McClure, P.J., Minton, N.P., 2016. The role of small acid-soluble proteins (SASPs) in protection of spores of *Clostridium botulinum* against nitrous acid. Int. J. Food Microbiol. 216, 25–30.

Myers, G.S., Rasko, D.A., Cheung, J.K., Ravel, J., Seshadri, R., DeBoy, R.T., Ren, Q., Varga, J., Awad, M.M., Brinkac, L.M., Daugherty, S.C., Haft, D.H., Dodson, R.J., Madupu, R., Nelson, W.C., Rosovitz, M.J., Sullivan, S.A., Khouri, H., Dimitrov, G.I., Watkins, K.L., Mulligan, S., Benton, J., Radune, D., Fisher, D.J., Atkins, H.S., Hiscox, T., Jost, B.H., Billington, S.J., Songer, J.G., McClane, B.A., Titball, R.W., Rood, J.I., Melville, S.B., Paulsen, I.T., 2006. Skewed genomic variability in strains of the toxigenic bacterial pathogen, *Clostridium perfringens*. Genome Res. 16, 1031–1040.

Nagano, N., Isomine, S., Kato, H., Sasaki, Y., Takahashi, M., Sakaida, K., Nagano, Y., Arakawa, Y., 2008. Human fulminant gas gangrene caused by *Clostridium chauvoei*. J. Clin. Microbiol. 46, 1545–1547.

Nowell, V.J., Kropinski, A.M., Songer, J.G., MacInnes, J.I., Parreira, V.R., Prescott, J.F., 2012. Genome sequencing and analysis of a type A *Clostridium perfringens* isolate from a case of bovine clostridial abomasitis. PLoS One 7, e32271.

Orsburn, B.C., Melville, S.B., Popham, D.L., 2010. EtfA catalyses the formation of dipicolinic acid in *Clostridium perfringens*. Mol. Microbiol. 75, 178–186.

Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., Edwards, R.A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A.R., Xia, F., Stevens, R., 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res. 42, D206–D214.

Paredes-Sabja, D., Raju, D., Torres, J.A., Sarker, M.R., 2008a. Role of small, acid-soluble spore proteins in the resistance of *Clostridium perfringens* spores to chemicals. Int. J. Food Microbiol. 122, 333–335.

Paredes-Sabja, D., Torres, J.A., Setlow, P., Sarker, M.R., 2008b. *Clostridium perfringens* spore germination: characterization of germinants and their receptors. J. Bacteriol. 190, 1190–1201.

Paredes-Sabja, D., Setlow, P., Sarker, M.R., 2009. Role of GerKB in germination and outgrowth of *Clostridium perfringens* spores. Appl. Environ. Microbiol. 75, 3813–3817.

Paredes-Sabja, D., Sarker, N., Sarker, M.R., 2011. *Clostridium perfringens* tpeL is expressed during sporulation. Microb. Pathog. 51, 384–388.

Pettit, L.J., Browne, H.P., Yu, L., Smits, W.K., Fagan, R.P., Barquist, L., Martin, M.J., Goulding, D., Duncan, S.H., Flint, H.J., Dougan, G., Choudhary, J.S., Lawley, T.D., 2014. Functional genomics reveals that *Clostridium difficile* Spo0A coordinates sporulation, virulence and metabolism. BMC Genomics 15, 1–15.

Porcheron, G., Garénaux, A., Proulx, J., Sabri, M., Dozois, C.M., 2013. Iron, copper, zinc, and manganese transport and regulation in pathogenic Enterobacteria: correlations between strains, site of infection and the relative importance of the different metal transport systems for virulence. Front. Cell. Infect. Microbiol. 3, 90.

Raju, D., Setlow, P., Sarker, M.R., 2007. Antisense-RNA-mediated decreased synthesis of small, acid-soluble spore proteins leads to decreased resistance of *Clostridium perfringens* spores to moist heat and UV radiation. Appl. Environ. Microbiol. 73, 2048–2053.

Rhoads, A., Au, K.F., 2015. PacBio sequencing and its applications. Genomics, Proteomics & Bioinformatics 13, 278–289.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.-A., Barrell, B., 2000. Artemis: sequence visualization and annotation. Bioinformatics (Oxford, England) 16, 944–945.

Sasaki, Y., Yamamoto, K., Kojima, A., Tetsuka, Y., Norimatsu, M., Tamura, Y., 2000. Rapid and direct detection of *Clostridium chauvoei* by PCR of the 16S-23S rDNA spacer region and partial 23S rDNA sequences. J. Vet. Med. Sci. 62, 1275–1281.

Sasaki, Y., Kojima, A., Aoki, H., Ogikubo, Y., Takikawa, N., Tamura, Y., 2002. Phylogenetic analysis and PCR detection of *Clostridium chauvoei*, *Clostridium haemolyticum*, *Clostridium novyi* types A and B, and *Clostridium septicum* based on the flagellin gene. Vet. Microbiol. 86, 257–267.

Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics (Oxford, England).

Selzer, J., Hofmann, F., Rex, G., Wilm, M., Mann, M., Just, I., Aktories, K., 1996. Clostridium novyi alpha-toxin-catalyzed incorporation of GlcNAc into Rho subfamily proteins. J. Biol. Chem. 271, 25173–25177.

Setlow, P., 1988. Small, acid-soluble spore proteins of *Bacillus* species: structure, synthesis, genetics, function, and degradation. Annu. Rev. Microbiol. 42, 319–338.

Setlow, P., 1994. Mechanisms which contribute to the long-term survival of spores of *Bacillus* species. Society for Applied Bacteriology symposium series 23 pp. 49s–60s.

Skarin, H., Håfström, T., Westerberg, J., Segerman, B., 2011. *Clostridium botulinum* group III: a group with dual identity shaped by plasmids, phages and mobile elements. BMC Genomics 12, 1–13.

Sommer, D.D., Delcher, A.L., Salzberg, S.L., Pop, M., 2007. Minimus: a fast, lightweight genome assembler. BMC Bioinformatics 8, 64.

Stackebrandt, E., Kramer, I., Swiderski, J., Hippe, H., 1999. Phylogenetic basis for a taxonomic dissection of the genus *Clostridium*. FEMS Immunol. Med. Microbiol. 24, 253–258.

Talukdar, P.K., Olguín-Araneda, V., Alnoman, M., Paredes-Sabja, D., Sarker, M.R., 2015. Updates on the sporulation process in *Clostridium* species. Res. Microbiol. 166, 225–235.

Tamura, Y., Tanaka, S., 1984. Effect of antiflagellar serum in the protection of mice against *Clostridium chauvoei*. Infect. Immun. 43, 612–616.

Tamura, Y., Kijima-Tanaka, M., Aoki, A., Ogikubo, Y., Takahashi, T., 1995. Reversible expression of motility and flagella in *Clostridium chauvoei* and their relationship to virulence. Microbiology 141 (Pt 3), 605–610.

Tanaka, M., Hirayama, N., Tamura, Y., 1987. Production, characterization, and protective effect of monoclonal antibodies to *Clostridium chauvoei* flagella. Infect. Immun. 55, 1779–1783.

Useh, N.M., Ibrahim, N.D., Nok, A.J., Esievo, K.A., 2006. Relationship between outbreaks of blackleg in cattle and annual rainfall in Zaria, Nigeria. Vet. Rec. 158, 100–101.

Usharani, J., Nagaleekar, V.K., Thomas, P., Gupta, S.K., Bhure, S.K., Dandapat, P., Agarwal, R.K., Singh, V.P., 2015. Development of a recombinant flagellin based ELISA for the detection of *Clostridium chauvoei*. Anaerobe 33, 48–54.

Vilei, E.M., Johansson, A., Schlatter, Y., Redhead, K., Frey, J., 2011. Genetic and functional characterization of the NanA sialidase from *Clostridium chauvoei*. Vet. Res. 42.

Weatherhead, J.E., Tweardy, D.J., 2012. Lethal human neutropenic enterocolitis caused by *Clostridium chauvoei* in the United States: tip of the iceberg? The Journal of Infection 64, 225–227.

Woudstra, C., Le Maréchal, C., Souillard, R., Bayon-Auboyer, M.-H., Mermoud, I., Desoutter, D., Fach, P., 2016. New insights into the genetic diversity of *Clostridium botulinum* group III through extensive genome exploration. Front. Microbiol. 7, 757.

Xiao, Y., van Hijum, S.A.F.T., Abee, T., Wells-Bennik, M.H.J., 2015. Genome-wide transcriptional profiling of *Clostridium perfringens* SM101 during sporulation extends the core of putative sporulation genes and genes determining spore properties and germination characteristics. PLoS One 10, e0127036.

Zhou, Y., Liang, Y., Lynch, K.H., Dennis, J.J., Wishart, D.S., 2011. PHAST: a fast phage search tool. Nucleic Acids Res. 39, W347–352.

Zong, Y., Mazmanian, S.K., Schneewind, O., Narayana, S.V., 2004. The structure of sortase B, a cysteine transpeptidase that tethers surface protein to the *Staphylococcus aureus* cell wall. Structure (London, England: 1993) 12, 105–112.