Non contiguous-finished genome sequence and description of 
Clostridium jeddahense sp. nov.

Jean-Christophe Lagier¹, Fehmida Bibi², Dhamodharan Ramasamy¹, Esam I. Azhar²,³, Catherine Robert¹, Muhammad Yasir², Asif A. Jiman-Fatani⁴, Khalid Z. Alshali⁵, Pierre-Edouard Fournier¹, Didier Raoult¹,²

¹ Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, UMR CNRS 7278 – IRD 198, Institut Hospitalo-Universitaire Méditerranée-Infection, Faculté de médecine, Aix-Marseille Université, France
²Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia
³Department of Medical Laboratory Technology, Faculty of Applied Medical Sciences, King Abdulaziz University, Jeddah, Saudi Arabia
⁴Department of Medical Microbiology and Parasitology, Faculty of Medicine, King Abdulaziz University, Jeddah, Saudi Arabia
⁵Department of Medicine, Faculty of Medicine, King Abdulaziz University, PO Box 80215, Jeddah, Postal code 21589, Saudi Arabia

*Correspondence: Didier Raoult (didier.raoult@gmail.com)

Keywords: Clostridium jeddahense, genome, culturomics, taxonogenomics

Clostridium jeddahense strain JCDᵀ (=CSUR P693 = DSM 27834) is the type strain of C. jeddahense sp. nov. This strain, whose genome is described here, was isolated from the fecal flora of an obese 24 year-old Saudian male (BMI=52 kg/m²). Clostridium jeddahense strain JCDᵀ is an obligate Gram-positive bacillus. Here we describe the features of this organism, together with the complete genome sequence and annotation. The 3,613,503 bp long genome (1 chromosome, no plasmid) exhibits a G+C content of 51.95% and contains 3,462 protein-coding and 53 RNA genes, including 4 rRNA genes.

Introduction

Clostridium jeddahense strain JCDᵀ (=CSUR P693 = DSM 27834), is the type strain of Clostridium jeddahense sp. nov. This bacterium is a Gram-positive, anaerobic, spore-forming indole, positive bacillus that was isolated from the stool of an obese 24 year-old Saudian individual, as a part of a culturomics study as previously reported. The usual parameters used to delineate a bacterial species include 16S rDNA sequence identity and phylogeny [1,2], genomic G + C content diversity, and DNA–DNA hybridization (DDH) [3,4]. Nevertheless, some limitations appeared notably because the cutoff values vary dramatically between species and genera [5]. The introduction of high-throughput sequencing techniques made genomic data for many bacterial species available [6]. We recently proposed a new method (taxonogenomics), which includes genomic data in a polyphasic approach to describe new bacterial species [6]. This strategy combines phenotypic characteristics, including MALDI-TOF MS spectrum, and genomic analysis [7-37].

Here, we present a summary classification and a set of features for C. jeddahense sp. nov. strain JCDᵀ (=CSUR P693 = DSM 27834), together with the description of the complete genome sequencing and annotation. These characteristics support the circumscription of the species C. jeddahense. The genus Clostridium was created in 1880 [38] and consists of obligate anaerobic rod-shaped bacilli able to produce endospores [38]. More than 200 species have been described to date (http://www.bacterio.cict.fr/c/clostridium.html). Members of the genus Clostridium are mostly environmental bacteria or associated with the commensal digestive flora of mammals. However, several are major human pathogens, including C. botulinum, C. difficile and C. tetani [38].
Classification and features

A stool sample was collected from an obese 24-year-old male Saudian volunteer patient living in Jeddah. The patient gave an informed and signed consent, and the agreement of the Ethical Committee of the King Abdulaziz University, King Fahd medical Research Centre, Saudi Arabia, and the local ethics committee of the IFR48 (Marseille, France) were obtained under agreement number 014-CEGMR-2-ETH-P and 09-022 respectively. The fecal specimen was preserved at -80°C after collection and sent to Marseille. Strain JCD\textsuperscript{T} (Table 1) was isolated in July 2013 by anaerobic cultivation on 5% sheep blood-enriched Columbia agar (BioMerieux, Marcy l’Etoile, France) after a 5-day preincubation on blood culture bottle with rumen fluid. This strain exhibited a 97.3% nucleotide sequence similarity with \textit{Clostridium sporosphae-roides} strain DSM 1294 (Figure 1). This value was lower than the 98.7% 16S rRNA gene sequence similarity threshold recommended by Stackebrandt and Ebers to delineate a new species without carrying out DNA-DNA hybridization [2] and was in the 78.4 to 98.9% range of 16S rRNA identity values observed among 41 \textit{Clostridium} species with validly published names [52].

Table 1. Classification and general features of \textit{Clostridium jeddahense} strain JCD\textsuperscript{T} according to the MIGS recommendations [39]

<table>
<thead>
<tr>
<th>MIGS ID</th>
<th>Property</th>
<th>Term</th>
<th>Evidence code$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Current classification</td>
<td>Domain Bacteria</td>
<td>TAS [40]</td>
</tr>
<tr>
<td></td>
<td>Phylum</td>
<td>Firmicutes</td>
<td>TAS [41-43]</td>
</tr>
<tr>
<td></td>
<td>Class</td>
<td>Clostridia</td>
<td>TAS [44,45]</td>
</tr>
<tr>
<td></td>
<td>Order</td>
<td>Clostridales</td>
<td>TAS [46,47]</td>
</tr>
<tr>
<td></td>
<td>Family</td>
<td>Clostridiaceae</td>
<td>TAS [46,48]</td>
</tr>
<tr>
<td></td>
<td>Genus</td>
<td>Clostridium</td>
<td>IDA [46,49,50]</td>
</tr>
<tr>
<td></td>
<td>Species</td>
<td>Clostridium jeddahense</td>
<td>IDA</td>
</tr>
<tr>
<td></td>
<td>Type</td>
<td>strain JCD\textsuperscript{T}</td>
<td>IDA</td>
</tr>
<tr>
<td></td>
<td>Gram stain</td>
<td>Positive</td>
<td>IDA</td>
</tr>
<tr>
<td></td>
<td>Cell shape</td>
<td>Rod</td>
<td>IDA</td>
</tr>
<tr>
<td></td>
<td>Motility</td>
<td>Motile</td>
<td>IDA</td>
</tr>
<tr>
<td></td>
<td>Sporulation</td>
<td>Sporulating</td>
<td>IDA</td>
</tr>
<tr>
<td></td>
<td>Temperature range</td>
<td>Mesophile</td>
<td>IDA</td>
</tr>
<tr>
<td></td>
<td>Optimum temperature</td>
<td>37°C</td>
<td>IDA</td>
</tr>
<tr>
<td>MIGS-6.3</td>
<td>Salinity</td>
<td>Unknown</td>
<td>IDA</td>
</tr>
<tr>
<td>MIGS-22</td>
<td>Oxygen requirement</td>
<td>Anaerobic</td>
<td>IDA</td>
</tr>
<tr>
<td></td>
<td>Carbon source</td>
<td>Unknown</td>
<td>IDA</td>
</tr>
<tr>
<td></td>
<td>Energy source</td>
<td>Unknown</td>
<td>IDA</td>
</tr>
<tr>
<td>MIGS-6</td>
<td>Habitat</td>
<td>Human gut</td>
<td>IDA</td>
</tr>
<tr>
<td>MIGS-15</td>
<td>Biotic relationship</td>
<td>Free living</td>
<td>IDA</td>
</tr>
<tr>
<td></td>
<td>Pathogenicity</td>
<td>Unknown</td>
<td>IDA</td>
</tr>
<tr>
<td></td>
<td>Biosafety level</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>MIGS-14</td>
<td>Isolation</td>
<td>Human feces</td>
<td>IDA</td>
</tr>
<tr>
<td>MIGS-4</td>
<td>Geographic location</td>
<td>Jeddah, Saudi Arabia</td>
<td>IDA</td>
</tr>
<tr>
<td>MIGS-5</td>
<td>Sample collection time</td>
<td>July 2013</td>
<td>IDA</td>
</tr>
<tr>
<td>MIGS-4.1</td>
<td>Latitude</td>
<td>21.422487</td>
<td>IDA</td>
</tr>
<tr>
<td>MIGS-4.1</td>
<td>Longitude</td>
<td>39.856184</td>
<td>IDA</td>
</tr>
<tr>
<td>MIGS-4.3</td>
<td>Depth</td>
<td>Surface</td>
<td>IDA</td>
</tr>
<tr>
<td>MIGS-4.4</td>
<td>Altitude</td>
<td>0 m above sea level</td>
<td>IDA</td>
</tr>
</tbody>
</table>

Evidence codes - IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [51]. If the evidence is IDA, then the property was directly observed for a live isolate by one of the authors or an expert mentioned in the acknowledgements.
Figure 1. A consensus phylogenetic tree highlighting the position of Clostridium jeddahense strain JCD\textsuperscript{T} relative to other type strains within the Clostridium genus. GenBank accession numbers are indicated in parentheses. Sequences were aligned using CLUSTALW, and phylogenetic inferences were obtained using the maximum-likelihood method in the MEGA software package. Numbers at the nodes are the percentages of bootstrap values from 500 replicates that support the node. Clostridium ramosum was used as outgroup. The scale bar represents a 2% nucleotide sequence divergence.

Four growth temperatures (25, 30, 37, 45°C) were tested; growth occurred between 25 and 37°C, but optimal growth was observed at 37°C, 24 hours after inoculation. No growth occurred at 45°C. Colonies were translucent and approximately 0.2 to 0.3 mm in diameter on 5% sheep blood-enriched Columbia agar (BioMerieux). Growth of the strain was tested on the same agar under anaerobic and microaerophilic conditions using GENbag anaer and GENbag microaer systems, respectively (BioMerieux), and in aerobic conditions, with or without 5% CO\textsubscript{2}. Growth was observed only anaerobically. No growth occurred in aerobic or microaerophilic conditions. Gram staining showed Gram-positive rods able to form spores (Figure 2). A motility test was positive. Cells grown on agar exhibit a mean diameter of 1 µm and a mean length of 1.22 µm in electron microscopy (Figure 3).

Strain JCD\textsuperscript{T} exhibited neither catalase nor oxidase activity (Table 2). Using an API Rapid ID 32A strip (BioMerieux), positive reactions were obtained for indole production, alkaline phosphatase, arginine arylamidase, proline arylamidase, alanine arylamidase, glycine arylamidase, histidine arylamidase, glutamyl glutamic acid arylamidase and serine arylamidase. Negative reactions were obtained for arginine dihydrolase, α-galactosidase, β-galactosidase, α-glucosidase, β-glucosidase, α-arabinosidase, N-acetyl-β-glucosaminidase, glutamic acid decarboxylase, α-fucosidase, nitrate reduction, leucyl glycine arylamidase, fermentation of mannose and raffinose, urease, β-galactosidase-6-phosphatase, β-glucuronidase, phenylalanine arylamidase, leucine arylamidase, pyroglutamic acid arylamidase and tyrosine arylamidase. Using an API 50CH strip (Biomerieux), strain JCD\textsuperscript{T} was asaccharolytic.
Clostridium jeddahense

*C. jeddahense* is susceptible to amoxicillin, amoxicillin-clavulanate, imipenem, metronidazole, doxycycline, rifampicin, vancomycin but resistant to ceftriaxone, ciprofloxacin and trimethoprim-sulfamethoxazole. The comparisons with other *Clostridium* species are summarized in Table 2.

![Figure 2. Gram stain of *Clostridium jeddahense* strain JCD](image1)

![Figure 3. Transmission electron micrograph of *C. jeddahense* strain JCD, taken using a Morgani 268D (Philips) at an operating voltage of 60kV. The scale bar represents 500 nm.](image2)
### Table 2: Differential characteristics of *Clostridium jeddahense* JCD\(^7\), *C. senegalense* JC122 [11], *C. dakarense* FF1 [34], *C. beijerinckii* NCIMB 8052, *C. difficile* B1, *C. cellulyticum* H10, *C. leptum* DSM 753 and *C. sporosphaeroides* DSM 1294 [53].

<table>
<thead>
<tr>
<th>Properties</th>
<th><em>C. jeddahense</em></th>
<th><em>C. sporosphaeroides</em></th>
<th><em>C. cellulyticum</em></th>
<th><em>C. leptum</em></th>
<th><em>C. senegalense</em></th>
<th><em>C. dakarense</em></th>
<th><em>C. beijerinckii</em></th>
<th><em>C. difficile</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell diameter (μm)</td>
<td>1.0</td>
<td>0.5-0.6</td>
<td>1.5</td>
<td>0.6-0.8</td>
<td>1.1</td>
<td>1.2</td>
<td>1.7</td>
<td>3.0</td>
</tr>
<tr>
<td>Oxygen requirement</td>
<td>Strictly anaerobic</td>
<td>Strictly anaerobic</td>
<td>Strictly anaerobic</td>
<td>Strictly anaerobic</td>
<td>Strictly anaerobic</td>
<td>Strictly anaerobic</td>
<td>Strictly anaerobic</td>
<td>Strictly anaerobic</td>
</tr>
<tr>
<td>Gram stain</td>
<td>Positive Motile</td>
<td>Positive Motile</td>
<td>Positive Motile</td>
<td>Positive Motile</td>
<td>Positive Motile</td>
<td>Positive Motile</td>
<td>Positive Motile</td>
<td>Positive Motile</td>
</tr>
<tr>
<td>Motility</td>
<td>+</td>
<td>Na</td>
<td>Na</td>
<td>-</td>
<td>+</td>
<td>Na</td>
<td>+</td>
<td>Na</td>
</tr>
<tr>
<td>Endospore formation</td>
<td>+</td>
<td>-</td>
<td>Na</td>
<td>-</td>
<td>Na</td>
<td>Na</td>
<td>Na</td>
<td>Na</td>
</tr>
<tr>
<td>Production of</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alkaline phosphatase</td>
<td>+</td>
<td>Na</td>
<td>Na</td>
<td>-</td>
<td>+</td>
<td>Na</td>
<td>Na</td>
<td>Na</td>
</tr>
<tr>
<td>Catalase</td>
<td>-</td>
<td>Na</td>
<td>Na</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Na</td>
</tr>
<tr>
<td>Oxidase</td>
<td>-</td>
<td>Na</td>
<td>Na</td>
<td>-</td>
<td>-</td>
<td>Na</td>
<td>Na</td>
<td>Na</td>
</tr>
<tr>
<td>Nitrate reductase</td>
<td>-</td>
<td>-</td>
<td>Na</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Urease</td>
<td>-</td>
<td>-</td>
<td>Na</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>Na</td>
</tr>
<tr>
<td>β-galactosidase</td>
<td>-</td>
<td>Na</td>
<td>Na</td>
<td>-</td>
<td>Na</td>
<td>Na</td>
<td>Na</td>
<td>Na</td>
</tr>
<tr>
<td>N-acetyl-glucosamine</td>
<td>-</td>
<td>Na</td>
<td>Na</td>
<td>+</td>
<td>Na</td>
<td>Na</td>
<td>Na</td>
<td>Na</td>
</tr>
<tr>
<td>Acid from</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L-Arabinose</td>
<td>-</td>
<td>-</td>
<td>Na</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ribose</td>
<td>-</td>
<td>-</td>
<td>Na</td>
<td>W</td>
<td>Na</td>
<td>Na</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Mannose</td>
<td>-</td>
<td>-</td>
<td>Na</td>
<td>-</td>
<td>Na</td>
<td>Na</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Mannitol</td>
<td>-</td>
<td>-</td>
<td>Na</td>
<td>-</td>
<td>Na</td>
<td>Na</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Sucrose</td>
<td>-</td>
<td>-Na</td>
<td>Na</td>
<td>W</td>
<td>Na</td>
<td>Na</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>D-glucose</td>
<td>-</td>
<td>Na</td>
<td>Na</td>
<td>Na</td>
<td>+</td>
<td>Na</td>
<td>Na</td>
<td>Na</td>
</tr>
<tr>
<td>D-fructose</td>
<td>-</td>
<td>Na</td>
<td>Na</td>
<td>Na</td>
<td>+</td>
<td>Na</td>
<td>Na</td>
<td>Na</td>
</tr>
<tr>
<td>D-maltose</td>
<td>-</td>
<td>Na</td>
<td>Na</td>
<td>Na</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>D-lactose</td>
<td>-</td>
<td>Na</td>
<td>Na</td>
<td>Na</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>G+C content (%)</td>
<td>52</td>
<td>41</td>
<td>51</td>
<td>26.8</td>
<td>27.98</td>
<td>28</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>Habitat</td>
<td>Human gut</td>
<td>Environment</td>
<td>Compost</td>
<td>Human gut</td>
<td>Human gut</td>
<td>Human gut</td>
<td>Human gut</td>
<td>Human gut</td>
</tr>
</tbody>
</table>

\(^7\) (Na = data not available; w = weak, v = variable reaction)
Matrix-assisted laser-desorption/ionization time-of-flight (MALDI-TOF) MS protein analysis was carried out as previously described [54]. Briefly, a pipette tip was used to pick one isolated bacterial colony from a culture agar plate and spread it as a thin film on a MTP 384 MALDI-TOF target plate (Bruker Daltonics, Leipzig, Germany). Twelve distinct deposits from twelve isolated colonies were performed for strain JCD\textsuperscript{T}. Each smear was overlaid with 2 µL of matrix solution (saturated solution of alpha-cyano-4-hydroxycinnamic acid) in 50% acetonitrile, 2.5% tri-fluoracetic acid, and allowed to dry for 5 minutes. Measurements were performed with a Microflex spectrometer (Bruker). Spectra were recorded in the positive linear mode for the mass range of 2,000 to 20,000 Da (parameter settings: ion source 1 (IS1), 20kV; IS2, 18.5 kV; lens, 7 kV). A spectrum was obtained after 675 shots with variable laser power. The time of acquisition was between 30 seconds and 1 minute per spot. The twelve JCD\textsuperscript{T} spectra were imported into the MALDI BioTyper software (version 2.0, Bruker) and analyzed by standard pattern matching (with default parameter settings) against the main spectra of 3,769 bacteria, including 228 spectra from 96 \textit{Clostridium} species. The method of identification included the m/z from 3,000 to 15,000 Da. For every spectrum, a maximum of 100 peaks were compared with spectra in database. The resulting score enabled the identification of tested species, or not: a score \(\geq 2\) with a validly published species enabled identification at the species level, a score \(\leq 1.7\) but \(< 2\) enabled identification at the genus level, and a score \(< 1.7\) did not enable any identification. No significant MALDI-TOF score was obtained for strain JCD\textsuperscript{T} against the Bruker database, suggesting that our isolate was not a member of a known species. We added the spectrum from strain JCD\textsuperscript{T} to our database (Figure 4). Finally, the gel view showed the spectral differences with other members of the genus \textit{Clostridium} (Figure 5).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4.png}
\caption{Reference mass spectrum from \textit{C. jeddahense} strain JCD\textsuperscript{T}. Spectra from 12 individual colonies were compared and a reference spectrum was generated.}
\end{figure}
Figure 5: Gel view comparing *C. jeddahense* strain JCD\(^1\) to other *Clostridium* species. The gel view displays the raw spectra of loaded spectrum files arranged as a pseudo-electrophoretic gel. The x-axis records the m/z value. The left y-axis displays the running spectrum number originating from subsequent spectra loading. The peak intensity is expressed by a grey scale scheme code. The grey scale bar on the right y-axis indicates the relation between the shade of grey a peak is displayed with and the peak intensity in arbitrary units. Species names are shown on the left.

---

**Genome sequencing information**

**Genome project history**

The organism was selected for sequencing on the basis of its phylogenetic position and 16S rDNA similarity to members of the genus *Clostridium*, and is part of a study of the human digestive flora aiming at isolating all bacterial species in human feces [55]. It was the 101\(^{st}\) genome of a *Clostridium* species and the first genome of *C. jeddahense* sp. nov. The GenBank accession number is CBYL00000000. The assembly consists of 104 contigs. Table 3 shows the project information and its association with MIGS version 2.0 compliance [39].

---

**Table 3. Project information**

<table>
<thead>
<tr>
<th>MIGS ID</th>
<th>Property</th>
<th>Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIGS-31</td>
<td>Finishing quality</td>
<td>High-quality draft</td>
</tr>
<tr>
<td>MIGS-28</td>
<td>Libraries used</td>
<td>Paired end and Mate pair</td>
</tr>
<tr>
<td>MIGS-29</td>
<td>Sequencing platform</td>
<td>MySeq Illumina</td>
</tr>
<tr>
<td>MIGS-31.2</td>
<td>Fold coverage</td>
<td>94.91×</td>
</tr>
<tr>
<td>MIGS-30</td>
<td>Assemblers</td>
<td>Newbler</td>
</tr>
<tr>
<td>MIGS-32</td>
<td>Gene calling method</td>
<td>PRODIGAL</td>
</tr>
<tr>
<td></td>
<td>Genbank Date of Release</td>
<td>February 12, 2014</td>
</tr>
<tr>
<td></td>
<td>Genbank project ID</td>
<td>CBYL00000000</td>
</tr>
<tr>
<td>MIGS-13</td>
<td>Project relevance</td>
<td>Study of the human gut microbiome</td>
</tr>
</tbody>
</table>
**Growth conditions and DNA isolation**

*C. jeddahense* sp. nov., strain JCD (= CSUR P693 = DSM 27834) was grown on 5% sheep blood-enriched Columbia agar (BioMerieux) at 37°C in anaerobic atmosphere. Bacteria grown on three Petri dishes were harvested and resuspended in 4x100µL of TE buffer. Then, 200 µL of this suspension was diluted in 1ml TE buffer for lysis treatment that included a 30- minute incubation with 2.5 µg/µL lysozyme at 37°C, followed by an overnight incubation with 20 µg/µL proteinase K at 37°C. Extracted DNA was then purified using 3 successive phenol-chloroform extractions and ethanol precipitation at -20°C overnight. After centrifugation, the DNA was resuspended in 160 µL TE buffer.

**Genome sequencing and assembly**

Genomic DNA of *Clostridium jeddahense* was sequenced on a MiSeq sequencer (Illumina, Inc, San Diego CA 92121, USA) with 2 applications: paired end and mate pair. The paired end and the mate pair strategies were barcoded in order to be mixed respectively with 14 other genomic projects constructed according the Nextera XT library kit (Illumina) and 11 others projects with the nextera Mate pair kit (Illumina).

The gDNA was quantified by a Qubit assay with the high sensitivity kit (Life technologies, Carlsbad, CA, USA) to 11.1 ng/µL and dilution was performed such that 1ng of each strain's gDNA was used to construct the paired end library. The “tagmentation” step fragmented and tagged the DNA. Then limited cycle PCR amplification completed the tag adapters and introduced dual-index barcodes. After purification on Ampure beads (Life Technologies, Carlsbad, CA, USA), the libraries were normalized on specific beads according to the Nextera XT protocol (Illumina). Normalized libraries are pooled into a single library for sequencing on the MiSeq. The pooled single strand library was loaded onto the reagent cartridge and then onto the instrument along with the flow cell. Automated cluster generation and sequencing run was performed in a single 39-hour run at a 2x250 bp read length. Total information of 3.9 Gb was obtained from a 399 K/mm2 density with 97.9% (7,840,000 clusters) of the clusters passing quality control (QC) filters. Within this pooled run, the index representation for *Clostridium jeddahense* was determined to be 6.54%.

From this genome sequencing process, the 501,426 produced Illumina reads for *Clostridium jeddahense* were filtered according to the read qualities.

**Genome annotation**

Open Reading Frames (ORFs) were predicted using Prodigal [56] with default parameters. However, the predicted ORFs were excluded if they spanned a sequencing gap region. The predicted bacterial protein sequences were searched against the GenBank [57] and Clusters of Orthologous Groups (COG) databases using BLASTP. The tRNAs and rRNAs were predicted using the tRNAscan-SE [58] and RNAmmer [59] tools, respectively. Signal peptides and numbers of transmembrane helices were predicted using SignalP [60] and TMHMM [61], respectively. Mobile genetic elements were predicted using PHAST [62] and RAST [63]. ORFans were identified if their BLASTP E-value was
was lower than 1e-03 for alignment length greater than 80 amino acids. If alignment lengths were smaller than 80 amino acids, we used an E-value of 1e-05. Such parameter thresholds have already been used in previous work to define ORFans. Artemis [64] and DNA Plotter [65] were used for data management and visualization of genomic features, respectively. The Mauve alignment tool (version 2.3.1) was used for multiple genomic sequence alignment [66].

To estimate the mean level of nucleotide sequence similarity at the genome level between \textit{C. jeddahense} and 7 other members of the genus \textit{Clostridium}, we used the Average Genomic Identity Of gene Sequences (AGIOS) home-made software [6]. Briefly, this software combines the Proteinortho software [67] for detecting orthologous proteins between pairs of genomes, then retrieves the corresponding genes and determines the mean percentage of nucleotide sequence identity among orthologous ORFs using the Needleman-Wunsch global alignment algorithm. \textit{C. jeddahense} strain JCD\textsuperscript{T} was compared to \textit{C. senegalense} strain JC122, \textit{C. dakarense} strain FF1, \textit{Clostridium beijerinckii} strain NCIMB 8052, \textit{C. difficile} strain B1, \textit{Clostridium cellulolyticum} strain H10, \textit{Clostridium leptum} strain DSM 753, and \textit{Clostridium sporosphaeroides} strain DSM 1294 (see Table 6B).

**Genome properties**

The genome is 3,613,503 bp long (1 chromosome, but no plasmid) with a 51.95% G+C content (Figure 6 and Table 4). Of the 3,515 predicted genes, 3,462 were protein-coding genes and 53 were RNAs, including 4 rRNAs. A total of 2,193 genes (62.38%) were assigned a putative function and 81 genes were identified as ORFans (2.3%). The properties and statistics of the genome are summarized in Tables 4 and 5. The distribution of genes into COG functional categories is presented in Table 5.

---

**Figure 6.** Graphical circular map of the chromosome. From the outside in: open reading frames oriented in the forward (colored by COG categories) direction, open reading frames oriented in the reverse (colored by COG categories) direction, RNA operon (red), and tRNAs (green), GC content plot, and GC skew (purple: negative values, olive: positive values).
Table 4. Nucleotide content and gene count levels of the genome

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Value</th>
<th>% of total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genome size (bp)</td>
<td>3,613,503</td>
<td></td>
</tr>
<tr>
<td>DNA G+C content (bp)</td>
<td>1,877,214</td>
<td>51.95</td>
</tr>
<tr>
<td>DNA Coding region (bp)</td>
<td>3,152,277</td>
<td>87.23</td>
</tr>
<tr>
<td>Number of replicons</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Extra chromosomal element</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Total genes</td>
<td>3,515</td>
<td>100</td>
</tr>
<tr>
<td>RNA genes</td>
<td>53</td>
<td>1.51</td>
</tr>
<tr>
<td>Protein-coding genes</td>
<td>3,462</td>
<td>98.49</td>
</tr>
<tr>
<td>Genes with function prediction</td>
<td>2,193</td>
<td>61.02</td>
</tr>
<tr>
<td>Genes assigned to COGs</td>
<td>2,515</td>
<td>71.55</td>
</tr>
<tr>
<td>Genes with peptide signals</td>
<td>135</td>
<td>3.84</td>
</tr>
<tr>
<td>Genes with transmembrane helices</td>
<td>887</td>
<td>25.23</td>
</tr>
</tbody>
</table>

* The total is based on either the size of the genome in base pairs or the total number of protein-coding genes in the annotated genome.

Table 5. Number of genes associated with the 25 general COG functional categories

<table>
<thead>
<tr>
<th>Code</th>
<th>Value</th>
<th>% age*</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>J</td>
<td>154</td>
<td>4.45</td>
<td>Translation</td>
</tr>
<tr>
<td>A</td>
<td>0</td>
<td>0</td>
<td>RNA processing and modification</td>
</tr>
<tr>
<td>K</td>
<td>296</td>
<td>8.55</td>
<td>Transcription</td>
</tr>
<tr>
<td>L</td>
<td>138</td>
<td>3.98</td>
<td>Replication, recombination and repair</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>0.03</td>
<td>Chromatin structure and dynamics</td>
</tr>
<tr>
<td>D</td>
<td>24</td>
<td>0.69</td>
<td>Cell cycle control, mitosis and meiosis</td>
</tr>
<tr>
<td>Y</td>
<td>0</td>
<td>0</td>
<td>Nuclear structure</td>
</tr>
<tr>
<td>V</td>
<td>73</td>
<td>2.11</td>
<td>Defense mechanisms</td>
</tr>
<tr>
<td>T</td>
<td>156</td>
<td>4.5</td>
<td>Signal transduction mechanisms</td>
</tr>
<tr>
<td>M</td>
<td>116</td>
<td>3.35</td>
<td>Cell wall/membrane biogenesis</td>
</tr>
<tr>
<td>N</td>
<td>62</td>
<td>1.79</td>
<td>Cell motility</td>
</tr>
<tr>
<td>Z</td>
<td>0</td>
<td>0</td>
<td>Cytoskeleton</td>
</tr>
<tr>
<td>W</td>
<td>0</td>
<td>0</td>
<td>Extracellular structures</td>
</tr>
<tr>
<td>U</td>
<td>48</td>
<td>1.38</td>
<td>Intracellular trafficking and secretion</td>
</tr>
<tr>
<td>O</td>
<td>66</td>
<td>1.9</td>
<td>Posttranslational modification, protein turnover, chaperones</td>
</tr>
<tr>
<td>C</td>
<td>154</td>
<td>4.45</td>
<td>Energy production and conversion</td>
</tr>
<tr>
<td>G</td>
<td>237</td>
<td>6.84</td>
<td>Carbohydrate transport and metabolism</td>
</tr>
<tr>
<td>E</td>
<td>328</td>
<td>9.47</td>
<td>Amino acid transport and metabolism</td>
</tr>
<tr>
<td>F</td>
<td>56</td>
<td>1.61</td>
<td>Nucleotide transport and metabolism</td>
</tr>
<tr>
<td>H</td>
<td>92</td>
<td>2.66</td>
<td>Coenzyme transport and metabolism</td>
</tr>
<tr>
<td>I</td>
<td>85</td>
<td>2.45</td>
<td>Lipid transport and metabolism</td>
</tr>
<tr>
<td>P</td>
<td>164</td>
<td>4.74</td>
<td>Inorganic ion transport and metabolism</td>
</tr>
<tr>
<td>Q</td>
<td>53</td>
<td>1.53</td>
<td>Secondary metabolites biosynthesis, transport and catabolism</td>
</tr>
<tr>
<td>R</td>
<td>346</td>
<td>10</td>
<td>General function prediction only</td>
</tr>
<tr>
<td>S</td>
<td>195</td>
<td>5.63</td>
<td>Function unknown</td>
</tr>
<tr>
<td></td>
<td>-</td>
<td>947</td>
<td>27.35 Not in COGs</td>
</tr>
</tbody>
</table>

* The total is based on the total number of protein-coding genes in the annotated genome.
Genome comparison with other *Clostridium* genomes

We compared the genomes of *C. jeddahense* JCD\textsuperscript{T}, *C. sporosphaeroides* DSM 1294, *C. leptum* DSM 753, *C. beijerincki* NCIMB 8052, *C. cellulolyticum* H10, *C. difficile* B1, *C. senegalense* DSM 25507, *C. dakarense* DSM 27086 (Table 6A).

The draft genome of *C. jeddahense* (3.61 Mb) is larger than *C. sporosphaeroides* and *C. leptum* (3.17 and 3.27 Mb respectively) but smaller than *C. beijerincki*, *C. cellulolyticum*, *C. difficile*, *C. senegalense* and *C. dakarense* (6.0, 4.07, 4.46, 3.89, 3.73 Mb respectively). It exhibits a higher G+C content than all other compared genome except *C. sporosphaeroides* (53.5%). *C. jeddahense* has a higher gene content (3,462) than *C. sporosphaeroides*, *C. difficile* (2,951 and 3,390 respectively) but smaller than *C. leptum*, *C. beijerincki*, *C. cellulolyticum*, *C. senegalense* and *C. dakarense* (3,591, 3,818, 3,923, 3,704, 5,020 respectively). *C. jeddahense* shared 1,573, 876, 816, 847, 1,030, 770 and 1,044 orthologous genes with *C. sporosphaeroides*, *C. cellulolyticum*, *C. dakarense*, *C. difficile*, *C. leptum*, *C. senegalense* and *C. beijerincki* respectively.

When we compared *C. jeddahense* with other species, AGIOS values ranged from 57.52 with *C. senegalense* to 91.97% with *C. sporosphaeroides*. Although the AGIOS value was elevated between *C. jeddahense* and *C. sporosphaeroides*, we believe that the remarkable phenotypic differences, including motility, indole production (Table 2), and protein profile (Figure 7), enable the classification of *C. jeddahense* as a new species.

<table>
<thead>
<tr>
<th>Species</th>
<th>Strain</th>
<th>Genome accession number</th>
<th>Genome size (Mb)</th>
<th>G+C content</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>C. jeddahense</em></td>
<td>JCD\textsuperscript{T}</td>
<td>CBYL000000000</td>
<td>3.61</td>
<td>51.95</td>
</tr>
<tr>
<td><em>C. sporosphaeroides</em></td>
<td>DSM 1294</td>
<td>ARTA010000000</td>
<td>3.17</td>
<td>53.5</td>
</tr>
<tr>
<td><em>C. cellulolyticum</em></td>
<td>H10</td>
<td>NC_011898</td>
<td>4.07</td>
<td>37.4</td>
</tr>
<tr>
<td><em>C. dakarense</em></td>
<td>DSM 27086</td>
<td>CBTZ010000000</td>
<td>3.73</td>
<td>27.98</td>
</tr>
<tr>
<td><em>C. difficile</em></td>
<td>B1</td>
<td>NC_017179</td>
<td>4.46</td>
<td>28.4</td>
</tr>
<tr>
<td><em>C. leptum</em></td>
<td>DSM 753</td>
<td>ABCB020000000</td>
<td>3.27</td>
<td>50.2</td>
</tr>
<tr>
<td><em>C. senegalense</em></td>
<td>DSM 25507</td>
<td>CAEV01000001</td>
<td>3.89</td>
<td>26.8</td>
</tr>
<tr>
<td><em>C. beijerincki</em></td>
<td>NCIMB 8052</td>
<td>NC_009617</td>
<td>6.0</td>
<td>29.0</td>
</tr>
</tbody>
</table>

\textsuperscript{A}: Species, Strain, GenBank accession number, genome size and G+C content of all compared genomes
Table 6B. Genomic comparison of *C. jeddahense* with 7 other *Clostridium* species†

<table>
<thead>
<tr>
<th></th>
<th><em>C. jeddahense</em></th>
<th><em>C. sporosphaeroides</em></th>
<th><em>C. cellulolyticum</em></th>
<th><em>C. dakarense</em></th>
<th><em>C. difficile</em></th>
<th><em>C. leptum</em></th>
<th><em>C. senegalense</em></th>
<th><em>C. beijerincki</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>C. jeddahense</em></td>
<td>3,462</td>
<td>1,573</td>
<td>876</td>
<td>816</td>
<td>847</td>
<td>1,030</td>
<td>770</td>
<td>1,044</td>
</tr>
<tr>
<td><em>C. sporosphaeroides</em></td>
<td>91.97</td>
<td>2,951</td>
<td>854</td>
<td>776</td>
<td>819</td>
<td>1,016</td>
<td>745</td>
<td>1,015</td>
</tr>
<tr>
<td><em>C. cellulolyticum</em></td>
<td>61.60</td>
<td>60.62</td>
<td>3,923</td>
<td>806</td>
<td>851</td>
<td>754</td>
<td>814</td>
<td>946</td>
</tr>
<tr>
<td><em>C. dakarense</em></td>
<td>57.30</td>
<td>56.34</td>
<td>65.70</td>
<td>5,020</td>
<td>1,271</td>
<td>665</td>
<td>1,110</td>
<td>1,142</td>
</tr>
<tr>
<td><em>C. difficile</em></td>
<td>57.56</td>
<td>56.80</td>
<td>65.65</td>
<td>77.74</td>
<td>3,390</td>
<td>714</td>
<td>1,098</td>
<td>1,171</td>
</tr>
<tr>
<td><em>C. leptum</em></td>
<td>67.98</td>
<td>68.07</td>
<td>61.94</td>
<td>58.55</td>
<td>58.84</td>
<td>3,591</td>
<td>651</td>
<td>780</td>
</tr>
<tr>
<td><em>C. senegalense</em></td>
<td>57.52</td>
<td>56.73</td>
<td>65.71</td>
<td>70.18</td>
<td>69.41</td>
<td>58.72</td>
<td>3,704</td>
<td>1,125</td>
</tr>
<tr>
<td><em>C. beijerincki</em></td>
<td>58.63</td>
<td>57.95</td>
<td>65.93</td>
<td>68.98</td>
<td>68.48</td>
<td>59.58</td>
<td>71.37</td>
<td>3,818</td>
</tr>
</tbody>
</table>

†Numbers of orthologous proteins shared between genomes (above diagonal), AGIOS values (below diagonal) and numbers of proteins per genome (bold numbers)
Figure 7. Distribution of predicted genes of *C. jeddahense* and 7 other *Clostridium* species into COG categories. *C.jdm* = *C. jeddahense*, *C.spo* = *C. sporosphaeroides*, *C. lep* = *C. leptum*, *C.bej* = *C. beijerinckii*, *C. cel* = *C. cellulolyticum*, *C. diff* = *C. difficile*, *C. sen* = *C. senegalense*, *C. dak* = *C. dakarensense*.

**Conclusion**

On the basis of phenotypic, phylogenetic and genomic analyses (taxonomics), we formally propose the creation of *Clostridium jeddahense* sp. nov. that contains strain JCD\(^T\). This strain was isolated from the fecal flora of an obese 24-year-old Saudi individual living in Jeddah.

**Description of *C. jeddahense* sp. nov.**

*Clostridium jeddahense* (jed.dah.en’se L.gen. neutr. n. combination of Jeddah, the city in Saudi Arabia where the specimen was obtained from an obese Saudi patient sample.) Transparent colonies were 0.2 to 0.3 mm in diameter on blood-enriched agar. *C. jeddahense* is a Gram-positive, obligate anaerobic, endospore-forming bacterium with a mean diameter of 1 µm. Optimal growth on axenic medium was observed at 37°C.

*C. jeddahense* is catalase negative and oxidase negative. Alkaline phosphatase, arginine arylamidase, proline arylamidase, alanine arylamidase, glycine arylamidase, histidine arylamidase, glutamyl glutamic acid arylamidase and serine arylamidase activities were positive. Arginine dihydrolase, α-galactosidase, β-galactosidase, α-glucosidase, β-glucosidase, α-arabinosidase, N-acetyl-β-glucosaminidase, glutamic acid decarboxylase, α-fucosidase, reduction of nitrate, leucyl glycine arylamidase, fermentation of mannose and raffinose, urease, β-galactosidase-6-phosphatase, β-glucuronidase, phenylalanine arylamidase, leucine arylamidase, pyrog glutamic acid arylamidase and tyrosine arylamidase activities were negative. Asaccharolytic. Positive for indole. Cells are susceptible to amoxicillin, amoxicillin-clavulanate, imipenem, metronidazole, doxycycline, rifampicin, vancomycin but resistant to ceftriaxone, ciprofloxacin and trimethoprim-sulfamethoxazole.

The G+C content of the genome is 51.95%. The 16S rDNA and genome sequences are deposited in GenBank under accession numbers HG726040 and CBYL00000000, respectively. The type strain is JCD\(^T\) (= CSUR P693 = DSM 27834).

http://standardsingenomics.org
Acknowledgements

This work was funded by the Deanship of Scientific Research (DSR), King Abdulaziz University, under grant No. (1-141/1433 HiCi). The authors, therefore, acknowledge technical and financial support of KAU. The authors thank the Xegen Company (www.xegen.fr) for automating the genome annotation process.

References


16. Ramasamy D, Kokcha S, Lagier JC, N’Guyen TT, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of


44. List of new names and new combinations previously effectively, but not validly, published. List no. 132. *Int J Syst Evol Microbiol* 2010; 60:469-472. [http://dx.doi.org/10.1099/ijs.0.022855-0](http://dx.doi.org/10.1099/ijs.0.022855-0)


