



EXTENDED GENOME REPORT

Open Access



# Non contiguous-finished genome sequence and description of *Bacillus jeddahensis* sp. nov.

Fadi Bittar<sup>1\*</sup>, Fehmida Bibi<sup>2</sup>, Dhamodharan Ramasamy<sup>1</sup>, Jean-Christophe Lagier<sup>1</sup>, Esam I. Azhar<sup>2,3</sup>, Asif A. Jiman-Fatani<sup>4</sup>, Ahmed K. Al-Ghamdi<sup>3</sup>, Ti Thien Nguyen<sup>1</sup>, Muhammad Yasir<sup>2</sup>, Pierre-Edouard Fournier<sup>1</sup> and Didier Raoult<sup>1,2</sup>

## Abstract

Strain JCE<sup>T</sup> was isolated from the fecal sample of a 24-year-old obese man living in Jeddah, Saudi Arabia. It is an aerobic, Gram-positive, rod-shaped bacterium. This strain exhibits a 16S rRNA nucleotide sequence similarity of 97.5 % with *Bacillus niacini*, the phylogenetically closest species with standing nomenclature. Moreover, the strain JCE<sup>T</sup> presents many phenotypic differences, when it is compared to other *Bacillus* species, and shows a low MALDI-TOF Mass Spectrometry score that does not allow any identification. Thus, it is likely that this strain represents a new species. Here we describe the features of this organism, together with the complete genome sequence and annotation. The 4,762,944 bp long genome (1 chromosome but no plasmid) contains 4,654 protein-coding and 98 RNAs genes, including 92 tRNA genes. The strain JCE<sup>T</sup> differs from most of the other closely *Bacillus* species by more than 1 % in G + C content. In addition, digital DNA-DNA hybridization values for the genome of the strain JCE<sup>T</sup> against the closest *Bacillus* genomes range between 19.5 to 28.1, that confirming again its new species status. On the basis of these polyphasic data made of phenotypic and genomic analyses, we propose the creation of *Bacillus jeddahensis* sp. nov. that contains the strain JCE<sup>T</sup>.

**Keywords:** *Bacillus jeddahensis*, Genome, Taxono-genomics, Culturomics, Human feces

## Introduction

Currently, a polyphasic approach that combines proteomic by MALDI-TOF spectra analysis, genomic data and phenotypic characterization is used widely to describe new bacterial species [1–13].

The genus *Bacillus*, described by Cohn [14] more than 140 years ago, includes actually 310 species names (296 validly and 14 not-validly published species) [15]. Species belonging to this genus are Gram-positive or variable and mostly motile and spore-forming bacteria. *Bacillus* spp. are ubiquitous bacteria isolated from various environmental sources but it could be involved in human infections [16].

Strain JCE<sup>T</sup> (= CSUR P732 = DSM 28281) is the type strain of *Bacillus jeddahensis* sp. nov. This bacterium is a Gram-positive, flagellated, facultatively anaerobic, indole-negative bacillus that has rounded-ends. It was isolated from the stool sample of a 24-year-old obese man living in Jeddah, Saudi Arabia as part of a culturomics study aiming

at cultivating bacterial species within human feces. By applying large scale of culture conditions, culturomics allowed previously the isolation of many new bacterial species from human stool samples [17–19].

Here we present a summary classification and a set of features for *B. jeddahensis* sp. nov. strain JCE<sup>T</sup> together with the description of the complete genome sequence and annotation. These characteristics support the circumscription of the species *B. jeddahensis* [20].

## Organism information

### Classification and features

In April 2013, a fecal sample was collected from a 24-year-old obese (body mass index 52 kg/m<sup>2</sup>) man living in Jeddah, Saudi Arabia (Table 1). Written assent was obtained from this individual. Both the study and the assent procedure were approved by Ethical Committee of the King Abdulaziz University, King Fahd medical Research Centre, Saudi Arabia (agreement number 014-CEGMR-2-ETH-P) and the Ethical Committee of the Institut Fédératif de Recherche IFR48, Faculty of Medecine, Marseille, France (agreement numbers 09–022 and 11–017). The fecal

\* Correspondence: fadi.bittar@univ-amu.fr

<sup>1</sup>URMITE, Aix-Marseille Université, Faculté de médecine, Marseille, France  
Full list of author information is available at the end of the article

**Table 1** Classification and general features of *Bacillus jeddahensis* strain JCE<sup>T</sup>

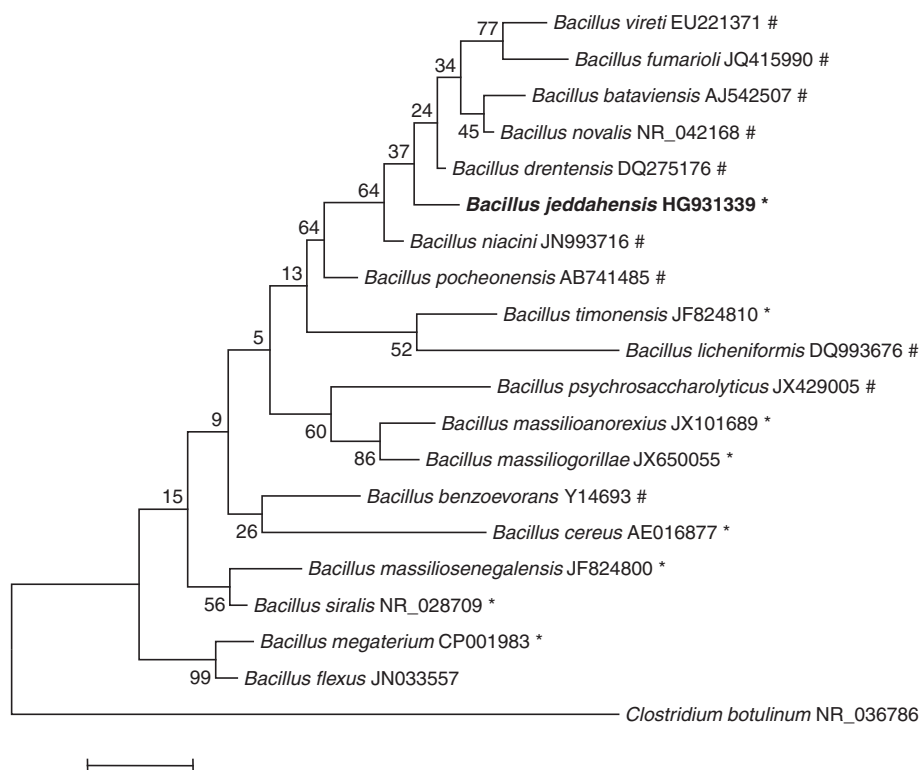
MIGS ID	Property	Term	Evidence code <sup>a</sup>
	Current classification	Domain: <i>Bacteria</i>	TAS [39]
		Phylum: <i>Firmicutes</i>	TAS [40–42]
		Class: <i>Bacilli</i>	TAS [43, 44]
		Order: <i>Bacillales</i>	TAS [45, 46]
		Family: <i>Bacillaceae</i>	TAS [45, 47]
		Genus: <i>Bacillus</i>	TAS [4, 45, 48]
		Species: <i>Bacillus jeddahensis</i>	IDA
		Type strain: JCE <sup>T</sup>	IDA
	Gram stain	Positive	IDA
	Cell shape	Rod-shaped	IDA
	Motility	Non-motile	IDA
	Sporulation	Sporulating	IDA
	Temperature range	Mesophile	IDA
	Optimum temperature	37 °C	IDA
	pH range; Optimum	Not determined	
MIGS-6.3	Salinity	growth in BHI medium + 3 % NaCl	IDA
MIGS-22	Oxygen requirement	Facultative Anaerobic	IDA
	Carbon source	varied (see Additional file 1: Table S1)	IDA
	Energy source	chemoorganoheterotrophic	IDA
MIGS-6	Habitat	Human gut	IDA
MIGS-15	Biotic relationship	Free living	IDA
MIGS-14	Pathogenicity	Unknown	
	Biosafety level	2	NAS
	Isolation	Human faeces	IDA
MIGS-4	Geographic location	Saudi Arabia	IDA
MIGS-5	Sample collection time	July 2013	IDA
MIGS-4.1	Latitude	21° 25' 20.953"	IDA
MIGS-4.1	Longitude	39° 49' 34.262"	IDA
MIGS-4.3	Depth	unknown	
MIGS-4.4	Altitude	unknown	

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 [49]. 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

specimen was preserved at  $-80^{\circ}\text{C}$  after collection and sent to Marseille. Strain JCE<sup>T</sup> (Table 1) was isolated in July 2013 by cultivation on blood culture bottle (Becton Dickinson, Temse, Belgique) supplemented with rumen fluid and sheep blood. This strain exhibited a 97.5 % 16S rRNA nucleotide sequence similarity with *Bacillus niacini*, the phylogenetically closest validly published *Bacillus* species (Fig. 1), when it was compared against NCBI database and Ribosomal Database Project. This value was equal to the percentage of 16S rRNA gene sequence threshold recommended by Meier-Kolthoff et al. for Firmicutes to delineate a new species without carrying

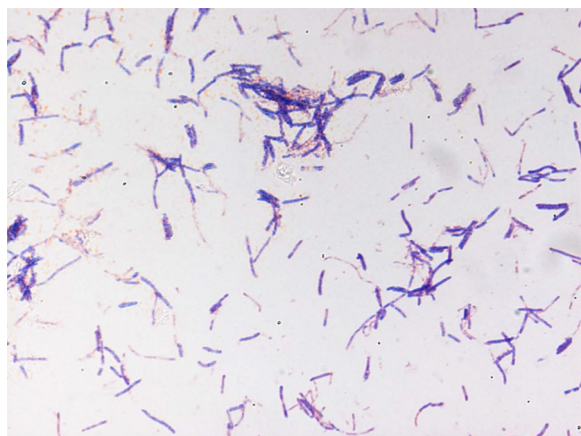
out DNA-DNA hybridization with maximum error probability of 0.01 % [21].

Different growth temperatures (28, 30, 37, 45, 56 °C) were tested. Growth occurred for the temperatures (28–45 °C), but the optimal growth was observed at 37 °C. Colonies were 0.4–0.5 mm in diameter on Columbia agar, appear smooth and grey in color at 37 °C. Growth of the strain was tested under anaerobic and microaerophilic conditions using GENbag anaer and GENbag microaer systems, respectively (BioMérieux), and in aerobic conditions, with or without 5 % CO<sub>2</sub>. Growth was achieved under aerobic (with and without CO<sub>2</sub>), microaerophilic



**Fig. 1** Phylogenetic tree highlighting the position of *Bacillus jeddahensis* strain JCE<sup>T</sup> relative to other type strains within the *Bacillus* genus. GenBank accession numbers are indicated in parentheses. Sequences were aligned using MUSCLE, and phylogenetic inferences obtained using the maximum-likelihood method and Kimura 2-parameter model within the MEGA 6 software [50]. Numbers at the nodes are percentages of bootstrap values obtained by repeating the analysis 1,000 times to generate a majority consensus tree. *Clostridium botulinum* was used as outgroup. The scale bar represents a rate of substitution per site of 0.01. \*Indicates the strains used in the tree have a sequenced genome. # indicates that a sequenced genome is available for this species but not for the strain used to build the tree

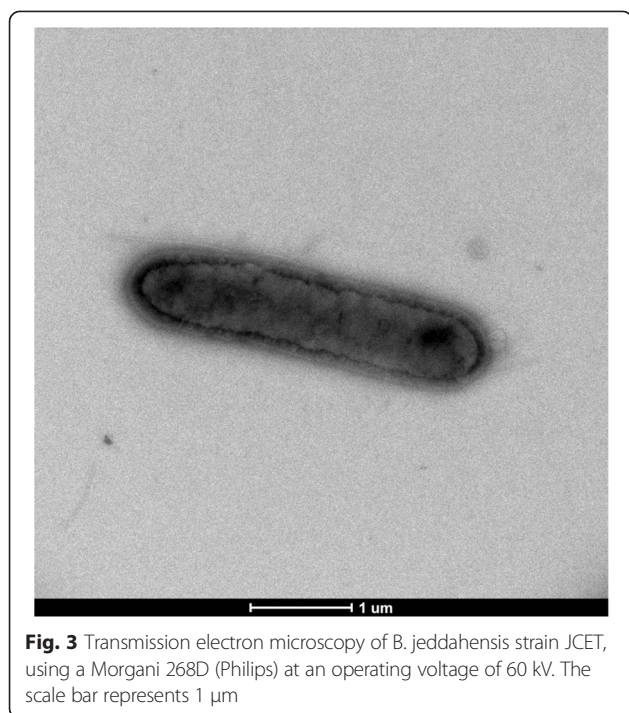
and anaerobic conditions. Gram staining showed Gram positive bacilli (Fig. 2). A motility test was negative. Cells grown on agar sporulate and the rods have a length ranging from 3.83 to 4.71  $\mu\text{m}$  (mean 4.14  $\mu\text{m}$ ) and a diameter ranging from 0.75 to 0.95  $\mu\text{m}$  (mean 0.87  $\mu\text{m}$ ). Both the



**Fig. 2** Gram staining of *B. jeddahensis* strain JCE<sup>T</sup>

length and the diameter were determined by negative staining transmission electron microscopy (Fig. 3).

Strain JCE<sup>T</sup> exhibited oxidase activity but not catalase activity. Using API 50CH system (BioMerieux), a positive reaction was observed for D-arabinose, L-arabinose, D-xylose, D-glucose, D-fructose, D-mannose, N-acetylglucosamine, esculin, D-maltose, D-trehalose, and weak reaction for D-melezitose. Negative reactions were observed for the remaining carbohydrate tests (i.e. glycerol, erythritol, D-ribose, L-xylose, D-adonitol, methyl- $\beta$ -D-xylopyranoside, D-galactose, L-sorbose, L-rhamnose, dulcitol, inositol, D-mannitol, D-sorbitol, methyl- $\alpha$ -D-mannopyranoside, methyl- $\alpha$ -D-glucopyranoside, amygdalin, arbutin, salicin, D-cellobiose, D-lactose, D-melibiose, D-saccharose, inulin, D-raffinose, amidon, glycogen, xylitol, gentiobiose, D-turanose, D-lyxose, D-tagatose, D-fucose, L-fucose, D-arabitol, L-arabitol, potassium gluconate, potassium 2-ketogluconate and potassium 5-ketogluconate). Using API ZYM, positive reactions were observed for esterase (C 4), esterase lipase (C 8), acid phosphatase, naphthol-AS-BI-phosphohydrolase and  $\beta$ -glucosidase. Negative reactions were observed for alkaline phosphatase, lipase (C 14), leucine arylamidase, valine arylamidase, cystine



arylamidase, trypsin,  $\alpha$ -chymotrypsin,  $\alpha$ -galactosidase,  $\beta$ -galactosidase,  $\beta$ -glucuronidase,  $\alpha$ -glucosidase, N-acetyl- $\beta$ -glucosaminidase,  $\alpha$ -mannosidase and  $\alpha$ -fucosidase. Using API NE system, nitrates were reduced to nitrites, the urease reaction, indole production, arginine dihydrolase and gelatin hydrolysis were negative, the following carbon sources were assimilated: D-glucose, D-mannose, N-acetylglucosamine and D-maltose, and the following carbon sources were not assimilated: L-arabinose, D-mannitol, potassium gluconate, capric acid, adipic acid, malic acid, trisodium citrate and phenylacetic acid. *B. jeddahensis* is susceptible to imipenem, doxycycline, amoxicillin, amoxicillin-clavulanate and gentamycin, but resistant to metronidazole, trimethoprim/sulfamethoxazole, rifampicin, vancomycin, erythromycin, ceftriaxone, ciprofloxacin and benzylpenicillin.

When compared to other *Bacillus* species [18, 22–24], *B. jeddahensis* sp. nov. strain JCE<sup>T</sup> exhibited the phenotypic differences detailed in Additional file 1: Table S1.

Matrix-assisted laser-desorption/ionization time-of-flight (MALDI-TOF) MS protein analysis was carried out as previously described [2] using a Microflex spectrometer (Bruker Daltonics, Leipzig, Germany). Twelve distinct deposits were done for strain JCE<sup>T</sup> from 12 isolated colonies. The twelve JCE<sup>T</sup> spectra were imported into the MALDI BioTyper software (version 2.0, Bruker) and analyzed by standard pattern matching (with default parameter settings) against 6,335 bacterial spectra including 210 spectra from 110 *Bacillus* species, used

as reference data, in the BioTyper database. Interpretation of scores was as follows: a score  $\geq 2$  enabled the identification at the species level, a score  $\geq 1.7$  but  $< 2$  enabled the identification at the genus level; and a score  $< 1.7$  did not enable any identification (These scores were established by the manufacturer Bruker Daltonics). For strain JCE<sup>T</sup>, the obtained scores ranged from 1.4 to 1.6, thus suggesting that our isolate was not a member of a known species. We incremented our database with the spectrum from strain JCE<sup>T</sup> (Fig. 4). Spectrum differences with other of *Bacillus* species are shown in Fig. 5.

## Genome sequencing information

### Genome project history

On the basis of phenotypic characteristics of this strain and because of the low 16S rRNA similarity to other members of the genus *Bacillus*, it is likely that the strain represents a new species and thus it was chosen for genome sequencing. It was the 348th genome of a *Bacillus* species (Genomes Online Database) and the first genome of *Bacillus jeddahensis* sp. nov. sequenced. A summary of the project information is shown in Table 2. The Genbank accession number is CCAS00000000 (Table 2) and consists of 149 contigs. Table 2 shows the project information and its association with MIGS version 2.0 compliance [25].

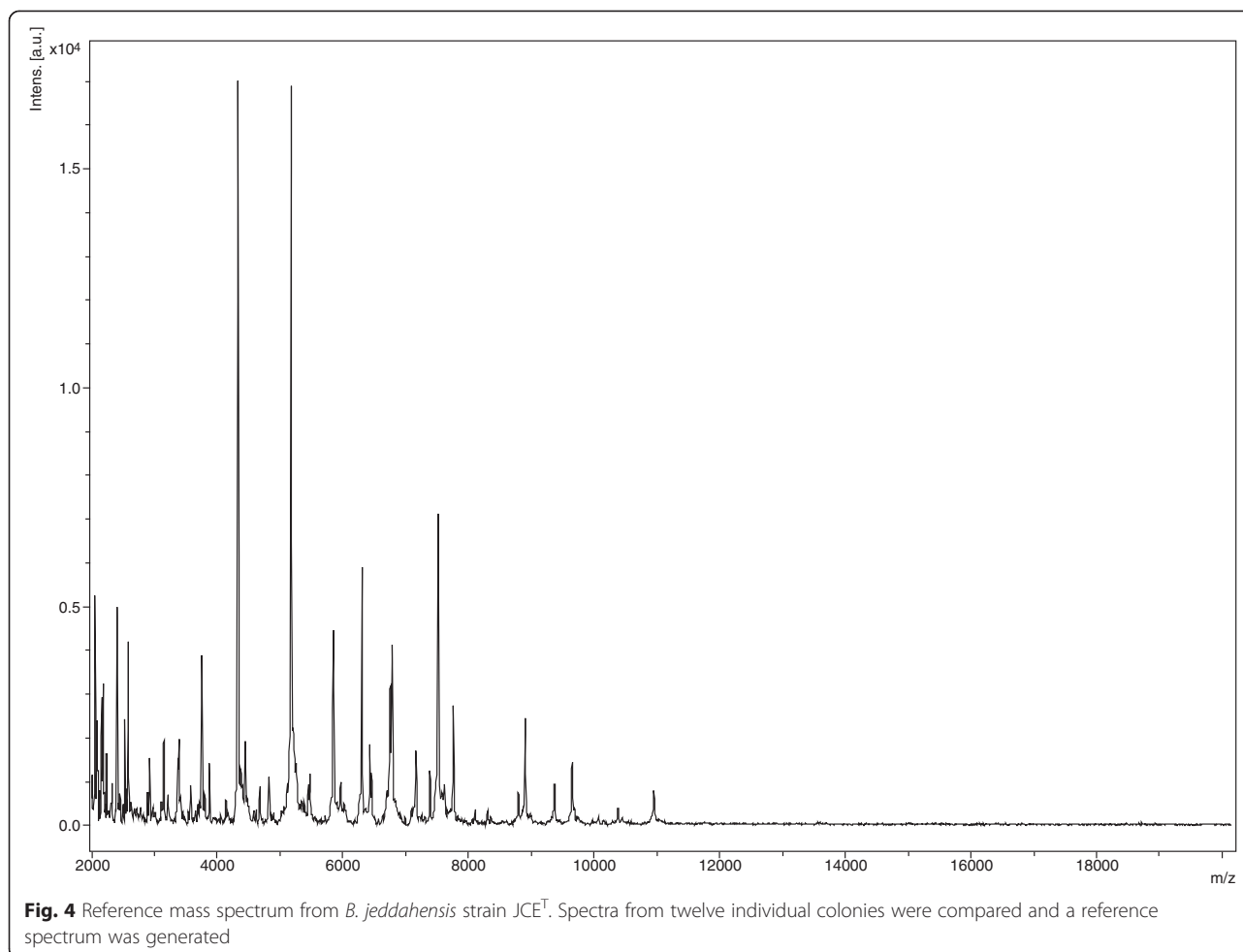
### Growth conditions and genomic DNA preparation

*B. jeddahensis* sp. nov. strain JCE<sup>T</sup>, CSUR P732, DSM 28281, was grown aerobically on 5 % sheep blood-enriched Columbia agar at 37 °C. Four Petri dishes were spread and resuspended in 3  $\times$  500  $\mu$ l of TE buffer and stored at 80 °C. Then, 500  $\mu$ l of this suspension were thawed, centrifuged 3 min at 10,000 rpm and resuspended in 3  $\times$  100  $\mu$ l of G2 buffer (EZ1 DNA Tissue kit, Qiagen). A first mechanical lysis was performed by glass powder on the Fastprep-24 device (Sample Preparation system, MP Biomedicals, USA) using 2  $\times$  20 s cycles. DNA was then treated with 2.5  $\mu$ g/ $\mu$ l lysozyme (30 min at 37 °C) and extracted using the BioRobot EZ1 Advanced XL (Qiagen). The DNA was then concentrated and purified using the Qiasm kit (Qiagen). The yield and the concentration was measured by the Quant-it Picogreen kit (Invitrogen) on the Genios Tecan fluorometer at 50 ng/ $\mu$ l.

### Genome sequencing and assembly

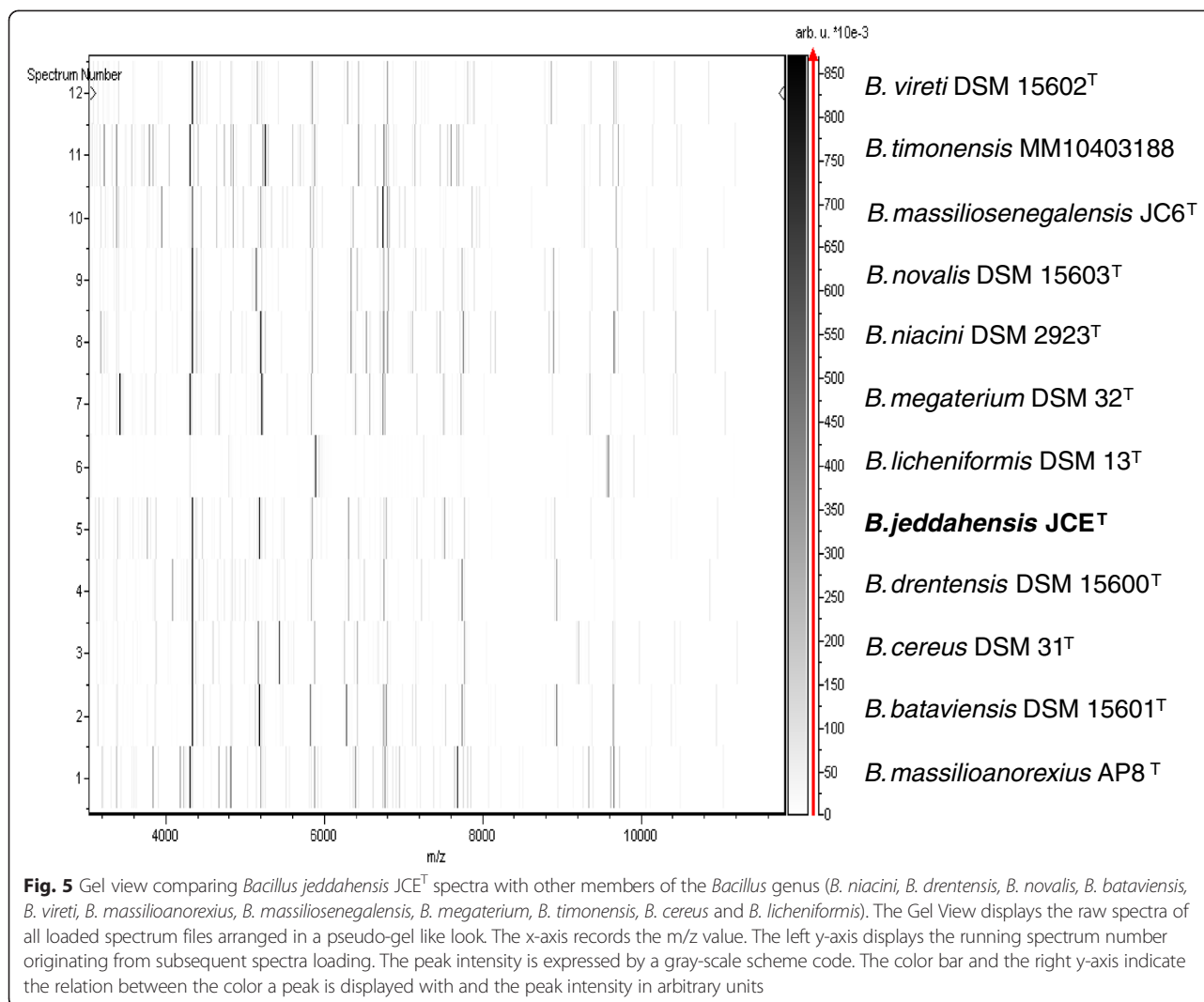
Genomic DNA of *B. jeddahensis* was sequenced on the MiSeq Technology (Illumina Inc, San Diego, CA, USA) with the 2 applications: paired end and mate pair. The paired end and the mate pair strategies were barcoded in order to be mixed respectively with 14 others genomic projects prepared with the Nextera XT DNA sample prep kit (Illumina) and eleven others projects with the Nextera Mate Pair sample prep kit (Illumina). The DNAg was





quantified by a Qubit assay with the high sensitivity kit (Life technologies, Carlsbad, CA, USA) to 16 ng/μl and dilution was performed to require 1ng of each genome as input to prepare the paired end library. The « tagmentation » step fragmented and tagged the DNA. Then limited cycle PCR amplification (twelve cycles) completed the tag adapters and introduced dual-index barcodes. After purification on AMPure XP beads (Beckman Coulter Inc, Fullerton, CA, USA), the libraries were then normalized on specific beads according to the Nextera XT protocol (Illumina). Normalized libraries were 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 paired end sequencing with dual index reads were performed in a single 39-h run in 2 × 250-bp. Total information of 5.3 Gb was obtained from a 574 K/mm<sup>2</sup> cluster density with a cluster passing quality control filters of 95.4 % (11,188,000 clusters). Within this run, the index representation for *B. jeddahensis* was determined to 10.3 %. The 1,062,432 reads were filtered according to the read qualities. The mate pair

library was prepared with 1 μg of genomic DNA using the Nextera mate pair Illumina guide. The genomic DNA sample was simultaneously fragmented and tagged with a mate pair junction adapter. The profile of the fragmentation was validated on an Agilent 2100 BioAnalyzer (Agilent Technologies Inc, Santa Clara, CA, USA) with a DNA 7500 labchip. The DNA fragments ranged in size from 1 kb up to 11 kb with an optimal size at 5 kb. No size selection was performed and 600 ng of tagmented fragments were circularized. The circularized DNA was mechanically sheared to small fragments with an optimal at 692 bp on the Covaris device S2 in microtubes (Covaris, Woburn, MA, USA). The library profile was visualized on a High Sensitivity Bioanalyzer LabChip (Agilent Technologies Inc, Santa Clara, CA, USA). The libraries were normalized at 2 nM and pooled. After a denaturation step and dilution at 10 pM, the pool of libraries was loaded onto the reagent cartridge and then onto the instrument along with the flow cell. Automated cluster generation and sequencing run were performed in a single 42-h run in a 2 × 250-bp. Total information of 3.9 Gb was obtained from a 399 K/mm<sup>2</sup> cluster density with a

**Table 2** Project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	High-quality draft
MIGS-28	Libraries used	Paired end and mate pair
MIGS-29	Sequencing platform	MiSeq Technology (Illumina Inc)
MIGS-31.2	Sequencing coverage	94.91x
MIGS-30	Assemblers	Newbler version 2.5.3
MIGS-32	Gene calling method	Prodigal
	EMBL Date of Release	2014
	EMBL ID	CCAS00000000
MIGS-13	Source material identifier	JCE <sup>T</sup>
	Project relevance	Study of the human gut microbiome

cluster passing quality control filters of 97.9 % (7,840,000 clusters). Within this run, the index representation for *B. jeddahensis* was determined to 9.37 %. The 718,848 reads were filtered according to the read qualities. The passed filter sequences were assembled using Newbler with 90 % identity and 40-bp as overlap. The final assembly identified 149 large contigs (>1.5 kb) generating a genome size of 4.76 Mb which corresponds to a genome coverage of 94.91x.

#### Genome annotation

Open Reading Frames (ORFs) were predicted using Prodigal [26] with default parameters but the predicted ORFs were excluded if they spanned a sequencing gap region. The predicted bacterial protein sequences were searched against the GenBank database [27] and the Clusters of Orthologous Groups (COG) databases using BLASTP. The tRNAScanSE tool [28] was used to find tRNA genes, whereas ribosomal RNAs were found by using RNAmmer

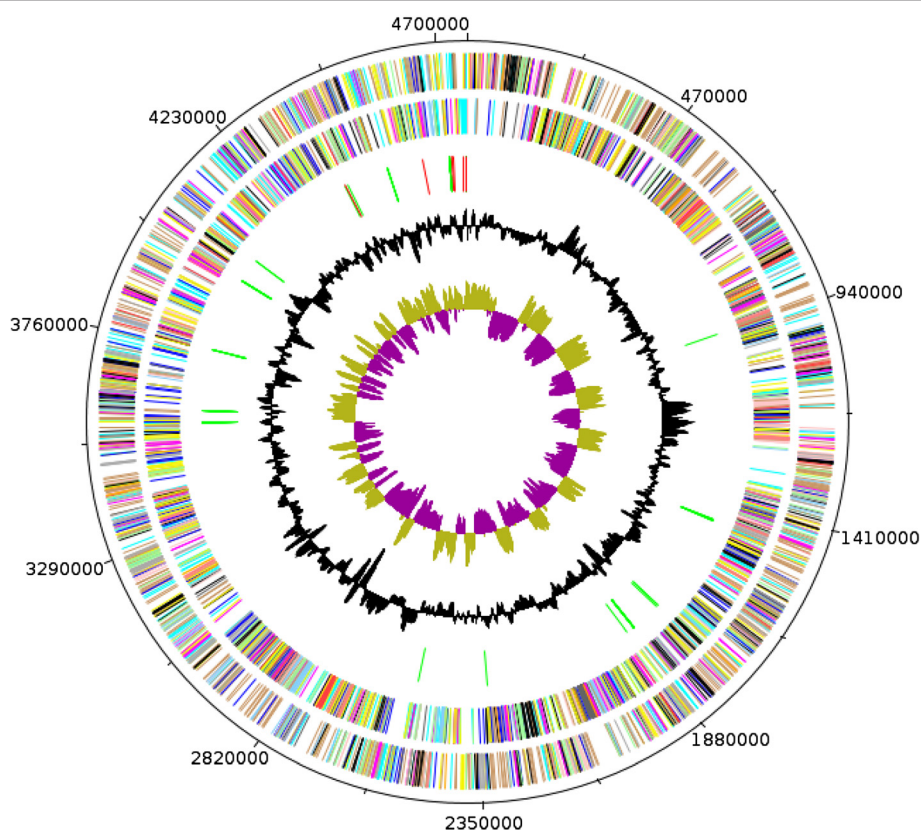
[29] and BLASTn against the GenBank database. Signal peptides and numbers of transmembrane helices were predicted using SignalP [30] and TMHMM [31], respectively. ORFans were identified if their BLASTP *E*-value was lower than  $1e-03$  for alignment length greater than 80 amino acids. If alignment lengths were smaller than 80 amino acids, we used an *E*-value of  $1e-05$ . Such parameter thresholds have already been used in previous works to define ORFans. Artemis [32] and DNA Plotter [33] were used for data management and visualization of genomic features, respectively. Mauve alignment tool (version 2.3.1) was used for multiple genomic sequence alignment [34].

To estimate the mean level of nucleotide sequence similarity at the genome level between *B. jeddahensis* sp nov. strain JCE<sup>T</sup> and nine other members of the genus *Bacillus*, we use the Average Genomic Identity of orthologous gene Sequences (AGIOS) program. Briefly, this software combines the Proteinortho software [35] to detect orthologous proteins between genomes compared on a pair-wise basis, then retrieves the corresponding genes and determines the mean percentage of nucleotide sequence identity among orthologous ORFs using the

Needleman-Wunsch global alignment algorithm. Moreover, we used Genome-to-Genome Distance Calculator (GGDC) web server available at (<http://ggdc.dsmz.de>) to estimate of the overall similarity among the compared genomes and to replace the wet-lab DNA-DNA hybridization (DDH) by a digital DDH (dDDH) [36, 37]. GGDC 2.0 BLAST+ was chosen as alignment method and the recommended formula 2 was taken into account to interpret the results.

### Genome properties

The genome 4,762,944 bp long (1 chromosome, but no plasmid) with a 39.42 % G + C content (Fig. 6 and Table 3). It is composed of 149 contigs. Of the 4,741 predicted genes, 4,654 were protein-coding genes and 98 were RNAs including 6 rRNA (1 gene is 16S rRNA, 1 gene is 23S rRNA and 5 genes are 5S rRNA). A total of 3,410 genes (71.92 %) were assigned a putative function (by COGs or by NR blast) and 147 genes were identified as ORFans (3.17 %). The distribution of genes into COGs functional categories is presented in Table 4. The properties and statistics of the genome are summarized in Tables 3 and 4.



**Fig. 6** Graphical circular map of the chromosome. From outside to the center: Genes on the forward strand colored by COG categories (only genes assigned to COG), genes on the reverse strand colored by COG categories (only gene assigned to COG), RNA genes (tRNAs green, rRNAs red), G + C content and GC skew. Purple and olive indicating negative and positive values, respectively

**Table 3** Nucleotide content and gene count levels of the genome

Attribute	Genome (total)	
	Value	% of total <sup>a</sup>
Size (bp)	4,762,944	100
G + C content (bp)	1,876,599	39.42
Coding region (bp)	4,065,588	85.36
DNA scaffolds	ND	
Total genes	4,741	100
RNA genes	98	2.07
Protein-coding genes	4,654	97.72
Pseudo genes	ND	
Genes in internal clusters	ND	
Genes with function prediction	3,410	71.92
Genes assigned to COGs	2,902	61.21
Genes with Pfam domains	ND	
Genes with peptide signals	93	1.96
Genes with transmembrane helices	1,301	27.44
CRISPR repeats	ND	

<sup>a</sup>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. ND not determined

### Insights from the genome sequence

Here, we compared the genome of *B. jeddahensis* strain JCE<sup>T</sup> with those of “*Bacillus massiliosenegalensis*” strain JC6<sup>T</sup>, “*Bacillus massilioanorexius*” strain AP8<sup>T</sup>, “*Bacillus timonensis*” strain MM10403188, *Bacillus cereus* strain ATCC 14579, *Bacillus megaterium* strain DSM 319, *Bacillus licheniformis* strain ATCC 14580, *Bacillus bataviensis* strain DSM 15601<sup>T</sup>, *Bacillus vireti* DSM 15602<sup>T</sup>, *Bacillus niacini* JAM F8 and *Bacillus niacini* DSM 2923<sup>T</sup> (Tables 5 and 6). The draft genome sequence of *B. jeddahensis* strain is larger in size than those of “*B. massilioanorexius*”, “*B. timonensis*”, *B. licheniformis* and *B. niacini* DSM 2923<sup>T</sup> (4.76 vs 4.59, 4.66, 4.22 and 2.2 Mb, respectively), but smaller than those of “*B. massiliosenegalensis*”, *B. cereus*, *B. megaterium*, *B. bataviensis*, *B. vireti* and *B. niacini* JAM F8 (4.76 vs 4.97, 5.43, 5.10, 5.37, 5.29 and 6.37 Mb, respectively) (Table 5). *B. jeddahensis* has a lower G + C content than those of *B. licheniformis*, *B. bataviensis* and *B. vireti* (39.42 vs 46.19, 39.6 and 39.74 %, respectively) and higher than those of “*B. massiliosenegalensis*”, “*B. massilioanorexius*”, “*B. timonensis*”, *B. cereus*, *B. megaterium*, *B. niacini* JAM F8 and *B. niacini* DSM 2923<sup>T</sup> (39.42 vs 37.58, 34.10, 37.28, 35.29, 38.13, 37.83 and 38.30 %, respectively) (Table 5). As it was reported recently that the G + C content varies no more than 1 % within species [38] and because the strain JCE<sup>T</sup> differs from most of the other closely strains by more than 1 % in G + C content, this might provide an

additional argument for the new taxon described herein. The protein content of *B. jeddahensis* is higher than those of “*B. massilioanorexius*”, “*B. timonensis*”, *B. licheniformis* and *B. niacini* DSM 2923<sup>T</sup> (4654 vs 4436, 4647, 4173 and 2184, respectively) but lower than those of “*B. massiliosenegalensis*”, *B. cereus*, *B. megaterium*, *B. bataviensis*, *B. vireti* and *B. niacini* JAM F8 (4654 vs 4935, 5231, 5100, 5207, 5092 and 6103, respectively) (Table 6). The distribution of genes into COG categories was not entirely similar in all the nine compared genomes (Fig. 7). In addition, *B. jeddahensis* shares 2075, 1786, 1930, 1729, 1894, 1715, 2494, 2433, 2404 and 914 orthologous genes with those of “*B. massiliosenegalensis*”, “*B. massilioanorexius*”, “*B. timonensis*”, *B. cereus*, *B. megaterium*, *B. licheniformis*, *B. bataviensis*, *B. vireti*, *B. niacini* JAM F8 and *B. niacini*

**Table 4** Number of genes associated with the 25 general COG functional categories

Code	Value	% of total <sup>a</sup>	Description
J	179	3.86	Translation
A	0	0	RNA processing and modification
K	342	7.38	Transcription
L	195	4.21	Replication, recombination and repair
B	1	0.02	Chromatin structure and dynamics
D	39	0.84	Cell cycle control, mitosis and meiosis
Y	0	0	Nuclear structure
V	81	1.75	Defense mechanisms
T	255	5.5	Signal transduction mechanisms
M	192	4.14	Cell wall/membrane biogenesis
N	57	1.23	Cell motility
Z	0	0	Cytoskeleton
W	0	0	Extracellular structures
U	48	1.04	Intracellular trafficking and secretion
O	119	2.57	Posttranslational modification, protein turnover, chaperones
C	224	4.83	Energy production and conversion
G	300	6.48	Carbohydrate transport and metabolism
E	425	9.17	Amino acid transport and metabolism
F	89	1.92	Nucleotide transport and metabolism
H	131	2.83	Coenzyme transport and metabolism
I	139	3	Lipid transport and metabolism
P	261	5.63	Inorganic ion transport and metabolism
Q	82	1.77	Secondary metabolites biosynthesis, transport and catabolism
R	500	10.79	General function prediction only
S	348	7.51	Function unknown
-	1223	26.4	Not in COGs

<sup>a</sup>The total is based on the total number of protein coding genes in the annotated genome



**Table 5** Genomic comparison of *B. jeddahensis* sp. nov., strain JCE<sup>T</sup> with other *Bacillus* species. Species and strain names, GenBank genome accession numbers, sizes and G + C contents

Species	Strain	Genome accession number	Genome size (Mb)	G + C content
<i>Bacillus jeddahensis</i>	JCE <sup>T</sup>	CCAS000000000	4.76	39.42
" <i>Bacillus massiliosenegalensis</i> "	JC6 <sup>T</sup>	CAHJ000000000	4.97	37.58
" <i>Bacillus massilioanorexius</i> "	AP8 <sup>T</sup>	CAPG000000000	4.59	34.10
" <i>Bacillus timonensis</i> "	MM10403188	CAET000000000	4.66	37.28
<i>Bacillus cereus</i>	ATCC 14579	NC_004722	5.43	35.29
<i>Bacillus megaterium</i>	DSM 319	NC_014103	5.10	38.13
<i>Bacillus licheniformis</i>	ATCC 14580	NC_006270	4.22	46.19
<i>Bacillus bataviensis</i>	DSM 15601 <sup>T</sup>	AJLS000000000	5.37	39.60
<i>Bacillus vireti</i>	DSM 15602 <sup>T</sup>	ALAN000000000	5.29	39.74
<i>Bacillus niacini</i>	JAM F8	BAWM000000000	6.37	37.83
<i>Bacillus niacini</i>	DSM 2923 <sup>T</sup>	JRYQ000000000	2.20	38.30

Bold numbers indicate numbers of proteins per genome

DSM 2923<sup>T</sup>, respectively. Among compared genomes except *B. jeddahensis*, AGIOS values range from 64.44 between *B. cereus* and *B. licheniformis* to 83.91 % between *B. niacini* JAM F8 and *B. niacini* DSM 2923<sup>T</sup>. When *B. jeddahensis* was compared to other species, AGIOS values range from 65.50 with *B. licheniformis* to 78.49 % with *B. bataviensis* (Table 6). dDDH estimation of the strain JCE<sup>T</sup> against the compared genomes ranged between 19.50 to 28.10. These values are very low and below the cutoff of 70 %, thus confirming again the new species status of the strain JCE<sup>T</sup>. Table 5 summarizes the number of orthologous genes and the average percentage of nucleotide sequence identity between the different genomes studied.

## Conclusions

On the basis of phenotypic characteristics (Additional file 1: Table S1), phylogenetic position (Fig. 1), genomic

analyses (taxonogenomics) (Table 5) and GGDC results, we formally propose the creation of *Bacillus jeddahensis* sp. nov. that contains the strain JCE<sup>T</sup>. This strain has been found in obese human feces collected from Jeddah, Saudi Arabia.

## Description of *Bacillus jeddahensis* sp. nov. strain JCE<sup>T</sup>

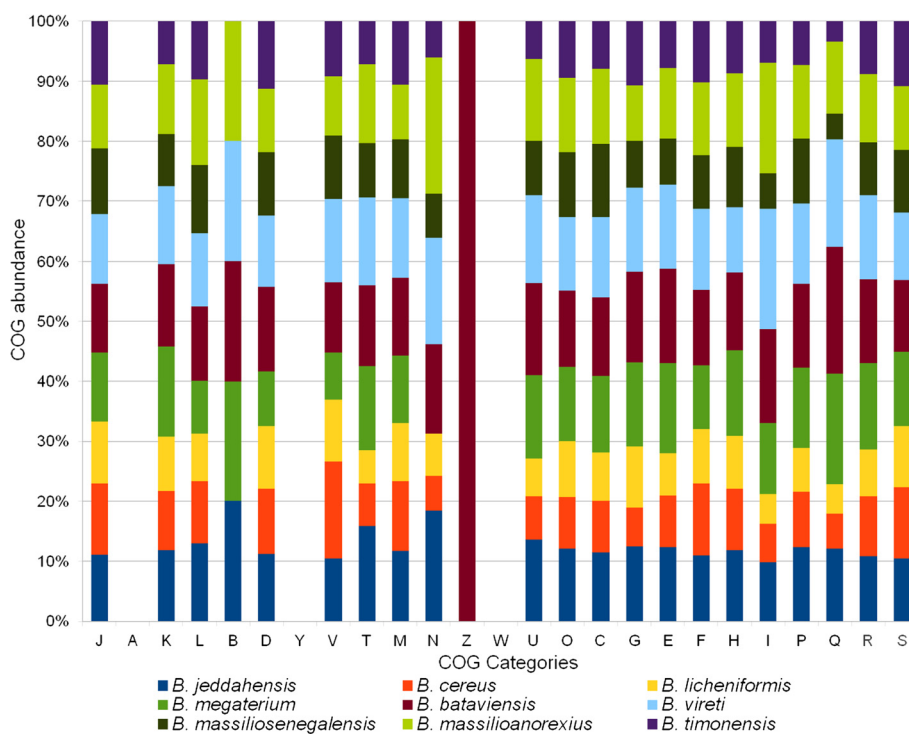
*Bacillus jeddahensis* (jed.dah. en 'sis L. gen. neutr. n. *jeddahensis*, pertaining to, or originating from Jeddah, the capital of Saudi Arabia, where the type strain was isolated).

*B. jeddahensis* is a Gram-positive. Optimal growth is achieved aerobically. But growth is also observed in microaerophilic or anaerobic conditions. Growth occurs on axenic media between 28 and 45 °C, with optimal growth observed at 37 °C. Cells stain Gram-positive, are

**Table 6** Genomic comparison of *B. jeddahensis* sp. nov., strain JCE<sup>T</sup> with other *Bacillus* species. Numbers of orthologous proteins shared between genomes (upper right triangle), average percentage similarity of nucleotides corresponding to orthologous proteins shared between genomes (lower left triangle)

	JCE <sup>T</sup>	JC6 <sup>T</sup>	AP8 <sup>T</sup>	MM10403188	ATCC 14579	DSM 319	ATCC 14580	DSM 15601 <sup>T</sup>	DSM 15602 <sup>T</sup>	JAM F8	DSM 2923 <sup>T</sup>
JCE <sup>T</sup>	<b>4654</b>	2075	1786	1930	1729	1894	1715	2494	2433	2404	914
JC6 <sup>T</sup>	69.86	<b>4935</b>	1858	1960	1669	1862	1746	2118	2132	2169	836
AP8 <sup>T</sup>	68.13	68.67	<b>4436</b>	1726	1674	1778	1636	1848	1854	1911	711
MM10403188	68.48	68.73	68.49	<b>4647</b>	1631	1811	1691	2092	2047	2103	829
ATCC 14579	66.99	67.35	67.83	67.99	<b>5231</b>	1837	1678	1776	1792	1837	726
DSM 319	67.26	67.43	67.80	68.24	67.94	<b>5100</b>	1871	1978	1941	2033	805
ATCC 14580	65.50	65.33	64.32	65.72	64.44	66.06	<b>4173</b>	1761	1794	1784	716
DSM 15601 <sup>T</sup>	78.49	69.51	67.94	68.53	66.69	67.18	65.54	<b>5207</b>	2699	2639	989
DSM 15602 <sup>T</sup>	77.41	69.37	67.69	68.32	66.58	66.97	65.61	79.72	<b>5092</b>	2528	956
JAM F8	74.84	70.04	68.49	68.90	67.21	67.42	65.18	74.95	74.98	<b>6103</b>	1102
DSM 2923 <sup>T</sup>	74.17	69.58	68.48	68.64	67.05	67.39	65.54	74.55	74.46	83.91	<b>2184</b>

Bold numbers indicate numbers of proteins per genome



**Fig. 7** Distribution of predicted genes of *B. jeddahensis* and eight other *Bacillus* species into COG categories

rod-shaped, endospore-forming and non-motile with a mean diameter of 0.87  $\mu\text{m}$  (range 0.75 to 0.95  $\mu\text{m}$ ) and a mean length of 4.1  $\mu\text{m}$  (range 3.8 to 4.7  $\mu\text{m}$ ). Colonies are smooth grey and 0.4–0.5 mm in diameter on blood-enriched Columbia agar.

Catalase negative, oxidase positive. A positive reaction is obtained for D-arabinose, L-arabinose, D-xylose, D-glucose, D-fructose, D-mannose, N-acetylglucosamine, esculin, D-maltose, D-trehalose, and weak reaction for D-melezitose. Negative reactions are obtained for the remaining carbohydrate tests (i.e. glycerol, erythritol, D-ribose, L-xylose, D-adonitol, methyl- $\beta$ -D-xylopyranoside, D-galactose, L-sorbose, L-rhamnose, dulcitol, inositol, D-mannitol, D-sorbitol, methyl- $\alpha$ -D-mannopyranoside, methyl- $\alpha$ -D-glucopyranoside, amygdalin, arbutin, salicin, D-cellobiose, D-lactose, D-melibiose, D-saccharose, inulin, D-raffinose, amidon, glycogen, xylitol, gentiobiose, D-turanose, D-lyxose, D-tagatose, D-fucose, L-fucose, D-arabitol, L-arabitol, potassium gluconate, potassium 2-ketogluconate and potassium 5-ketogluconate). Positive reactions are observed for esterase (C 4), esterase lipase (C 8), acid phosphatase, naphthol-AS-BI-phosphohydrolase and  $\beta$ -glucosidase. Negative reactions are obtained for alkaline phosphatase, lipase (C 14), leucine arylamidase, valine arylamidase, cystine arylamidase, trypsin,  $\alpha$ -chymotrypsin,  $\alpha$ -galactosidase,  $\beta$ -galactosidase,  $\beta$ -glucuronidase,  $\alpha$ -glucosidase, N-acetyl- $\beta$ -glucosaminidase,  $\alpha$ -mannosidase and  $\alpha$ -fucosidase. Nitrates are reduced to

nitrites, the urease reaction, indole production, arginine dihydrolase and gelatin hydrolysis are negative, the following carbon sources are assimilated: D-glucose, D-mannose, N-acetylglucosamine and D-maltose, and the following carbon sources were not assimilated: L-arabinose, D-mannitol, potassium gluconate, capric acid, adipic acid, malic acid, trisodium citrate and phenylacetic acid. *B. jeddahensis* is susceptible to imipenem, doxycycline, amoxicillin, amoxicillin-clavulanate and gentamycin, but resistant to metronidazole, trimethoprim/sulfamethoxazole, rifampicin, vancomycin, erythromycin, ceftriaxone, ciprofloxacin and benzylpenicillin.

The G + C content of the genome is 39.42 %. The 16S rRNA and genome sequences are deposited in GenBank under accession numbers HG931339 and CCAS00000000, respectively. The type strain JCE<sup>T</sup> (= CSUR P732 = DSM 28281) was isolated from the fecal flora of an obese man from Jeddah in Saudi Arabia.

### Additional file

**Additional file 1: Table S1.** Differential phenotypic characteristics between *B. jeddahensis* sp. nov. strain JCE<sup>T</sup> and phylogenetically close *Bacillus* species.

### Abbreviations

URMITE: Unité de Recherchesur les Maladies Infectieuses et Tropicales Emergentes; CSUR: Collection de Souches de l'Unité des Rickettsies;

DSM: Deutsche Sammlung von Mikroorganismen; MALDI-TOF MS: Matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry; TE buffer: Tris-EDTA buffer; GGDC: Genome-to-Genome Distance Calculator; dDDH: digital DNA-DNA hybridization.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

F Bittar, PE Fournier and D Raoult conceived the study, participated in its design and drafted the manuscript. F Bittar, F Bibi, JC Lagier, El Azhar, AA Jiman-Fatani, AK Al-Ghamdi, M Yasir and D Raoult contributed materials and analyses. F Bittar and JC Lagier performed the phenotypic characterization of the bacterium. D Ramasamy performed the genomic analyses and helped to draft the manuscript. TT Nguyen performed the genomic sequencing and helped to draft the manuscript. All authors read and approved the final manuscript.

### 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. Fadi Bittar was supported by a Chair of Excellence IRD provided by the Institut de Recherche pour le Développement / Méditerranée-Infection foundation.

### Author details

<sup>1</sup>URMITE, Aix-Marseille Université, Faculté de médecine, Marseille, France. <sup>2</sup>Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia. <sup>3</sup>Department of Medical Laboratory Technology, Faculty of Applied Medical Sciences, King Abdulaziz University, Jeddah, Saudi Arabia. <sup>4</sup>Department of Medical Microbiology and Parasitology, Faculty of Medicine, King Abdulaziz University, Jeddah, Saudi Arabia.

Received: 23 May 2014 Accepted: 21 May 2015

Published online: 01 August 2015

### References

- Mishra AK, Lagier JC, Rivet R, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of "*Paenibacillus senegalensis*" sp. nov. *Stand Genomic Sci.* 2012;7:70–81.
- Mamadou Bhoie K, Seydina D, Catherine R, Didier R, Pierre-Edouard F, Fadi B. Non-contiguous finished genome sequence and description of *Bacillus massiliogorillae* sp. nov. *Stand Genomic Sci.* 2013;9:93–105.
- Mishra AK, Lagier JC, Robert C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Clostridium senegalense* sp. nov. *Stand Genomic Sci.* 2012;6:386–95.
- Lagier JC, Armougom F, Mishra AK, Ngyuen TT, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Alistipes timonensis* sp. nov. *Stand Genomic Sci.* 2012;6:315–24.
- Lagier JC, El Karkouri K, Nguyen TT, Armougom F, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Anaerococcus senegalensis* sp. nov. *Stand Genomic Sci.* 2012;6:116–25.
- Roux V, El Karkouri K, Lagier JC, Robert C, Raoult D. Non-contiguous finished genome sequence and description of *Kurthia massiliensis* sp. nov. *Stand Genomic Sci.* 2012;7:221–32.
- Mishra AK, Lagier JC, Robert C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Peptoniphilus timonensis* sp. nov. *Stand Genomic Sci.* 2012;7:1–11.
- Hugon P, Ramasamy D, Lagier JC, Rivet R, Couderc C, Raoult D, et al. Non contiguous-finished genome sequence and description of *Alistipes obesi* sp. nov. *Stand Genomic Sci.* 2013;7:427–39.
- Ramasamy D, Lagier JC, Nguyen TT, Raoult D, Fournier PE. Non contiguous-finished genome sequence and description of *Dielma fastidiosa* gen. nov., sp. nov., a new member of the Family *Eysipelotrichaceae*. *Stand Genomic Sci.* 2013;8:336–51.
- Mishra AK, Lagier JC, Robert C, Raoult D, Fournier PE. Genome sequence and description of *Timonella senegalensis* gen. nov., sp. nov., a new member of the suborder *Micrococinae*. *Stand Genomic Sci.* 2013;8:318–35.
- Mishra AK, Hugon P, Lagier JC, Nguyen TT, Couderc C, Raoult D, et al. Non contiguous-finished genome sequence and description of *Enorma massiliensis* gen. nov., sp. nov., a new member of the Family *Coriobacteriaceae*. *Stand Genomic Sci.* 2013;8:290–305.
- Ramasamy D, Lagier JC, Gorlas A, Raoult D, Fournier PE. Non contiguous-finished genome sequence and description of *Bacillus massilosenegalensis* sp. nov. *Stand Genomic Sci.* 2013;8:264–78.
- Hugon P, Mishra AK, Lagier JC, Nguyen TT, Couderc C, Raoult D, et al. Non-contiguous finished genome sequence and description of *Brevibacillus massiliensis* sp. nov. *Stand Genomic Sci.* 2013;8:1–14.
- Cohn F. Untersuchungen über Bakterien. *Beiträge zur Biologie der Pflanzen Heft.* 1872;1:127–224.
- Abstract for the genus *Bacillus*. NamesforLife, LLC. Retrieved January 30, 2014. <http://doi.namesforlife.com/10.1601/tx.4857>
- Mandell GL, Bennett JE, Dolin R. *Principles and Practice of Infectious Diseases*. Elsevier. 2010; 4320 p. CHURCHILL LIVINGSTONE, Pennsylvania.
- Lagier JC, Armougom F, Million M, Hugon P, Pagnier I, Robert C, et al. Microbial culturomics: paradigm shift in the human gut microbiome study. *Clin Microbiol Infect.* 2012;18:1185–93.
- Dubourg G, Lagier JC, Armougom F, Robert C, Hamad I, Brouqui P, et al. The gut microbiota of a patient with resistant tuberculosis is more comprehensively studied by culturomics than by metagenomics. *Eur J Clin Microbiol Infect Dis.* 2013;32:637–45.
- Pfleiderer A, Lagier JC, Armougom F, Robert C, Vialettes B, Raoult D. Culturomics identified 11 new bacterial species from a single anorexia nervosa stool sample. *Eur J Clin Microbiol Infect Dis.* 2013;32:1471–81.
- Sentaosa E, Fournier PE. Advantages and limitations of genomics in prokaryotic taxonomy. *Clin Microbiol Infect.* 2013;19:790–5.
- Meier-Kolthoff JP, Göker M, Spröer C, Klenk HP. When should a DDH experiment be mandatory in microbial taxonomy? *Arch Microbiol.* 2013;6:413–8.
- Heyrman J, Vanparys B, Logan NA, Balcaen A, Rodríguez-Díaz M, Felske A, et al. *Bacillus novalis* sp. nov., *Bacillus vireti* sp. nov., *Bacillus soli* sp. nov., *Bacillus bataviensis* sp. nov. and *Bacillus dretnensis* sp. nov., from the Drentse A grasslands. *Int J Syst Evol Microbiol.* 2004;54:47–57.
- Mishra AK, Pfeleiderer A, Lagier JC, Robert C, Raoult D, Fournier PE. Non-contiguous finished genome sequence and description of *Bacillus massilioanorexius* sp. nov. *Stand Genomic Sci.* 2013;8:465–79.
- Nagel M, Andreesen JR. *Bacillus niacini* sp. nov. a nicotinate-metabolizing mesophile isolated from soil. *Int J Syst Bacteriol.* 1991;41:134–9.
- Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, et al. The minimum information about a genome sequence (MIGS) specification. *Nat Biotechnol.* 2008;26:541–7.
- Prodigal. <http://prodigal.ornl.gov>
- GenBank database. <http://www.ncbi.nlm.nih.gov/genbank>
- Lowe TM, Eddy SR. t-RNAscan-SE: a program for improved detection of transfer RNA gene in genomic sequence. *Nucl Acids Res.* 1997;25:955–64.
- Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucl Acids Res.* 2007;35:3100–8.
- Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol.* 2004;340:783–95.
- Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol.* 2001;305:567–80.
- Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, et al. Artemis: sequence visualization and annotation. *Bioinformatics.* 2000;16:944–5.
- Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics.* 2009;25:119–20.
- Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genome sequence with rearrangements. *Genome Res.* 2004;14:1394–403.
- Lechner M, Findeib S, Steiner L, Marz M, Stadler PF, Prohaska SJ. Proteinortho: detection of (Co-) orthologs in large-scale analysis. *BMC Bioinformatics.* 2011;12:124.
- Auch AF, von Jan M, Klenk HP, Göker M. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand Genomic Sci.* 2010;2:117–34.
- Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics.* 2013;14:60.
- Meier-Kolthoff JP, Klenk HP, Göker M. Taxonomic use of DNA G + C content and DNA-DNA hybridization in the genomic age. *Int J Syst Evol Microbiol.* 2014;64(Pt 2):352–6.

39. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains *Archae*, *Bacteria*, and *Eukarya*. *Proc Natl Acad Sci U S A*. 1990;87:4576–9.
40. Gibbons NE, Murray RGE. Proposals concerning the higher taxa of *Bacteria*. *Int J Syst Bacteriol*. 1978;28:1–6.
41. Garrity GM, Holt JG. The road map to the manual. In: Garrity GM, Boone DR, Castenholz RW, editors. *Bergey's Manual of Systematic Bacteriology*, vol. 1. 2nd ed. New York: Springer; 2001. p. 119–69.
42. Murray RGE. The higher taxa, or, a place for everything...? In: Holt JG, editor. *Bergey's Manual of Systematic Bacteriology*, vol. 1. 1st ed. Baltimore: The Williams and Wilkins Co.; 1984. p. 31–4.
43. Ludwig W, Schleifer KH, Whitman WB. Class I. *Bacilli* class nov. In: De Vos P, Garrity G, Jones D, Krieg NR, Ludwig W, Rainey FA, Schleifer KH, Whitman WB, editors. *Bergey's Manual of Systematic Bacteriology*. Second ed. New York: Springer; 2009. p. 19–20.
44. Euzéby J. 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.
45. Skerman VBD, McGowan V, Sneath PHA. Approved lists of bacterial names. *Int J Syst Bacteriol*. 1980;30:225–420.
46. Prévot AR. In: Hauderoy P, Ehringer G, Guillot G, Magrou J, Prévot AR, Rosset D, Urbain A (eds), *Dictionnaire des Bactéries Pathogènes*, Second Edition, Masson et Cie, Paris, 1953, p. 1–692
47. Fischer A. Untersuchungen über bakterien. *Jahrbücher für Wissenschaftliche Botanik*. 1895;27:1–163.
48. Gibson T, Gordon RE. Genus I. *Bacillus* Cohn 1872, 174; Nom. gen. cons. *Nomencl. Comm. Intern. Soc. Microbiol*. 1937, 28; *Opin. A. Jud. Comm*. 1955, 39. In: Buchanan RE, Gibbons NE, editors. *Bergey's Manual of Determinative Bacteriology*. Eighthth ed. Baltimore: The Williams and Wilkins Co; 1974. p. 529–50.
49. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. *The gene ontology consortium*. *Nat Genet*. 2000;25:25–9.
50. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. 2011;28:2731–9.

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

