Noncontiguous finished genome sequence and description of *Bacillus testis* strain SIT10 sp. nov.

T. Cimmino[1], S. I. Traore[1,3], C. Valentini[1], S. le Page[1], C. Sokhna[2,4], A. Diallo[5], D. Raoult[1] and J. M. Rolain[1]

1) *Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, UM 63, CNRS 7278, IRD 198, Inserm U1095, Institut Hospitalo-Universitaire Méditerranée-Infection, Faculté de médecine,* 2) *Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes IRD 198, CNRS 7278, Aix-Marseille Université, Marseille, France,* 3) *Département d'Epidémiologie des Affections Parasitaires, Faculté de médecine et de pharmacie de Bamako, Bamako, Mali,* 4) *Campus Commun UCAD-IRD of Hann and* 5) *URMITE, UMR CNRS 6236, IRD 198, Aix Marseille Université, Marseille, Joint Campus UCAD-IRD of Hann, BP 1386 CP 18524, Dakar, Senegal*

Abstract

Bacillus testis strain SIT10 (= CSUR P1492 = DSMZ 101190) is the new type strain collected from stool from a 2-year-old boy from Senegal during a culturomics study. This Gram-positive bacterium is a facultative anaerobic rod and a member of the Bacillaceae family. We describe here the features of this bacterium, together with the complete genome sequence and annotation. The 3 987 349 bp long genome (one chromosome but no plasmid) with 42.8% GC content contains 4005 protein-coding and 171 sRNA genes, including 19 5S rRNA gene, 15 16S rRNA genes and ten 23S rRNA genes.

Keywords: *Bacillus testis*, culturomics, taxonogenomics
Original Submission: 12 March 2016; **Accepted:** 22 March 2016
Article published online: 13 April 2016

Corresponding author: J. M. Rolain, Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, UM 63, CNRS 7278, IRD 198, Inserm U1095, Institut Hospitalo-Universitaire Méditerranée-Infection, Faculté de médecine, Aix-Marseille Université, 27 Boulevard Jean Moulin, 13385 Marseille cedex 05, France
E-mail: jean-marc.rolain@univ-amu.fr

Introduction

Culturomics was developed in 2012 in order to extend knowledge of the human gut repertoire [1]. A polyphasic approach that combines proteomic by matrix-assisted desorption ionization–time of flight mass spectrometry (MALDI-TOF MS) analysis, genomic data and phenotypic characterization is widely used to describe new bacterial species [2].

Bacillus testis strain SIT10 was isolated from a 2-year-old boy in Senegal as part of a culturomics study aiming to isolate all bacterial species present in the human gut [1]. The genus *Bacillus* currently comprises 268 species and seven subspecies, although a few of these have been assigned to other genera commonly found in the environment and as laboratory contaminants; however, few species have been linked to infections in humans (http://www.bacterio.net/) [3]. Two *Bacillus* species are considered medically significant: *B. anthracis*, which causes anthrax, and *B. cereus*, which causes a foodborne illness [4]. This genus is one of the largest and most ubiquitous, and it has gained notoriety with taxonomists for its extreme phenotypic diversity and heterogeneity. Bacilli are ubiquitous bacteria that exploit a wide variety of organic and inorganic substrates as nutrient sources [3].

Herein we present a summary of the classification and set of features for *Bacillus testis* sp. nov. strain SIT10 (= CSUR P1492 = DSMZ 101190) together with the description of the complete genomic sequencing and annotation. These characteristics support the creation of this *Bacillus testis* species.

Organism Information

Classification and features

A stool sample was collected from a 2-year-old boy living in Senegal. The study was approved by the ethics committee of

http://dx.doi.org/10.1016/j.nmni.2016.03.007

the Institut Fédératif de Recherche IFR48, Faculty of Medicine, Marseille, France, under agreement 09-002.

The faecal specimen was preserved at −80°C after collection. Strain SIT10 was isolated on Columbia agar supplemented with 5% sheep's blood (bioMérieux, Marcy-l'Étoile, France) in aerobic and anaerobic condition using GasPak EZ Anaerobe Container System Sachets (Becton Dickinson (BD), San Diego, CA, USA) at 37°C. Strain SIT10 exhibited a 97.6% 16S rRNA sequence identity with *Bacillus massiliogorillae* strain G2 (NZ_CAVL000000000.2), the phylogenetically closest bacterial species withstanding in nomenclature (Fig. 1). Its 16S rRNA sequence was deposited in GenBank under accession number LN827531. This value was lower than the 98.7% 16S rRNA gene sequence threshold recommended by Stackebrandt and Ebers [5] to delineate a new species without carrying out DNA-DNA hybridization. The spectrum from SIT10 was added to our MALDI-TOF MS database.

Growth conditions and identification

Growth at different temperatures (25, 37, 45°C) was tested. Growth of the strain was tested in 5% sheep's blood–enriched Columbia agar (bioMérieux) under anaerobic using GENbag anaer systems (bioMérieux). Growth was achieved only both aerobically and anaerobically. Gram staining and electron microscopy were performed with a TechnaiG^2 Cryo (FEI Company, Limeil-Brevannes, France) at an operating voltage of 200 keV (Fig. 2). Cells were grown on 5% sheep's blood agar for 24 hours. A bacterial suspension was prefixed in 5% (v/v) glutaraldehyde in phosphate buffer (Thermo Fisher Scientific Life Sciences, Waltham, MA, USA) for at least 1 hour at room temperature, washed in the same buffer and stained with 1% (w/v) ammonium molybdate 1%. Catalase activity was determined by an ID Color catalase test kit (bioMérieux), and oxidase activity was assayed by applying the cells to moistened discs impregnated with dimethyl-p-phenylenediamine (bioMérieux). Biochemical tests were performed with the commercially available API ZYM and API 50 CH strips and were used to characterize the biochemical properties of the strain according to the manufacturer's instructions. The antibiotic susceptibility was tested using SirScan Discs antibiotics (i2a, Montpellier, France).

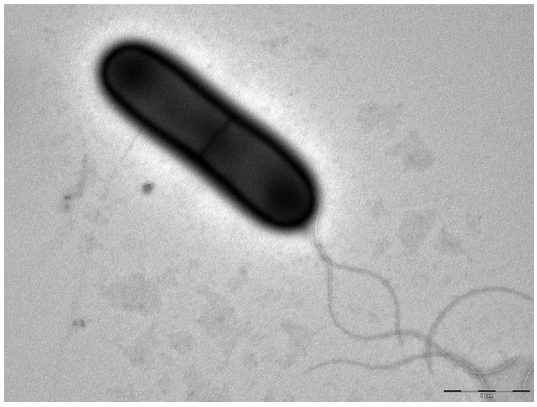

FIG. 2. Transmission electron microscopy of *Bacillus testis* strain SIT10 using a Morgani 268D (Philips, Amsterdam, The Netherlands) at operating voltage of 60 kV. Scale bar = 500 nm.

Extended Features Descriptions

MALDI-TOF MS [6] protein analysis was carried out as previously described. The SIT10 spectra were imported into the MALDI BioTyper software (version 3.0; Bruker Daltonics, Leipzig, Germany) and analysed by standard pattern matching (with default parameter settings) against 7765 spectra of bacteria, including 231 spectra from *Bacillus* genus. The method of identification included the *m/z* from 3000 to 15 000 Da. A maximum of 100 peaks were compared with spectra in the database for every spectrum. The resulting score enabled the identification (or not) of tested species: a score of $\geq$2 with a validly published species enabled identification at the species

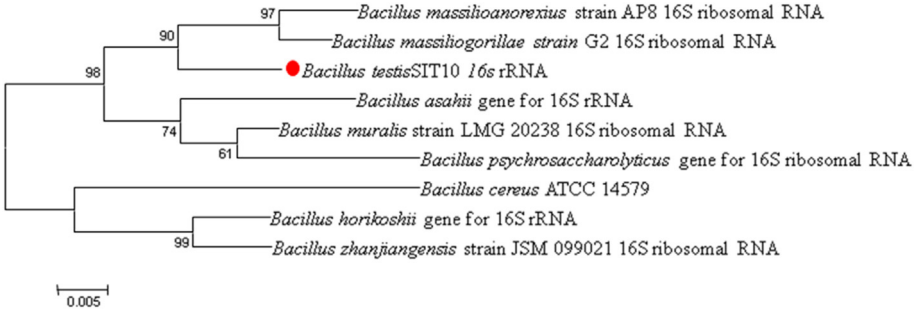

FIG. 1. Phylogenetic tree highlighting position of *Bacillus testis* strain SIT10 relative to other type strains within *Bacillus* genus. Sequences were aligned using MUSCLE, and phylogenetic inferences were obtained using maximum-likelihood method within MEGA6. Numbers at nodes are percentages of bootstrap values obtained by repeating analysis to generate majority consensus tree 1000 times.

level, a score of ≥1.7 but <2 enabled identification at the genus level, and a score of <1.7 did not enable any identification. No significant MALDI-TOF MS score was obtained for strain SIT10 against the Bruker database, suggesting that our isolate was not a member of a known species. We added the spectrum from strain SIT10 to our database

Genome sequencing and assembly

Genomic DNA of *Bacillus testis* strain SIT10 was sequenced on the MiSeq Technology (Illumina, San Diego, CA, USA) with the mate pair strategy. The assembly was performed using Soap De Novo with 179% coverage, and 1 910 663 paired reads were filtered according to the read qualities. Its leads to scaffolds and large contigs (>1500 bp) generated a genome size of 3.99 Mb.

Genome annotation and comparison

Genome was annotated by Rapid Annotation using Subsytem Technology (RAST) bioserver [7]. The predicted bacterial protein sequences were searched against the GenBank database and the Clusters of Orthologous Groups (COGs) databases using BLASTP (*E* value 1e-03, coverage 0.7 and identity percentage 30%). The tRNAScanSE tool was used to find tRNA genes [8], whereas ribosomal RNAs were found by using RNAmmer and BLASTn against the GenBank database [9]. The resistome was analysed with ARG-ANNOT (Antibiotic Resistance Gene-ANNOTation) database [10]. The exhaustive search for bacteriocin was performed using the database available in our laboratories (Bacteriocins of the URMITE database BUR; http://drissifatima.wix.com/bacteriocins). Protein sequences from this database allowed putative bacteriocins from human gut microbiota to be identified using BLASTp methodology [11]. Analysis of the presence of polyketide synthase and nonribosomal peptide synthase was performed by discriminating the gene with large size using a database realized in our laboratory, and predicted proteins were compared against the nonredundant (nr) GenBank database using BLASTp.

Phylogenetic relationships with closely related species were determined by MEGA6. The evolutionary history was inferred by using the maximum likelihood method based on the JTT matrix-based model. We compared the genome sequence of *Bacillus testis* strain SIT10 with those of *Bacillus massiliogorillae* strain G2b (NZ_CAVL000000000.2), *Bacillus cereus* ATCC 14579 (NC_004722.1) and *Bacillus psychrosaccharolyticus* ATCC 23296 (NZ_AJTN00000000.2).

Results

Phenotypic properties

Bacillus testis strain SIT10 grew in anaerobic and aerobic conditions on 5% sheep's blood–Columbia agar at 37°C. Colonies were 0.4 to 0.5 mm in diameter on Columbia agar, and they appeared smooth and grey in color at 37 °C. No growth was observed at 45°C. Gram staining showed Gram-positive bacilli (Fig. 2). A motility test was positive. Strain SIT10 exhibits positive catalase and negative oxidase activity. Acid production was observed for the following carbohydrates with API 50 CH strip (bioMérieux): L-arabinose, D-ribose, D-xylose, D-glucose, D-fructose, D-mannose, N-acetylglucosamine, amygdalin, arbutin, aesculin, salicin, cellobiose and maltose (Table 1). By using API ZYM, positive reactions were observed for phosphatase alkaline, esterase, leucine aminopeptidase, valine

TABLE 1. Differential characteristics of *Bacillus testis* strain SIT10, *B. massiliogorillae*, *B. cereus* ATCC 14579 and *B. psychrosaccharolyticus* ATCC 23296

Test	*B. testis* SIT10	*B. massiliogorillae* G2^T	*B. cereus* ATCC 14579	*B. psychrosaccharolyticus* ATCC 23296
Glycerol	+/−	−	ND	+
L-Arabinose	+	−	ND	+
D-Ribose	+	+	ND	+
D-Xylose	+	−	−	+
D-Galactose	−	−	+	ND
D-Glucose	+	+	+	+
D-Fructose	+	+	ND	+
D-Mannose	+	−	ND	+
Inositol	−		ND	+
D-Mannitol	−	−	−	+
Methyl β-D-glucopyranoside	−	ND	ND	ND
N-acetylglucosamine	+	+	ND	+
Amygdalin	+	+	ND	+
Aesculin	+	+	ND	+
Salicin	+	+	ND	+
Cellobiose	+	+	ND	+
Lactose	−	+	+	+
Melibiose	−	ND	ND	+
Raffinose	−	ND	ND	+
Gentiobiose	+/−	ND	ND	+
Turanose	−	ND	ND	+
D-Lyxose	−	ND	ND	−

+, positive result; −, negative result; NA, data not available.

TABLE 2. Genome features of *Bacillus testis* strain SIT10

Attribute	Value
Size (bp)	3 987 349
G+C content (bp)	42.8
RNAs gene	171
5S rRNA	19
16S rRNA	15
23S rRNA	10
Protein coding gene	4005
Genes with unknown function	264
Genes assigned to COGs	2673
CRISPRs	0
Genes associated to PKS or NRPS	2
Genes associated to toxin/antitoxin	6
Genes associated to resistome	3

COGs, Clusters of Orthologous Groups database; CRISPR, clustered regularly interspaced short palindromic repeat; G+C, guanine cytosine; NRPS, nonribosomal peptide synthase; PKS, polyketide synthase; rRNA, ribosomal RNA.

aminopeptidase, cystine aminopeptidase, chymotrypsin, phosphatase acid, α-galactosidase, β-galactosidase, α-glucosidase and β-glucosidase (Table 1). Phenotypically, *Bacillus testis* strain SIT10 was resistant to trimethoprim–sulfamethoxazole, cephalosporins (ceftazidime, ceftriaxone) and ticarcillin–clavulanic acid but was susceptible to oxacillin fosfomycin, erythromycin, clindamycin, vancomycin, teicoplanin, linezolid, gentamicin, ciprofloxacin, doxycycline, rifampicin and colistin.

Genome properties

The genome size of *Bacillus testis* strain SIT10 is 3 987 349 bp with a 42.8% G+C content and is assembled into nine scaffolds (28 large contig). A total of 4005 protein-coding genes are annotated; 171 were RNAs (19 genes were 5S rRNA, 15 were 16S rRNA, ten were 23S rRNA and 124 were tRNA). A total of 2673 genes were assigned as putative function, while 264 were assigned as unknown function (by COGs or by nr BLAST) (Table 2). The distribution of genes into COGs functional categories is presented in Table 3. The properties and comparisons of the genome are summarized in Table 4. It contains two intact phages 68.2 and 20.5 kb in size with 39% and 35.90% GC content, respectively.

Resistome

The resistome of *Bacillus testis* strain SIT10 includes (*bla*) *AmpS* β-lactamase encoding gene and the major facilitator superfamily (MFS).

Specific features

Analysis of the genome revealed the absence of nonribosomal polyketide synthesis but the presence of two bacteriocin peptidases that showed 65% similarity with *Clostridium sulfidigenes* (Fig. 3) and a pseudouridine synthase that showed 81% similarity with *Bacillus massiliogorillae*. The genome of *B. testis* contains the T2SS (type 2 secretion system) operon and approximately 36 genes distributed in five clusters encoding flagellar system disseminated in the genome.

TABLE 3. Number of genes associated with 25 general COGs functional categories

Code	Value	% of total	Description
J	168	4.45	Translation
A	0	0	RNA processing and modification
K	216	5.72	Transcription
L	181	4.79	Replication, recombination and repair
B	1	0.026	Chromatin structure and dynamics
D	32	0.84	Cell cycle control, mitosis and meiosis
Y	0	0	Nuclear structure
V	48	1.27	Defense mechanisms
T	105	2.78	Signal transduction mechanisms
M	103	2.73	Cell wall/membrane biogenesis
N	48	1.27	Cell motility
Z	0	0	Cytoskeleton
W	0	0	Extracellular structures
U	42	1.11	Intracellular trafficking and secretion
O	97	2.57	Posttranslational modification, protein turnover, chaperones
C	159	4.21	Energy production and conversion
G	173	4.58	Carbohydrate transport and metabolism
E	304	8.05	Amino acid transport and metabolism
F	86	2.27	Nucleotide transport and metabolism
H	89	2.35	Coenzyme transport and metabolism
I	124	3.28	Lipid transport and metabolism
P	203	5.38	Inorganic ion transport and metabolism
Q	71	1.88	Secondary metabolites biosynthesis, transport and catabolism
R	423	11.21	General function prediction only
S	264	6.99	Function unknown
—	1256	33.29	Not in COGs

COGs, Clusters of Orthologous Groups database.

TABLE 4. Genomic comparison of *Bacillus testis* strain SIT10 with other *Bacillus* species

Species	Strain	Size (Mb)	GC%	Gene content
Bacillus testis	SIT10	3.98	42.7	3772
B. cereus	ATCC 14579	5.4	35.3	5494
B. psychrosaccharolyticus	ATCC 23296	4.59	38.8	4621
B. massiliogorillae	G2	5.45	34.9	5118

Genome comparison

Here we compared the genome of *Bacillus testis* strain SIT10 with those of *Bacillus massiliogorillae* G2, *Bacillus cereus* ATCC 14579 and *Bacillus psychrosaccharolyticus* ATCC 23296. The draft genome of *Bacillus testis* strain SIT10 is smaller in size than those of *B. massiliogorilae, B. psychrosaccharolyticus* and *B. cereus* (3.9 vs. 5.45 and 4.59 Mb, respectively). The G+C content of *Bacillus testis* is larger than those of *B. massiliogorilae, B. psychrosaccharolyticus* and *B. cereus* (42.7% vs. 34.9%, 38.8% and 35.51%).

Conclusions

On the basis of phenotypic, phylogenetic and genomic analyses, we formally propose the creation of *Bacillus testis* strain SIT10 sp. nov. The strain was isolated from the stool sample of a 2-year-old boy from Senegal. Several other previously undescribed bacterial species were also cultivated from

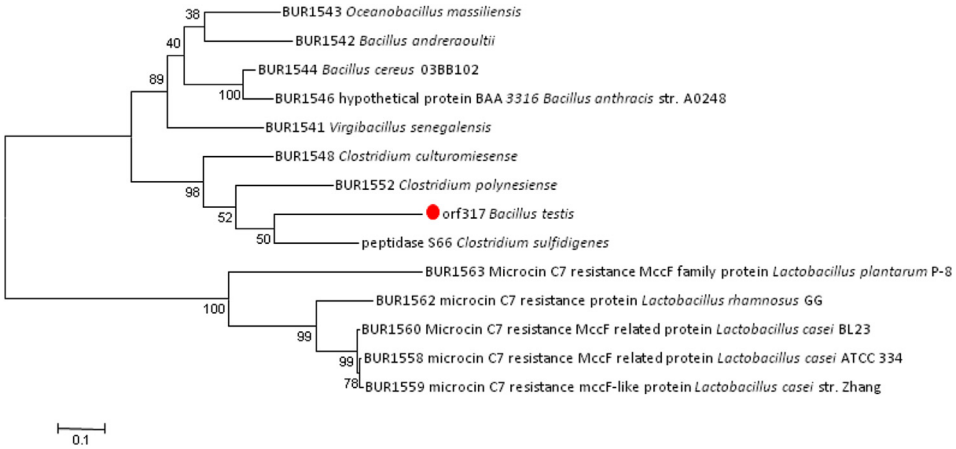

FIG. 3. Molecular phylogenetic analysis by maximum likelihood method of representatives of genus *Bacillus testis* strain SIT10 inferred from 16S rRNA gene sequence. Tree with highest log likelihood (−2930.4905) is shown. Percentage of trees in which associated taxa clustered together is shown next to branches. Initial tree or trees for heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, then selecting topology with superior log likelihood value. Tree is drawn to scale, with branch lengths measured in number of substitutions per site. Analysis involved 14 amino acid sequences. There were a total of 155 positions in the final data set. Evolutionary analyses were conducted in MEGA6.

different faecal samples through diversification of culture conditions [1].

Description of *Bacillus testis* strain SIT10 sp. nov

Bacillus testis strain SIT10 (= CSUR P1492 = DSMZ 101190) is the type strain of the genus *Bacillus*. It was isolated from a 2-year-old boy living in Senegal as part of a culturomics study aiming to isolate all bacterial species present in the human gut. The main scope of the culturomics study is to cultivate all the species within human faeces. *Bacillus testis* is a motile Gram-positive bacilli that exhibits positive catalase and negative oxidase activities. Colonies were 0.4 to 0.5 mm in diameter. It is a facultative anaerobic bacterium. Using API ZYM and API 50CH, positive reactions were found for phosphatase alkaline, esterase, leucine aminopeptidase, valine aminopeptidase, cystine aminopeptidase, chymotrypsin, phosphatase acid, α-galactosidase, β-galactosidase, α-glucosidase, β-glucosidase L-arabinose, D-ribose, D-xylose, D-glucose, D-fructose, D-mannose, N-acetylglucosamine, amygdalin, arbutin, aesculin, salicin, cellobiose and maltose. Antimicrobial susceptibility testing demonstrated that the cells were resistant to trimethoprim−sulfamethoxazole, cephalosporins (ceftazidime, ceftriaxone) and ticarcillin−clavulanic acid. The analysis of genome revealed the absence of nonribosomal polyketide synthesis but the presence of two bacteriocins.

Genome sequence accession number

The genome of *Bacillus testis* strain SIT10 has been submitted to the EBI database under bioproject ID PRJEB9400 with GenBank accession number CVQX00000000.1 and 16S RNA accession number LN827531.

Acknowledgements

We are grateful to L. Hadjadj for technical assistance and to Caroline for English-language editing. Funded by the French Centre National de la Recherche Scientifique (CNRS) and Infectiopole Sud Fundation.

Conflict of Interest

None declared.

References

[1] Lagier J, 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.

[2] Ramasamy D, Mishra AK, Lagier JC, Padhmanabhan R, Rossi M, Sentausa E, et al. A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. Int J Syst Evol Microbiol 2014;64:384–91.

[3] Koneman EW, Allen SD, Janda WM, Schreckenberger PC, Winn WJE. Koneman's color atlas and textbook of diagnostic microbiology. Lippincott Williams & Wilkins; 1997.
[4] Champoux James J, Neidhardt Frederick C, Lawrence Drew W, Plorde James J. Sherris medical microbiology. 4th ed. 2004.
[5] Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. Microbiol Today 2006;33:152–5.
[6] Seng P, Drancourt M, Gouriet F, La Scola B, Fournier PE, Rolain JM, et al. Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. Clin Infect Dis 2009;49:543–51.
[7] Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. BMC Genomics 2008;9:75.
[8] Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 1997;25:955–64.
[9] Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 2007;35:3100–8.
[10] Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L, et al. ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother 2014;58:212–20.
[11] Drissi F, Buffet S, Raoult D, Merhej V. Common occurrence of antibacterial agents in human intestinal microbiota. Front Microbiol 2015;6:441.