High-quality genome sequence and description of *Paenibacillus dakarensis* sp. nov.

C. I. Lo[1], S. A. Sankar[1], B. Fall[2], B. Sambe-Ba[2], O. Mediannikov[1,3], C. Robert[1], N. Faye[4], B. Wade[2], D. Raoult[1,3,5], P.-E. Fournier[1] and F. Fenollar[1,3]

1) *Aix-Marseille Université, URMITE, UM63, CNRS 7278, IRD 198, Inserm U1095, Faculté de médecine, Marseille, France,* 2) *Hôpital Principal de Dakar,* 3) *Campus International UCAD-IRD,* 4) *Université Cheikh Anta Diop de Dakar, Laboratoire de Parasitologie générale, Dakar, Senegal and* 5) *Special Infectious Agents Unit, King Fahd Medical Research Center, King Abdulaziz University, Jeddah, Saudi Arabia*

Abstract

Strain FF9T was isolated in Dakar (Senegal) from a blood-culture taken from a 16-month-old child. MALDI-TOF analysis did not allow for identification. After sequencing, strain FF9T exhibited 98.18% similarity with the 16SrRNA sequence of Paenibacillus uliginis. A polyphasic study of phenotypic and genomic analyses showed that strain FF9T is Gram variable, catalase-positive, and presents a genome of 4,569,428 bp (one chromosome but no plasmid) with 4,427genes (4,352 protein-coding and 75 RNA genes (including 3 rRNA operons). The G+C content is 45.7%. On the basis of these genomic and phenotypic data analyses, we propose the creation of Paenibacillus dakarensis strain FF9T.

Keywords: Culturomics, genome, *Paenibacillus dakarensis*, taxono-genomics
Original Submission: 15 October 2015; **Revised Submission:** 18 December 2015; **Accepted:** 14 January 2016
Article published online: 22 January 2016

Corresponding author: F. Fenollar, Aix-Marseille Université, URMITE, UM63, CNRS 7278, IRD 198, Inserm U1095, Faculté de médecine, 27 Boulevard Jean Moulin, 13385 Marseille cedex 05, France
E-mail: florence.fenollar@univ-amu.fr

Introduction

Paenibacillus species were originally classified within the *Bacillus* genus [1]. Members of this genus are often Gram variable, facultatively anaerobic, and endospore forming. These bacteria are frequently isolated from environments such as soil, water, vegetable matter, forage, larvae and insects but could be detected from clinical samples [2–4]. Bacteria belonging to this genus are able to produce polysaccharide-degrading enzymes and proteases [5], are beneficial to agriculture and horticulture and have industrial and medical applications [6]. Some species of this genus may be involved in human infections [7–10]. Currently this genus includes 165 validly published species and four subspecies [11].

Recently high-throughput genome sequencing and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) analyses of bacteria have given unprecedented access to an abundance of genetic and proteomic information [12,13]. Thus, a polyphasic approach is currently proposed to describe new bacterial taxa that includes their genome sequence, MALDI-TOF spectrum and major phenotypic characteristics such as Gram staining, culture, metabolic characteristics, habitat and, if applicable, pathogenicity [14].

The strain FF9^T (= CSUR P1429 = DSM 29777) was isolated from a blood culture of a 16-month-old child presenting at the Hôpital Principal de Dakar, Senegal. Strain FF9^T is a Gram-variable bacterium, facultatively anaerobic, motile and rod shaped.

Here we present a summary classification and a set of features for *Paenibacillus dakarensis* sp. nov., together with a description of the complete genome sequencing and annotation.

http://dx.doi.org/10.1016/j.nmni.2016.01.011

These characteristics support the circumscription of the species *Paenibacillus dakarensis*.

Classification and features

In March 2014 a blood culture was performed on a 16-month-old child presenting at the Hôpital Principal de Dakar, Senegal. Strain FF9^T (Table 1) was isolated from this blood culture by culture on 5% sheep's blood–enriched Columbia agar (bioMérieux, Marcy l'Etoile, France). Identification was not obtained using MALDI-TOF because the scores obtained by this strain were low [23].

Moreover, strain FF9 exhibited 98.18% 16S rRNA sequence similarity with *Paenibacillus uliginis* [24] (GenBank accession no. FN556467), the phylogenetically closest bacterial species with standing in nomenclature (Fig. 1). These values were lower than the 98.7% 16S rRNA sequence threshold recommended by Meier-Kolthoff *et al.* [25] in 2013 to delineate a new species within the *Firmicutes* phylum without carrying out DNA-DNA hybridization. Different growth temperatures (25, 28, 37, 45 and 56°C) were tested. Growth was obtained between 28 and 37°C, with optimal growth occurring at 37°C. Growth of the strain was also tested under anaerobic and microaerophilic conditions using GENbag anaer and GENbag microaer systems (bioMérieux), respectively, and under aerobic conditions, with or without 5% CO_2. Optimal growth was observed under aerobic conditions, but weak growth was observed under anaerobic and microaerophilic conditions. Strain FF9 shows transparent, white, small colonies on 5% sheep's blood–enriched Columbia agar (bioMérieux) approximately 1 mm in diameter. A motility test was positive. Cells are Gram-variable, endospore-forming rods with rounded ends (Fig. 2) and have a mean diameter of 0.6 µm (range, 0.5–0.7 µm) and a mean length of 2.8 µm (range, 2.1–3.5 µm) (Fig. 3).

TABLE 1. Classification and general features of *Paenibacillus dakarensis* strain FF9^T

MIGS ID	Property	Term	Evidence code[a]
	Current classification	Domain: *Bacteria*	TAS [15]
		Phylum: *Firmicutes*	TAS [16,17]
		Class: *Firmibacteria*	TAS [18,19]
		Order: *Bacillales*	TAS [15–20]
		Family: *Paenibacillaceae*	TAS [18–21]
		Genus: *Paenibacillus*	TAS [1]
		Species: *Paenibacillus dakarensis*	IDA
		Type strain: FF9	IDA
	Gram stain	Variable	IDA
	Cell shape	Rods	IDA
	Motility	Motile	IDA
	Sporulation	Non–spore forming	IDA
	Temperature range	28–37°C	IDA
	Optimum temperature	37°C	IDA
	pH range; optimum	7.3–8.2; 7.7	
	Carbon source	Unknown	
MIGS-6	Habitat	Human blood	IDA
MIGS-6.3	Salinity	Unknown	
MIGS-22	Oxygen requirement	Facultative anaerobic	IDA
MIGS-15	Biotic relationship	Free-living	IDA
MIGS-14	Pathogenicity	Unknown	
MIGS-4	Geographic location	Senegal	IDA
MIGS-5	Sample collection	March 2014	IDA
MIGS-4.1	Latitude	14.6937000	IDA
MIGS-4.1	Longitude	−17.4440600	IDA
MIGS-4.4	Altitude	12 m above sea level	IDA

MIGS, minimum information about a genome sequence.
[a]Evidence codes are as follows: IDA, inferred from direct assay; TAS, traceable author statement (i.e. a direct report exists in the literature); NAS, nontraceable 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 (http://www.geneontology.org/GO.evidence.shtml) [22]. If the evidence code is IDA, then the property should have been directly observed, for the purpose of this specific publication, for a live isolate by one of the authors, or by an expert or reputable institution mentioned in the acknowledgements.

Paenibacillus dakarensis is catalase positive and oxidase negative. Using an API 50CH strip (bioMérieux), positive reactions were observed for D-ribose, D-glucose, D-mannose, N-acetyl-D-glucosamine, amygdaline, esculin, D-cellobiose, D-lactose, D-saccharose, D-trehalose, D- melezitose, gentiobiose and D-lyxose. Using a API 20NE strip (bioMérieux), positive reactions were observed for esculin, β-galactosidase, glucose and mannose. Using a API ZYM strip (bioMérieux), negative reactions were observed for alkaline phosphatase, esterase, esterase-lipase, leucine arylamidase, acid phosphatase, naphthol-AS-BI-phosphohydrolase, cystine arylamidase, valine arylamidase, trypsin, α-glucosidase, β-glucosidase, α-galactosidase, β-galactosidase, β-glucuronidase, α-mannosidase, α-fucosidase and N-acetyl-β-glucosaminidase. Strain FF9^T is susceptible to amoxicillin/clavulanic acid, ticarcillin, ceftriaxone, cefalotin, imipenem, gentamicin and doxycycline but is resistant to penicillin, metronidazole and trimethoprim/sulfamethoxazole. The minimum inhibitory concentrations (MICs) for some antibiotics tested by *Paenibacillus dakarensis* strain FF9^T sp. nov. are listed in Table 2.

A comparison of phenotypic characteristics with *Paenibacillus polymyxa* [1], *Paenibacillus massiliensis* and *Paenibacillus sanguinis* [26] is summarized in Table 3.

MALDI-TOF protein analysis was performed using a Microflex LT (Bruker Daltonics, Leipzig, Germany), as previously reported [27,28]. The scores previously established by Bruker to identify or validate species compared to the instrument's database were applied. In short, a score $\geq$2.000 with a species with a validly published name allows for identification at the species level; a score $\geq$1.700 and <2.000 allows for identification at the genus level; and a score <1.700 does not allow for any identification to be made. We performed 12 distinct deposits from 12 isolated colonies of strain FF9^T. Two microlitres of matrix solution (saturated solution of α-cyano-4-hydroxycinnamic acid) in 50% acetonitrile and 2.5% trifluoroacetic acid were distributed on each smear and subjected to air drying for 5 minutes. The spectra from the 12 different colonies

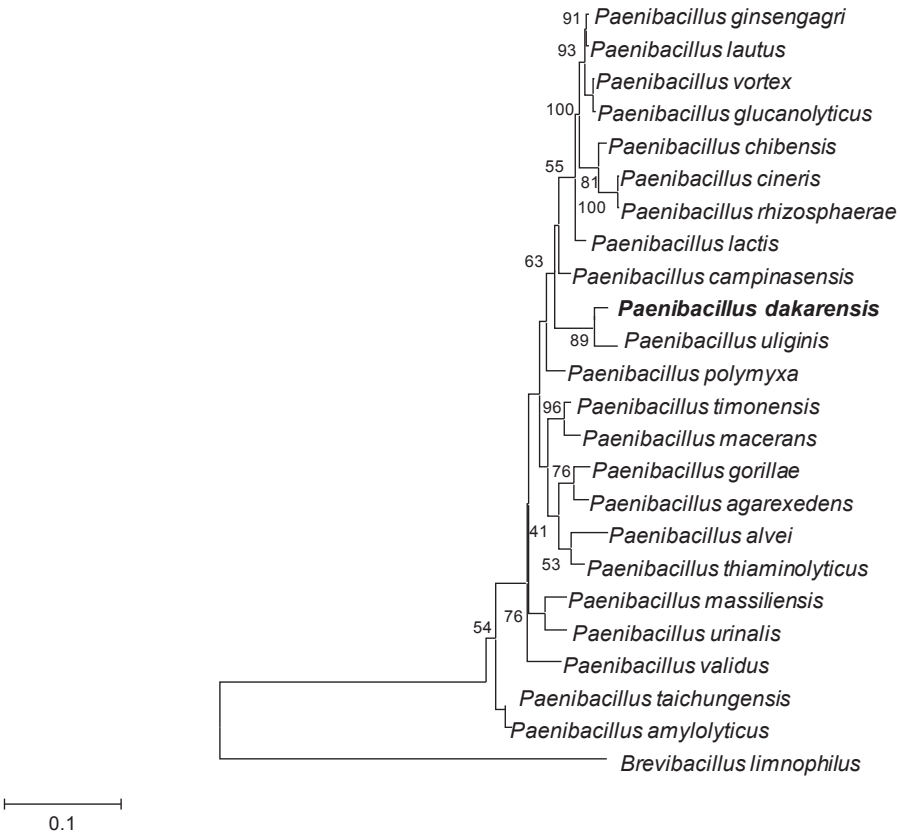

FIG. 1. Phylogenetic tree highlighting position of *Paenibacillus dakarensis* sp. nov. strain FF9^T relative to other type strains within *Paenibacillus* genus. Sequences were aligned using Clustal W, and phylogenetic inferences were obtained using maximum-likelihood method within MEGA 6. Numbers at nodes are percentages of bootstrap values obtained by repeating 1000 times analysis to generate majority consensus tree. *Brevibacillus limnophilus* strain was used as outgroup. Scale bar = 10% nucleotide sequence divergence.

were then imported into the MALDI BioTyper 2.0 software (Bruker) and analysed by standard pattern matching (with default parameter settings) against the main spectra of 6252 bacteria. Scores ranging from 1.225 to 1.456 were obtained for the FF9^T, suggesting that this strain was not a member of any known species. The reference mass spectrum from strain FF9^T was incremented in our database (Fig. 4). The gel view highlighted spectrum differences with other *Paenibacillaceae* species (Fig. 5).

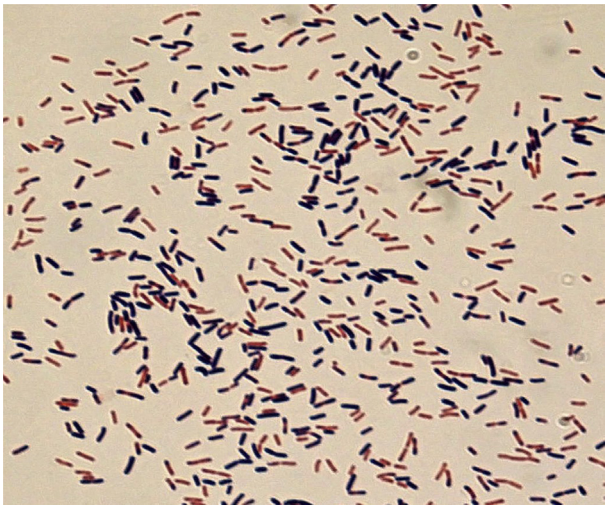

FIG. 2. Gram staining of *Paenibacillus dakarensis* sp. nov. strain FF9^T.

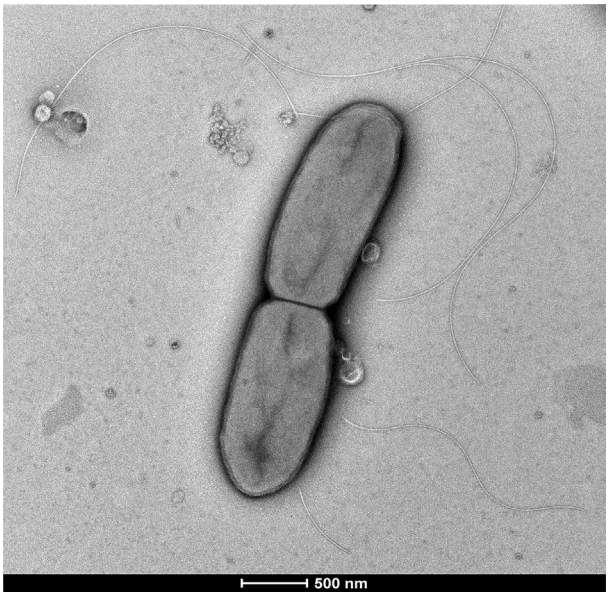

FIG. 3. Transmission electron microscopy of *Paenibacillus dakarensis* strain FF9^T. Cells were observed on Tecnai G2 transmission electron microscope operated at 200 keV. Scale bar = 500 nm.

TABLE 2. Antimicrobial susceptibility and MIC values of *Paenibacillus dakarensis* strain FF9^T sp. nov.

Antibiotic	MIC (mg/L)	Interpretation
Amoxicillin	2	Susceptible
Amoxicillin/clavulanic acid	2	Susceptible
Ticarcillin	2	Susceptible
Ceftriaxone	0.5	Susceptible
Imipenem	0.5	Susceptible
Ciprofloxacin	0.125	Susceptible
Gentamicin	1	Susceptible
Doxycycline	0.06	Susceptible

MIC, minimum inhibitory concentration.

Genome sequencing information

Genome project history

The organism was selected for sequencing on the basis of its phylogenetic position, 16S rRNA similarity and phenotypic differences with other members of the *Paenibacillaceae* family. There are more than 15 genomes for the *Paenibacillus* genus available in public genomic collections. Here we present the first *Paenibacillus dakarensis* sp. nov. genome. The GenBank accession number is CDSE01000001, and it consists of 102 contigs. Table 4 shows the project information and its association with minimum information about a genome sequence (MIGS) 2.0 compliance [29].

Growth conditions and DNA isolation

Paenibacillus dakarensis strain FF9^T (= CSUR P1429 = DSM 29777) was grown on 5% sheep's blood–enriched Columbia agar (bioMérieux) at 37°C. Bacteria grown on four petri dishes were resuspended in 5 × 100 µL of Tris-EDTA (TE) buffer; 150 µL of this suspension was diluted in 350 µL TE buffer 10×, 25 µL proteinase K and 50 µL sodium dodecyl sulfate for lysis treatment. This preparation was incubated overnight at 56°C. Extracted DNA was then purified using three successive phenol–chloroform extractions and ethanol precipitations at −20°C overnight. After centrifugation, DNA was suspended in 65 µL Elution buffer (EB) buffer. The genomic DNA concentration was measured at 452.7 ng/µL using the Qubit assay with the high sensitivity kit (Life Technologies, Carlsbad, CA, USA).

Genome sequencing and assembly

The mate pair library was prepared with 1.5 µg of genomic DNA using the Nextera mate pair Illumina guide (Illumina, San Diego, CA, USA). The genomic DNA sample was simultaneously fragmented and tagged with a mate pair junction adapter. The pattern of fragmentation was validated on an Agilent 2100 BioAnalyzer (Agilent Technologies, Santa Clara, CA, USA) using a DNA 7500 lab chip. The DNA fragments ranged in size from 1.5 to 11 kb, with an optimal size of 5.773 kb. No size selection was performed, and 600 ng of tagmented fragments were circularized. The circularized DNA was mechanically sheared into small fragments with an optimal size of 932 bp on the Covaris device S2 in T6 tubes (Covaris, Woburn, MA, USA). The library profile was visualized on a High Sensitivity Bioanalyzer LabChip (Agilent), and the final concentration library was measured at 19.07 nmol/L.

TABLE 3. Differential characteristics of *Paenibacillus dakarensis* strain FF9^T (data from this study) with *Paenibacillus polymyxa* [1], *Paenibacillus massiliensis* [26] and *Paenibacillus sanguinis* [26]

Character	*Paenibacillus dakarensis*	*Paenibacillus uliginis*	*Paenibacillus polymyxa*	*Paenibacillus massiliensis*	*Paenibacillus sanguinis*
Cell diameter (µm)	0.5	0.8	0.5	0.5	0.5
Oxygen requirement	Facultatively anaerobic	Facultatively anaerobic	Facultatively anaerobic	Facultatively anaerobic	Facultatively anaerobic
Gram stain	v	v	v	+	+
Motility	+	+	+	+	+
Endospore forming	+	+	+	+	+
Catalase	+	+	+	+	+
Oxidase	−	+	−	−	−
Alkaline phosphatase	−	NA	NA	NA	NA
Nitrate reductase	NA	(−/+)	+	+	+
Haemolysis	−	NA	+	−	−
Acid production from:					
Ribose	+	+	+	−	+
Glucose	+	+	+	−	+
Mannose	+	+	+	−	+
Rhamnose	−	−	−	−	−
Mannitol	−	−	+	+	+
Methyl β-D-xyloside	−	−	+	+	−
Methyl α-D-glucoside	−	+	+	+	−
N-acetyl-β-glucosaminidase	−	+	−	+	−
Utilization of:					
5-Keto-gluconate	−	−	NA	−	−
D-Xylose	−	−	+	−	+
D-Fructose	−	−	+	+	+
L-Fucose	−	−	−	−	−
D-Arabitol	−	−	−	−	−
Habitat	Blood culture	Fen peat of soil	Various soils	Blood culture	Blood culture

+, positive result; −, negative result; (−/+), strain-dependent reaction; v, variable result; NA, data not available.

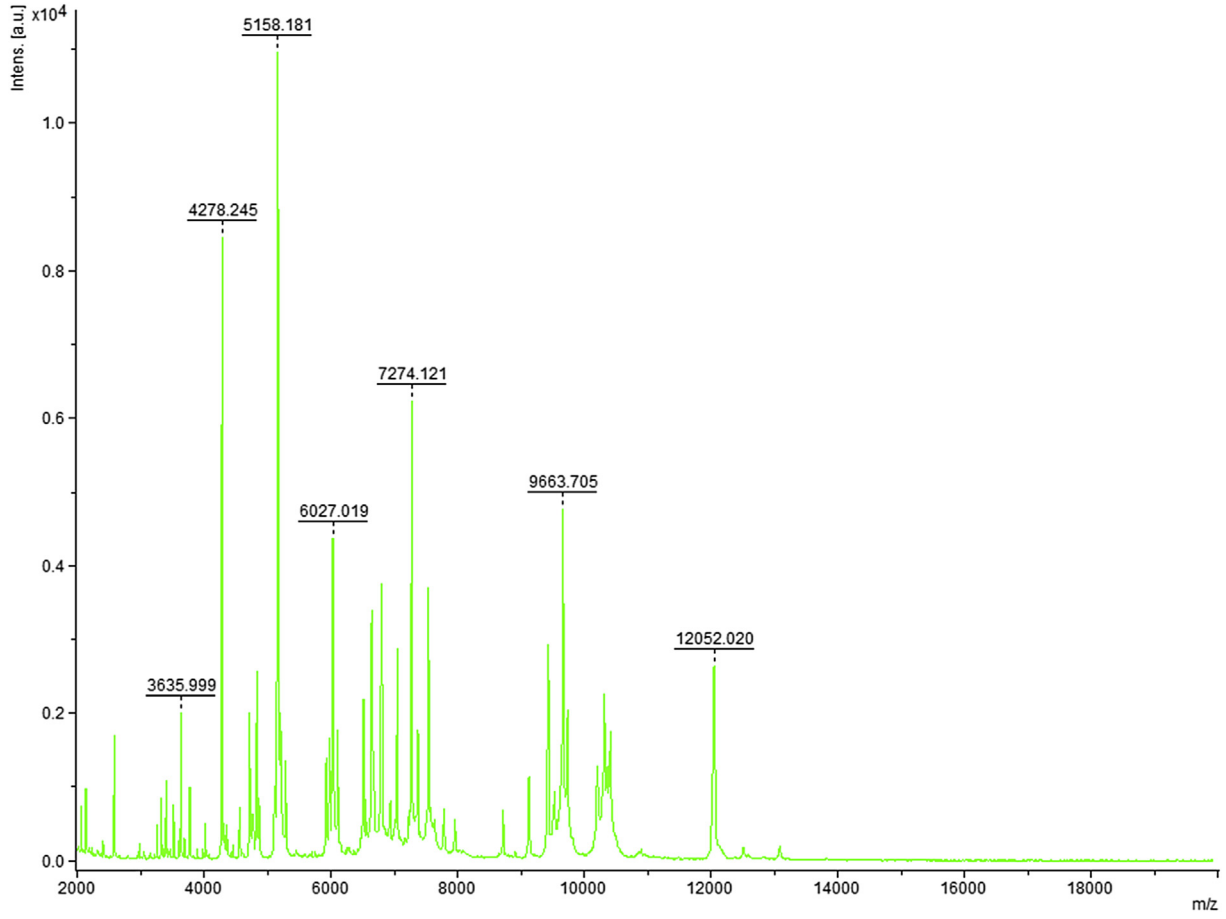

FIG. 4. Reference mass spectrum from *Paenibacillus dakarensis* strain FF9^T. Spectra from 12 individual colonies were compared and reference spectrum generated.

The libraries were normalized at 2 nM and pooled. After a denaturation step and dilution at 15 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 runs were performed in a single 39-hour run in a 2 × 251 bp read length.

Total information of 4.9 Gb was obtained from a 506K/mm^2 cluster density with a cluster passing quality control filters of 97% (9 954 000 clusters). Within this run, the index representation for *Paenibacillus dakarensis* FF9 was determined to be 8.98%. The 866 711 paired reads were filtered according to read quality. These reads were trimmed and then assembled.

Genome annotation

Open reading frames (ORFs) were predicted using Prodigal [30] 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 [31] and the Clusters of Orthologous Groups (COGs) database using BLASTP. The tRNAScanSE tool [32] was used to find tRNA genes, while ribosomal RNAs were found using RNAmmer [33] and BLASTn against the GenBank database. Lipoprotein signal peptides and the number of transmembrane helices were predicted using SignalP [34] and TMHMM [35] 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 been used in previous works to define ORFans. Artemis [36] was used for data management and DNA Plotter [37] for visualization of genomic features. The Mauve alignment tool (version 2.3.1) was used for multiple genomic sequence alignment [38]. To estimate the mean level of nucleotide sequence similarity at the genome level, we used the MAGI homemade software to calculate the average genomic identity of gene sequences (AGIOS) among compared genomes. Briefly, this software combines the Proteinortho software [39] for detecting orthologous proteins in pairwise

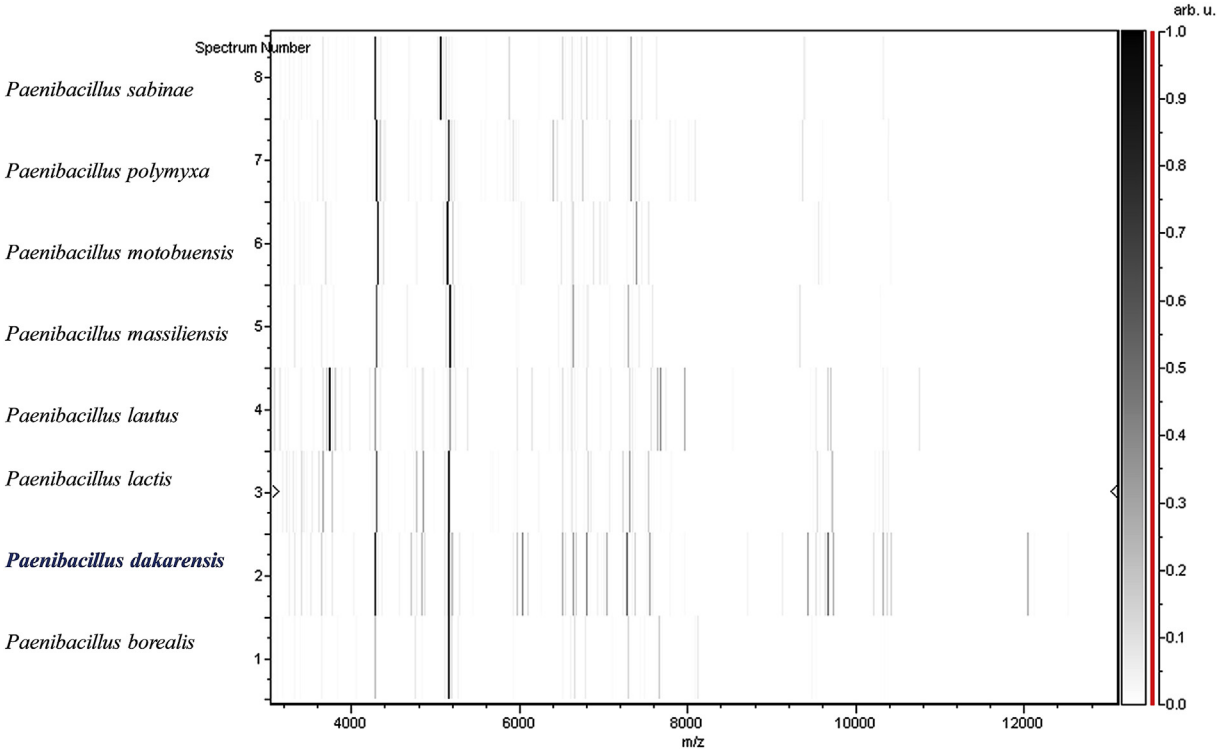

FIG. 5. Gel view comparing *Paenibacillus dakarensis* strain FF9^T to members of family *Paenibacillaceae*. Gel view displays raw spectra of all loaded spectrum files arranged in pseudo-gel-like look. *x*-axis records *m*/*z* value. Left *y*-axis displays running spectrum number originating from subsequent spectra loading. Peak intensity is expressed by greyscale scheme code. Color bar and right *y*-axis indicate relation between color peak, with peak intensity in arbitrary units. Displayed species are indicated on left.

genomic comparisons, then retrieves the corresponding genes and determines the mean percentage of nucleotide sequence identity among orthologous ORFs using the Needleman-Wunsch global alignment algorithm. Genomes from the *Paenibacillus* genus and closely related genera were used for the calculation of AGIOS values. The script created to calculate AGIOS values was named MAGi (Marseille Average genomic identity) and is written in Perl and Bioperl modules. Genome-to-Genome Distance Calculator (GGDC) analysis was also performed using the GGDC Web server (http://ggdc.dsmz.de) as previously reported [40,41]. Here, we compared the genome sequences of *P. dakarensis* strain FF9^T (GenBank accession no. CDSE01000001) with those of *Paenibacillus lactis* strain 154 (AGIP00000000), *Paenibacillus polymyxa* strain ATCC 842^T (AFOX00000000), *Paenibacillus massiliensis* strain 2301065^T (ARIL00000000), *Paenibacillus sabinae* strain T27^T (CP004078), *Paenibacillus borealis* strain DSM 13188^T (CP009285) and *Paenibacillus forsythiae* strain T98^T (ASSC00000000).

TABLE 4. **Project information**

MIGS ID	Property	Term
MIGS-31	Finishing quality	High-quality draft
MIGS-28	Libraries used	Mate-pair library
MIGS-29	Sequencing platforms	Illumina MiSeq
MIGS-31.2	Fold coverage	65.47×
MIGS-30	Assemblers	CLC GENOMICSWB4
MIGS-32	Gene calling method	Prodigal
	BioProject ID	PRJEB8435
	GenBank accession numbers	CDSE01000001-CDSE01000102
	GenBank Date of Release	9 April 2015
	Project relevance	MALDI-TOF implementation in Dakar

MALDI-TOF, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry; MIGS, minimum information about a genome sequence.

Genome properties

The genome of the *P. dakarensis* strain FF9^T is 4 569 428 bp long with a 45.7% G+C content (Fig. 6). Of the 4427 predicted genes, 4352 were protein-coding genes and 75 were RNA genes. Six rRNA genes (two 16S rRNA, two 23S rRNA and two 5S rRNA) and 69 predicted tRNA genes were identified in the genome. A total of 2820 genes (63.70%) were assigned a putative function. A total of 128 genes were identified as ORFans (2.89%). The remaining genes were annotated as hypothetical proteins. The properties and the statistics of the genome are summarized in Table 5. The distribution of genes into COGs functional categories is presented in Table 6.

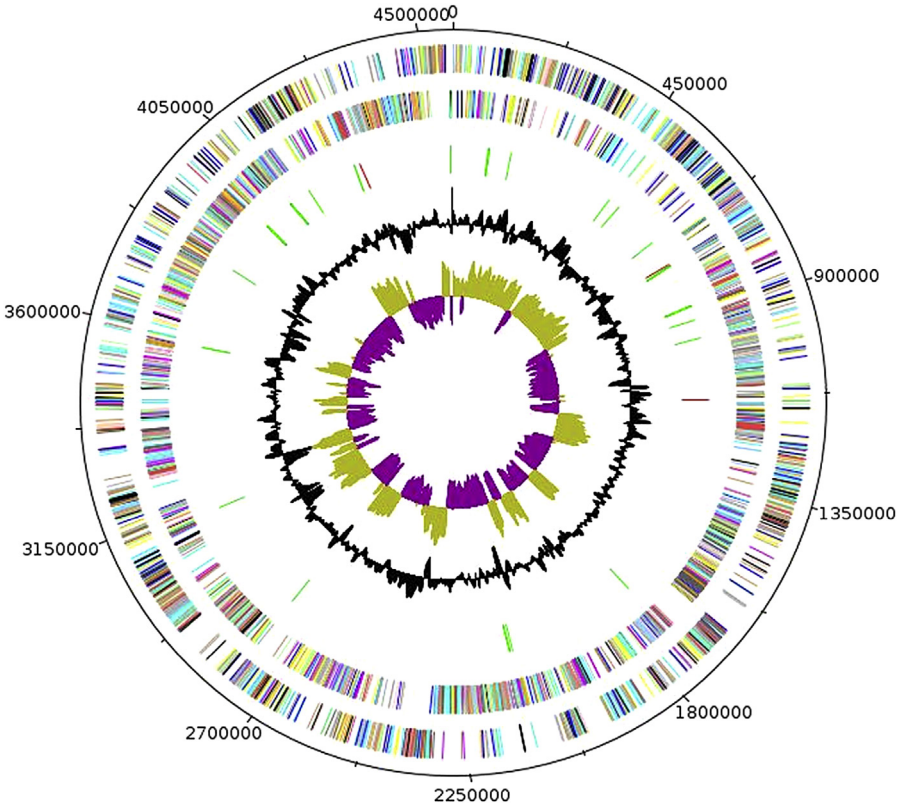

FIG. 6. Graphical circular map of *Paenibacillus dakarensis* strain FF9^T chromosome. From outside in, outer two circles show ORFs oriented in forward (coloured by COGs categories) and reverse (coloured by COGs categories) directions, respectively. Third circle marks tRNA genes (green). Fourth circle shows G+C% content plot. Innermost circle shows GC skew, with purple indicating negative values and olive positive values. COGs, Clusters of Orthologous Groups database; ORF, open reading frame.

Insights from genome sequence

Genomic comparison with other *Paenibacillus* species

The draft genome of *P. dakarensis* is smaller than that of *P. lactis, P. polymyxa, P. massiliensis, P. sabinae, P. borealis* and *P. forsythiae* (4.56, 6.81, 5.9, 6.39, 5.27, 8.16 and 5.08 Mb, respectively). The G+C content of *P. dakarensis* is higher than those of *P. polymyxa*

TABLE 5. Nucleotide content and gene count levels of the genome

	Genome (total)	
Attribute	**Value**	**% of total[a]**
Size (bp)	4 569 428	100
G+C content (bp)	2 056 242	45.7
Coding region (bp)	3 977 838	87.05
Total genes	4427	100
RNA genes	75	1.69
Protein-coding genes	4352	98.30
Genes with function prediction	3176	71.74
Genes assigned to COGs	2820	63.70
Genes with peptide signals	223	5.03
Genes with transmembrane helices	745	16.82
ORFan genes	128	2.89

COGs, Clusters of Orthologous Groups database.
[a]Total is based on either size of genome (bp) or total number of protein-coding genes in annotated genome.

TABLE 6. Number of genes associated with 25 general COGs functional categories

Code	Value	% of total[a]	Description
J	178	4.09	Translation
A	0	0	RNA processing and modification
K	334	7.67	Transcription
L	175	4.02	Replication, recombination and repair
B	0	0	Chromatin structure and dynamics
D	34	0.78	Cell cycle control, mitosis and meiosis
Y	0	0	Nuclear structure
V	88	2.02	Defense mechanisms
T	208	4.77	Signal transduction mechanisms
M	147	3.37	Cell wall/membrane biogenesis
N	64	1.47	Cell motility
Z	0	0	Cytoskeleton
W	0	0	Extracellular structures
U	45	1.03	Intracellular trafficking and secretion
O	111	2.55	Posttranslational modification, protein turnover, chaperones
C	156	3.58	Energy production and conversion
G	421	9.67	Carbohydrate transport and metabolism
E	271	6.22	Amino acid transport and metabolism
F	93	2.13	Nucleotide transport and metabolism
H	112	2.57	Coenzyme transport and metabolism
I	88	2.02	Lipid transport and metabolism
P	197	4.52	Inorganic ion transport and metabolism
Q	77	1.76	Secondary metabolites biosynthesis, transport and catabolism
R	531	12.20	General function prediction only
S	310	7.12	Function unknown
—	1532	35.20	Not in COGs

COGs, Clusters of Orthologous Groups database.
[a]Total is based on total number of protein-coding genes in annotated genome.

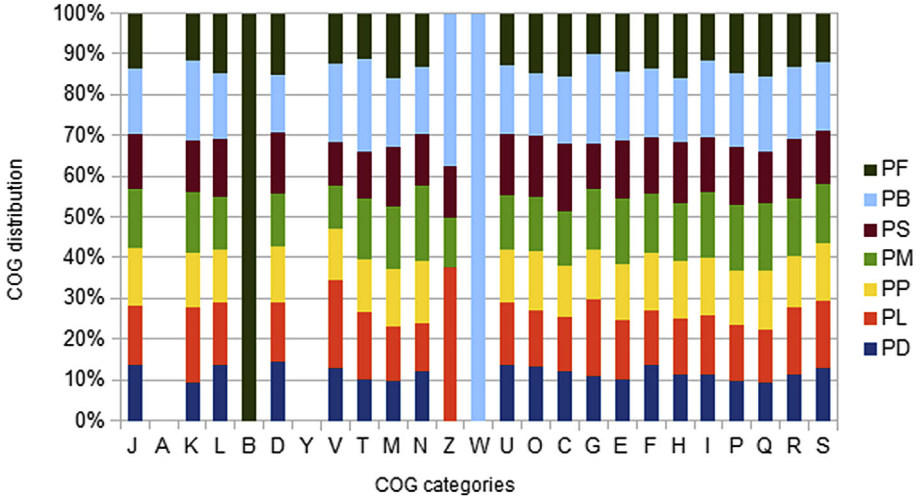

FIG. 7. Distribution of functional classes of predicted genes in genomes from various *Paenibacillus* spp. chromosomes according to clusters of orthologous groups of proteins. PA, *Paenibacillus antibiotocaphila;* PB, *Paenibacillus borealis;* PD, *Paenibacillus dakarensis;* PF, *Paenibacillus forsythiae;* PL, *Paenibacillus lactis;* PM, *Paenibacillus massiliensis;* PP, *Paenibacillus polymyxa;* PS, *Paenibacillus sabinae;* PSE, *Paenibacillus senegalense*.

TABLE 7. **Numbers of orthologous protein shared between genomes (upper right)**[a]

	PD	PL	PP	PM	PS	PB	PF
PD	**4352**	2627	2087	2159	2070	2408	1889
PL	73.59	**6149**	2463	2557	2340	2876	2147
PP	70.09	69.35	**5068**	2752	2279	2685	2138
PM	69.61	69.64	71.59	**5055**	2333	2843	2178
PS	69.76	71.12	69.27	69.33	**4788**	2904	2789
PB	69.74	70.32	69.27	69.46	74.49	**6213**	2656
PF	69.89	71.27	69.19	69.38	84.00	74.47	**5011**

PB, *Paenibacillus borealis;* PD, *Paenibacillus dakarensis;* PF, *Paenibacillus forsythia;* PL, *Paenibacillus lactis;* PM, *Paenibacillus massiliensis;* PP, *Paenibacillus polymyxa;* PS, *Paenibacillus sabinae.*
[a]Average percentage similarity of nucleotides corresponding to orthologous protein shared between genomes (lower left) and numbers of proteins per genome (bold).

(45.7 and 44.9%, respectively) but lower than those of *P. lactis, P. massiliensis, P. sabinae, P. borealis* and *P. forsythiae* (49.1, 48.5, 52.6, 51.4 and 52.9 respectively). The gene content of *P. dakarensis* is lower than those of *P. lactis, P. massiliensis, P. sabinae, P. borealis* and *P. forsythiae* (4427, 6234, 5206, 5193, 4896, 6382 and 5103 respectively). However, the distribution of genes into COGs categories was similar in all compared genomes (Fig. 7). In addition, *P. dakarensis* shared 4352, 6149, 5068, 5055, 4788, 6213 and 5011 orthologous genes with *P. lactis, P. polymyxa, P. massiliensis, P. sabinae, P. borealis* and *P. forsythiae* respectively (Table 7). Among species with standing in nomenclature, AGIOS values ranged from 69.19% between *P. polymyxa* and *P. forsythiae* to 84.00% between *P. forsythiae* and *P. sabinae.*

Conclusion

On the basis of phenotypic, phylogenetic and genomic analyses, we formally propose the creation of *Paenibacillus dakarensis* sp. nov., which contains strain FF9^T. The strain was isolated from a blood sample taken from a 16-month-old Senegalese child presenting at the Hôpital Principal de Dakar.

Taxonomic and nomenclatural proposals

Description of *Paenibacillus dakarensis* strain FF9^T sp. nov.

Paenibacillus dakarensis (da.kar.e′n.se. L. gen. neutr. n. *dakarensis,* or originating from Dakar, the capital of Senegal, where the type strain was isolated). The strain FF9^T is a facultative anaerobic, Gram variable bacterium, with small, white colonies on 5% sheep's blood–enriched Columbia agar. A motility test was positive. Cells have a mean diameter of 0.6 µm (range, 0.5–0.7 µm) and a mean length of 2.8 µm (range, 2.1–3.5 µm). The strain FF9^T is oxidase negative and catalase positive. Positive reactions were observed for D-ribose, D-glucose, D-mannose, N-acetyl-D-glucosamine, amygdaline, esculin, D-cellobiose, D-lactose, D-saccharose, D-trehalose, D-melezitose, gentiobiose, D-lyxose and β-galactosidase. *Paenibacillus dakarensis* strain FF9^T is susceptible to amoxicillin/clavulanic acid, ticarcillin, ceftriaxone, cefalotin, imipenem, gentamicin and doxycycline but resistant to metronidazole, penicillin and trimethoprim/sulfamethoxazole. The G+C content of the

genome is 45.7%. The 16S rRNA and genome sequences are deposited in GenBank under accession numbers LM652718 and CDSE01000001, respectively. The type strain FF9^T (= CSUR P1429 = DSM 29777) was isolated from a blood sample taken from a 16-month-old child presenting at the Hôpital Principal de Dakar, Senegal.

Conflict of interest

None declared.

Acknowledgements

We thank C. Couderc for help performing the MALDI-TOF analysis. We also thank F. Di Pinto for taking the electron microscope photos. The authors thank the Xegen Company (http://www.xegen.fr/) for automating the genomic annotation process. This study was funded by the Fondation Méditerranée Infection.

References

[1] Ash C, Priest FG, Collins MD. Molecular identification of rRNA group 3 bacilli (Ash, Farrow, Wallbanks and Collins) using a PCR probe test. Proposal for the creation of a new genus Paenibacillus. Antonie Van Leeuwenhoek 1993;64:253–60.

[2] Lal S, Tabacchioni S. Ecology and biotechnological potential of *Paenibacillus polymyxa*: a minireview. Indian J Microbiol 2009;49:2–10.

[3] Montes MJ, Mercade E, Bozal N, Guinea J. *Paenibacillus antarcticus* sp. nov., a novel psychrotolerant organism from the Antarctic environment. Int J Syst Evol Microbiol 2004;54:1521–6.

[4] Glaeser SP, Falsen E, Busse HJ, Kämpfer P. *Paenibacillus vulneris* sp. nov., isolated from a necrotic wound. Int J Syst Evol Microbiol 2013;63:777–82.

[5] Konishi J, Maruhashi K. 2-(2′-Hydroxyphenyl) benzene sulfinate desulfinase from the thermophilic desulfurizing bacterium *Paenibacillus* sp. strain A11-2: purification and characterization. Appl Microbiol Biotechnol 2003;62:356–61.

[6] McSpadden Gardener BB. Ecology of *Bacillus* and *Paenibacillus* spp. in agricultural systems. Phytopathology 2004;94:1252–8.

[7] Ouyang J, Pei Z, Lutwick L, Dalal S, Yang L, Cassai N, et al. Case report: *Paenibacillus thiaminolyticus*: a new cause of human infection, inducing bacteremia in a patient on hemodialysis. Ann Clin Lab Sci 2008;38:393–400.

[8] Roux V, Fenner L, Raoult D. *Paenibacillus provencensis* sp. nov., isolated from human cerebrospinal fluid, and *Paenibacillus urinalis* sp. nov., isolated from human urine. Int J Syst Evol Microbiol 2008;58:682–7.

[9] Leão RS, Pereira RHV, Ferreira AG, Lima AN, Albano RM, Marques EA. First report of *Paenibacillus cineris* from a patient with cystic fibrosis. Diagn Microbiol Infect Dis 2010;66:101–3.

[10] Anikpeh YF, Keller P, Bloemberg GV, Grünenfelder J, Zinkernagel AS. Spacecraft bacterium, *Paenibacillus pasadenensis*, causing wound infection in humans. BMJ Case Rep 2010:2010.

[11] Parte AC. LPSN—list of prokaryotic names with standing in nomenclature. Nucleic Acids Res 2014;42:D613–6.

[12] 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 new bacterial species. Int J Syst Evol Microbiol 2014;64:384–91.

[13] Sentausa E, Fournier PE. Advantages and limitations of genomics in prokaryotic taxonomy. Clin Microbiol Infect 2013;19:790–5.

[14] Lo CI, Padhmanabhan R, Mediannikov O, Terras J, Robert C, Faye N, et al. High-quality genome sequence and description of *Bacillus dielmoensis* strain FF4(T) sp. nov. Stand Genomic Sci 2015;10:41.

[15] Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains *Archaea, Bacteria*, and *Eukarya*. Proc Natl Acad Sci U S A 1990;87:4576–9.

[16] Skerman VBD, Sneath PHA. Approved list of bacterial names. Int J Syst Bacteriol 1980;30:225–420.

[17] Garrity GM, Holt J. The road map to the manual. In: Garrity GM, Boone DR, Castenholz RW, editors. Bergey's manual of systematic bacteriology. 2nd ed., vol. 1. New York: Springer; 2001. p. 119–69.

[18] List Editor. List of new names and new combinations previously effectively, but not validly, published. List no. 132. Int J Syst Evol Microbiol 2010;60:469–72.

[19] Ludwig W, Schleifer KH, Whitman WB. Class I. Bacilli class nov. In: De Vos P, Garrity G, Jones D, Krieg NR, Ludwig W, Rainey F, et al., editors. Bergey's manual of systematic bacteriology. 2nd ed., vol. 3. New York: Springer; 2009. p. 19–20.

[20] Prevot AR. In: Hauduroy P, Ehringer G, Guillot G, Magrou J, Prévot AR, Rosset, et al., editors. Dictionnaire des bactéries pathogènes. 2nd ed., Paris: Masson; 1953. p. 1–692.

[21] De Vos P, Ludwig W, Schleifer KH, Whitman WB. Family IV. *Paenibacillaceae* fam. nov. In: De Vos P, Garrity GM, Jones D, et al., editors. Bergey's manual of systematic bacteriology. The firmicutes. 2nd ed., vol. 3. New York: Springer; 2009. p. 269.

[22] 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.

[23] Fall B, Lo CI, Samb-Ba B, Perrot N, Diawara S, Gueye NW, et al. The ongoing revolution of MALDI-TOF mass spectrometry for microbiology reaches tropical Africa. Am J Trop Med Hyg 2015;92:641–7.

[24] Behrendt U, Schumann P, Stieglmeier M, Pukall R, Augustin J, Spröer C, et al. Characterization of heterotrophic nitrifying bacteria with respiratory ammonification and denitrification activity—description of *Paenibacillus uliginis* sp. nov., an inhabitant of fen peat soil and *Paenibacillus purispatii* sp. nov., isolated from a spacecraft assembly clean room. Syst Appl Microbiol 2010;33:328–36.

[25] Meier-Kolthoff JP, Göker M, Spröer C, Klenk HP. When should a DDH experiment be mandatory in microbial taxonomy? Arch Microbiol 2013;195:413–8.

[26] Roux V, Raoult D. *Paenibacillus massiliensis* sp. nov., *Paenibacillus sanguinis* sp. nov. and *Paenibacillus timonensis* sp. nov., isolated from blood cultures. Int J Syst Evol Microbiol 2004;54:1049–54.

[27] 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.

[28] Seng P, Rolain JM, Fournier PE, La Scola B, Drancourt M, Raoult D. MALDI-TOF–mass spectrometry applications in clinical microbiology. Future Microbiol 2010;5:1733–54.

[29] 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.

[30] Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 2010;11:119.

[31] Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW. GenBank. Nucleic Acids Res 2012;40:48–53.

[32] 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.
[33] Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 2007;35:3100–8.
[34] Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 2004;340:783–95.
[35] 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.
[36] Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, et al. Artemis: sequence visualization and annotation. Bioinformatics 2000;16:944–5.
[37] Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J. DNAPlotter: circular and linear interactive genome visualization. Bioinformatics 2009;25:119–20.
[38] Darling AC, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 2004;14:1394–403.
[39] 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.
[40] Auch AF, Klenk HP, Göker M. Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. Stand Genomic Sci 2010;2:142–8.
[41] 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.