

Clostridium niakharensense sp. nov., a novel species isolated from the breast milk of a Senegalese mother breastfeeding a healthy child

Salematou SARR

IHU-Méditerranée Infection

Ousmane NDOUR

IHU-Méditerranée Infection

Adama MBOW

IHU-Méditerranée Infection

Mamadou BEYE

IHU-Méditerranée Infection

Maryam TIDJANI ALOU

IHU-Méditerranée Infection

Anissa IDRISSE ABDOULAYE

IHU Méditerranée Infection

Georges DIATTA

IRD, MINES, Campus International IRD-UCAD

Hubert BASSENE

IRD, MINES, Campus International IRD-UCAD

Sara BELLALI

IHU Méditerranée Infection

Stephane ALIBAR

IHU-Méditerranée Infection

Simon ROBINNE

IHU Méditerranée Infection

Nicolas ORAIN

IHU Méditerranée Infection

Clémence THIAKANE

ICP Saint Martin Health Center

Cheikh SOKHNA

IHU-Méditerranée Infection

Makhtar CAMARA

Université Cheikh Anta Diop

Matthieu MILLION

matthieumillion@gmail.com

IHU Méditerranée Infection

Research Article

Keywords: breast milk, Clostridium niakharensis, microbiota, new species, Senegal, taxonogenomics

Posted Date: October 18th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11088235/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: Competing interest reported. Conflicts of Interest: Sara Bellali received personal fees through a collaborative contract with Hitachi High-Tech Corporation. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abstract

Strain Marseille-QA1073[☐] was identified from the breast milk of a healthy mother living in Senegal. Strain Marseille-QA1073[☐] is an endospore-forming, motile, Gram-stain-positive and anaerobic bacterium with fusiform cells. It is negative for catalase and oxidase, developing at temperatures between 20°C and 45°C under anaerobic conditions with weak growth under microaerophilic conditions. Based on 16S rRNA gene sequence analysis, strain Marseille-QA1073[☐] was most closely related to *Clostridium argentinense* ATCC 27322^T and *Clostridium faecium* N37[☐] (~ 99.00% similarity); as the 16S rRNA gene did not discriminate the strain from these relatives, species delineation relied on genome-based indices. With a G + C content of 28.55 mol%, and a genome size of 4.45 Mbp, strain Marseille-QA1073[☐] shared an average nucleotide identity (ANIb) of 94.3% and a digital DNA-DNA hybridization (dDDH) values of 55.3% with *Clostridium argentinense*, (its closest relative) and 87.4% / 34.1% with *Clostridium faecium*, all well below the 95% (ANI) and 70% (dDDH) species delineation thresholds. Genomic screening revealed a chromosomal *vanF*-type operon without any virulence factors or botulinum neurotoxin (*bont*) genes. These genomic and phenotypic data confirm that strain Marseille-QA1073[☐] represents a distinct species within the genus *Clostridium*, for which we propose the name *Clostridium niakharensense* sp. nov. (Type strain Marseille-QA1073[☐] = CSUR QA1073[☐] = CECT 31321[☐]), deposited in GenBank under accession number JBSVQM000000000.

Introduction

The genus *Clostridium* was first described in 1880 by Prazmowski (**Prazmowski, 1880**). To define the genus, Prazmowski relied on purely phenotypic criteria, such as a Gram-stain-positive cell wall structure, strictly anaerobic metabolism, the ability to form endospores, and an inability to reduce sulfate (**Wiegel et al., 2006**). For many years, this classification system led to any endospore-forming anaerobic rod exhibiting these characteristics being assigned to the genus *Clostridium*, so that a large and heterogeneous set of species accumulated under this name (**Rainey et al., 2009**). Consequently, the need for a polyphasic approach to achieve a more consistent classification became increasingly apparent. In 1975, Johnson and Francis grouped reference strains representing 56 *Clostridium* species into four clusters, primarily based on 23S rRNA homology (rRNA competition experiments) with DNA and G + C content as a complementary criterion through a purely phylogenetic approach (**Johnson et al., 1975**). Notably, the genus *Sarcina* (**Skerman et al., 1980**), which antedates *Clostridium* (**Prazmowski, 1880**), comprises species that fall within Cluster I, i.e. among the "true" clostridia, raising a question nomenclatural priority (**Skerman et al., 1980; Goodsir and Wilson, 1842**). Subsequently, in 1994, Collins et al. revisited this categorization in greater depth based on 16S rRNA gene sequences (**Collins et al., 1994**). This general classification system, previously based on phenotypic and phylogenetic approaches, was further refined by the work of Gupta and Gao. They utilised phylogenetic and comparative analyses of *Clostridium* genome sequences to clarify the evolutionary relationships between species within the genus (**Gupta and Gao, 2009; Yutin and Galperin, 2013**). However, in 2015, Rossi-Tamisier et al. observed that these classifications did not sufficiently meet the identity thresholds established for 16S

rRNA gene sequences (**Rossi-Tamisier et al., 2015**). This led to a major re-categorisation by Lawson and Rainey in 2016, which proposed formally restricting the genus *Clostridium* to the type species *C. butyricum* and related species within Cluster I (**Lawson, 2016; Lawson and Rainey, 2016**). This reorganisation resulted in the transfer of several species into and out of the genus *Clostridium* and the creation of the genera *Hathewayia* and *Clostridioides* (**Lawson et al., 2016**).

In 2019, genomic analysis of 779 strains confirmed the extreme heterogeneity of the group. This work highlighted major differences among members previously classified as a single species and allowed for a reassessment of taxonomy to address inconsistencies in toxin dynamics and unique central metabolic characteristics within Cluster I agents (**Cruz-Morales et al., 2019**). Currently, following these successive reclassifications, the genus *Clostridium* comprises 172 species with validly published names according to the List of Prokaryotic names with Standing in Nomenclature (LPSN, <https://lpsn.dsmz.de/genus/clostridium>) last accessed on August 22, 2026) (**Göker et al., 2025**).

Many species have been isolated from various environments, including milk (**Cerf and Bergère, 1968**), soil (**Suen et al., 1988**), marine sediments, and water, as well as from the digestive tracts of humans and animals (**Gilroy et al., 2021**). This includes significant pathogens such as *C. tetani* (**Brüggemann et al., 2003**) and *C. botulinum* (**Hauschild, 2018**), alongside organisms utilized in industrial and biotechnological applications (**Grosse-Herrenthey et al., 2008**). Here, we propose the description of a novel species within the genus *Clostridium*, isolated from human breast milk by culturomics using the taxonogenomics approach.

Materials and methods

Ethics approval and Sample Collection

Following administrative authorization ethical and scientific approval from the National Ethics Committee for Health Research (CNERS) of the Senegalese Ministry of Health and Public Hygiene (MSPH) under protocol no. 00000233/MSAS/CNERS/SP (approved on September 11, valid 2022–2024), and in accordance with the Declaration of Helsinki and the Nagoya Protocol, strain Marseille-QA1073 was isolated as part of a study focused on the identification of probiotic strains from the breast milk microbiota of Senegalese woman. After obtaining free and informed consent, 10 to 15 mL of breast milk was collected from a mother breastfeeding healthy child at the Niakhar site in the Fatick region of Senegal. The samples were immediately transported to the laboratory of the Institut de Recherche pour le Développement (IRD) in Dakar at + 4°C. Upon arrival, they were aliquoted into cryotubes, supplemented with 50 to 100 µL of C-Top preservation medium (Culture Top, Marseille, France), and stored at -80°C. Subsequently, the samples were shipped on dry ice to the Méditerranée Infection University Hospital Institute (IHU) in Marseille, France, for culture, sequencing and phenotypic characterization.

Isolation and identification of the strain

Breast milk was collected after cleansing of the breast with alcohol and preserved in C-Top medium (Culture Top, Marseille, France), whose antioxidants maintain bacterial viability. Strain Marseille-QA1073 was isolated by culturomics (**Lagier et al., 2018**): 100 µL of breast milk from a healthy breastfeeding mother was inoculated into YCFA (Yeast Casitone Fatty Acid) broth modified with 18 g/L lactose and enriched anaerobically at 37°C for 21 days, a prolonged anaerobic pre-incubation used in culturomics to recover fastidious, slow-growing anaerobes, rather than single passive incubation. To ensure that any growth originated from the milk and not from the medium, a sterility control was systematically performed prior to inoculation: an aliquot of the uninoculated YCFA broth was subcultured onto Columbia agar supplemented with 5% sheep blood (COS) (bioMérieux, Craponne, France) and incubated under the same conditions; only sterile media were used (no growth). After enrichment the culture was plated onto COS agar and incubated anaerobically at 37°C for 72 hours. Sixteen colonies were subcultured to obtain pure isolates, identified using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) with the Sirius system (Biotyper® Sirius, Bruker, Bremen, Germany). The spectra being compared using the MALDI Biotyper Compass software (Bruker) against the BDAL 12,438 MSPs (version 13) database as well as our in-house spectra database available at the MEPHI repository (<https://ihu-medinfection.com/urms-data-base/>). Identification scores were interpreted according to the manufacturer's thresholds: a score ≥ 2.0 was considered reliable identification to the species level (green), a score between 1.7 and 2.0 indicated identification to the genus level only (yellow), and a score < 1.7 was interpreted as no reliable identification (red).

Phenotypic and biochemical characteristics

Colony morphology of strain Marseille-QA1073 was recorded on blood agar after 24 to 72 hours of incubation, including size, shape, colour (opacity, transparency, or pigmentation), surface appearance (rough or smooth), odour and haemolysis. A panel of standardized phenotypic tests was applied. Gram staining was used to determine cell-wall type and cell arrangement. Motility was assessed on fresh colonies mounted between slide and coverslip. Sporulation was evaluated by subjecting a pure colony suspension in 1 mL of phosphate-buffered saline (PBS), to a heat shock at 80°C for 20 minutes. Catalase activity was tested using hydrogen peroxide (bioMérieux, Marcy-l'Étoile, France) and oxidase activity with oxidase discs (MASTDISCS™ ID, Mast Group). Growth was examined at 20°C, 28°C, 37°C, 45°C, and 56°C under aerobic, microaerophilic, and anaerobic conditions using the GENbag anaer and microaer systems (bioMérieux) to determine the optimal growth parameters. Tolerance to acidity and alkalinity was assessed in media adjusted pH 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8.0 and 8.5. Salt tolerance was assessed in media supplemented with 7.5%, 10%, 15%, and 20% (w/v) NaCl.

Carbohydrate fermentation and enzymatic activities were determined using API 20 A and API ZYM strips (bioMérieux) according to the manufacturer's instructions, using a bacterial suspension adjusted to 5–6 McFarland. Antimicrobial susceptibility was assessed and minimum inhibitor concentrations (MICs) were determined using gradient diffusion strips (Etest®, bioMérieux), interpreted according to the EUCAST 2025 guidelines (<https://www.eucast.org/bacteria/clinical-breakpoints-and->

interpretation/clinical-breakpoint-tables/). The following antibiotics were tested: amikacin, amoxicillin, benzylpenicillin, ceftazidime, ceftriaxone, ciprofloxacin, clindamycin, colistin, cotrimoxazole, doxycycline, fosfomycin, gentamicin, imipenem, linezolid, nitrofurantoin, oxacillin, rifampicin, teicoplanin, tobramycin, vancomycin.

Analysis of fatty acid methyl esters (FAME) by GC/MS

Cellular fatty acid methyl ester (FAME) analysis was performed by gas chromatography/mass spectrometry (GC/MS). Two samples were prepared, each containing approximately 150 mg of bacterial biomass harvested from several culture plates. Fatty acid methyl esters were prepared according to the method described by Sasser (2006). GC/MS analyses were carried out as previously described (**Dione et al., 2016**). Briefly, FAMEs were separated using an Elite 5-MS column and monitored by mass spectrometry (Clarus 500 - SQ 8 S, PerkinElmer, Courtaboeuf, France). Spectral database searches were performed using MS Search 2.0 operated with the Standard Reference Database 1A (NIST, Gaithersburg, USA) and the FAME mass spectral database (Wiley, Chichester, UK).

Morphological characterisation by scanning electron microscopy

The morphology of strain Marseille-QA1073 was examined using a Hitachi SU5000 scanning electron microscope (SEM) (Hitachi High-Tech Corporation, Japan). Fresh bacterial colonies were fixed in 2.5% glutaraldehyde (Electron Microscopy Sciences, USA) prepared in 0.1 M sodium cacodylate buffer (Thermo Fisher Scientific, USA) and stained with 10% phosphotungstic acid (PTA) (pH 7.4; Sigma-Aldrich, USA) solution at neutral pH (pH 7.4). Cell suspensions were Cytocentrifugated onto glass slides at 800 rpm for 7 minutes using a Cytospin 4 centrifuge (Thermo Electron Corporation - Shandon, United Kingdom). The preparation was subsequently coated with thin layer of platinum-palladium using an MC1000 ion sputter coater (Hitachi High-Tech, Japan) prior imaging. SEM micrographs were acquired at magnifications ranging from 1,000× to 30,000× using both backscattered electron (BSE) and secondary electron (SE) detectors to provide compositional contrast and complementary surface. Detailed acquisition parameters, including instrument type, acceleration voltage, working distance, magnification, and detection mode, are provided for each corresponding micrograph.

Genome sequencing, assembly and annotation

Genomic DNA was extracted with the EZ2 DNA Tissue Kit on an EZ2 Advanced XL workstation (Qiagen, Hilden, Germany), after mechanical (glass bead) and enzymatic (lysozyme) lysis and quantified with the Qubit dsDNA HS assay (Thermo Fisher Scientific).

The genome was sequenced using a hybrid strategy combining Illumina short reads (2 x 250 bp paired-end, Nextera XT library, MiSeq) and Oxford Nanopore long reads (using native barcoding EXP-NBD114.96,

FLO-MIN114 flow cell R10.4.1, Gridlon; ONT, Oxford, UK) for 72 hours, with DNA purification carried out using CleanNA beads (Coenecoop, Netherlands)

Genomic analysis

Illumina reads quality was assessed with FastQC v0.11.9 (**Andrews, 2010**) and low-quality bases were trimmed with Trimmomatic v0.39 (**Bolger et al., 2014**). For Oxford Nanopore reads, read-length distribution and quality were visualized using NanoPlot (**De Coster et al., 2018**) and High-quality reads were retained with Filtlong (<https://github.com/rrwick/Filtlong>). A hybrid assembly was generated with Unicycler v0.4.8 (**Wick et al., 2017**); redundant contigs contained within larger ones were then removed, and assembly completeness and contamination were assessed with CheckM2 v1.1.0 (**Chklovski et al., 2023**). The genome was annotated with Prokka v1.14.6 (**Seemann, 2014**) and functional categories were assigned against the Clusters of Orthologous Groups (COG) database using eggNOG-mapper v2.1.13 (**Huerta-Cepas et al., 2017; Cantalapiedra et al., 2021**) via the eggNOG database v5.0.2 (**Huerta-Cepas et al., 2019**). Transmembrane helices and signal peptides were predicted using TMHMM and SignalP respectively (**Teufel et al., 2022; Krogh et al., 2001**). Antibiotic-resistance genes were identified with ABRicate v1.4.0 (<https://github.com/tseemann/abricate>) (**Seemann, 2016**) and CRISPR-Cas systems with CRISPERCasFinder (**Couvin et al., 2018**). Species delineation relied on digital DNA-DNA hybridization (dDDH), computer with the Type (Strain) Genome Server (TYGS) (**Meier-Kolthoff and Göker, 2019**) and average nucleotide identity (ANIb) (**Goris et al., 2007**); the 70% dDDH and 95% ANI (**Richter et al., 2009**) thresholds were applied. The circular genome map was generated with Proksee (**Grant et al., 2023**).

For the 16S rRNA gene phylogeny, reference sequences of related type strains were retrieved from EzBioCloud (**Chalita et al., 2024**), aligned with MAFFT (**Katoh et al., 2019**), and a maximum-likelihood tree was inferred with PhyML using the model selected by SMS on NGPhylogeny.fr (**Lemoine et al., 2019**), with branch support estimated by bootstrap; the tree was visualized with ITOL (Interactive Tree of life) v7.6 (**Letunic and Bork, 2024**). Average nucleotide identity (ANIb) was computed with JSpeciesWS (**Richter et al., 2015**); applying the 95% species threshold (**Richter and Rossello, 2009**). Digital DNA-DNA hybridization (dDDH, formula d_4) and a Genome BLAST Distance Phylogeny (GBDP) tree were obtained with TYGS (**Meier-Kolthoff and Göker, 2019**), with a 70% threshold.

Ecological analysis IMNGS

To investigate the distribution, prevalence and ecological abundance of strain Marseille-QA1073^T in various human and animal habitats, raw metagenomic sequencing archives from IMNGS (<https://www.imngs.org>) were analyzed in September 2026 (**Lagkouvardos et al., 2016**).

Results and Discussion

Phenotypic and biochemical characteristics

Macroscopic examination revealed circular, convex colonies that were small and translucent during the initial hours of growth, reaching up to 1 cm in diameter with a slightly whitish tint after 48 to 72 hours. The colonies exhibited regular, smooth margins and a mucoid consistency. When fresh, cells are Gram-stain-positive, motile, endospore-forming, curved fusiform (spindle-to lemon-shaped) rods with a mean length of $7.3 \pm 0.8 \mu\text{m}$ (range 6.4–8.3 μm) (Fig. 1). SEM micrographs revealed clusters of cells exhibiting characteristic fusiform morphology with tapered ends and flagella-like appendages. Strain Marseille-QA1073 is anaerobic, growing optimally at 37°C (range 20–45°C); weak growth was also observed under microaerophilic conditions (28–45°C), whereas no growth occurred aerobically. The strain thrived in media adjusted to pH values ranging from 6.0 to 8.5; growth occurred on standard low-salt media but was inhibited at NaCl concentrations $\geq 7.5\%$. Both oxidase and catalase tests were negative. Results from the API ZYM gallery revealed activity for various substrates, including esterase (C4), leucine arylamidase, valine arylamidase, cystine arylamidase, α -chymotrypsin, acid phosphatase, naphthol-AS-BI-phosphohydrolase, and β -glucosidase. These data are summarized in Table 1. Additionally, the API 20 A galleries, designed for the biochemical characterization of anaerobic bacteria, showed positive results only for the hydrolysis of gelatin and esculin.

Table 1

Differential phenotypic and enzymatic characteristics between strain (1) Marseille-QA1073^T and its closest phylogenetic relative related species (2) *C. argentinense* ATCC 27322^T based on API ZYM tests standard biochemical assays

Properties/API ZYM	1	2
Cell shape	Curved, Fusiform (spindle-to lemon-shaped) rod	Straight to slightly curved rod
Cell length (µm)	6.3–8.3	1.6–9.4
Oxygen requirement	anaerobic	anaerobic
Gram stain	+	+
Flagellation	+	+
Motility	+	+
Endospore formation	+	+
Production of:		
Alkaline phosphatase	–	–
Acid phosphatase	+	–
Catalase	–	–
Oxidase	–	–
β-Glucosidase	+	–
α-Glucosidase	–	–
α-Galactosidase	–	–
β-Glucuronidase	–	–
β-Galactosidase	–	–
Esterase	+	–
Esterase lipase	–	–
Naphthol-AS-BI-phosphohydrolase	+	–
Lipase	–	–
Leucine arylamidase	+	–
N-acetyl-β-glucosaminidase	–	–
+ means positive reaction; – means negative reaction. Data taken from description of <i>C. argentinense</i> ATCC 27322 ^T (Suen et al., 1988)		

Properties/API ZYM	1	2
Valine arylamidase	+	–
Cystine arylamidase	+	–
Trypsine	–	–
α -chymotrypsine	+	–
α -fucosidase	–	–
α -mannosidase	–	–
Habitat	maternal milk	Soil
+ means positive reaction; – means negative reaction. Data taken from description of <i>C. argentinense</i> ATCC 27322 ^T (Suen et al., 1988)		

Fatty acid analysis

The results of the cellular fatty acid composition analysis for strain Marseille-QA1073[¶] demonstrated that the most abundant fatty acids were hexadecanoic acid (C_{16:0}, 35.8%) and 3-methylbutanoic acid (acid iso-C_{5:0}, 33.5%) followed by tetradecanoic acid (C_{14:0}, 9.3%) and 15-methylhexadecanoic acid (iso-C_{17:0}, 3.4%), together with minor branched and unsaturated fatty acids. A comparison of the fatty acid profiles of strain Marseille-QA1073[¶] with those of closely related species is provided in Table 2.

Table 2
Cellular fatty acid methyl ester (FAME) profiles (%) of strain Marseille-QA1073^T

Fatty acid	Name	(1) Mean relative (%) ^a
C _{16:0}	Hexadecanoic acid	35.8
C _{5:0 iso}	3-Methyl-butanoic acid	33.5
C _{14:0}	Tetradecanoic acid	9.3
C _{17:0 iso}	15-Methyl-Hexadecenoic acid	3.4
C _{15:0 iso}	13-Methyl-Hexadecenoic acid	3.0
C _{18:1n12}	6-Octadecenoic acid	2.8
C _{13:0 iso}	11- Methyl-dodecanoic acid	2.7
C _{18:0}	Octadecanoic acid	1.6
C _{15:0 anteiso}	12-Methyl-tetradecanoic acid	1.6
C _{16:1n7}	9-Hexadecenoic acid	1.3
C _{18:1n9}	9-Octadecenoic acid	1.2
C _{18:1n7}	11-Octadecenoic acid	1.1
C _{13:0 anteiso}	10-Methyl-dodecanoic acid	1.0
C _{12:0}	Dodecanoic acid	TR
C _{16:1n9}	7-Hexadecenoic acid	TR
C ₈ H ₈ O ₂	Phenylacetic acid	15.0*
* Mean peak area percentage; Values are percentages of total fatty acids. Only fatty acids representing ≥ 1% of the total are shown, while those amounting to < 1% are indicated by TR (Trace)		
* Value when aromatic acids were included in the total fatty acid profile		
All data taken from cellular fatty acid methyl ester (FAMES) analysis of this study.		

Antibiotic tests analysis

The results of the antibiotic susceptibility testing for strain Marseille-QA1073, performed in accordance with EUCAST guidelines (Gaur et al., 2023), are summarized in Table 3.

Table 3
Antibiotic susceptibility profile of strain Marseille-QA1073^T

Antibiotic	MIC (µg/mL)
Amikacin	> 256
Amoxicillin	0.064
Benzylpenicillin	0.25
Ceftazidime	> 256
Ceftriaxone	4
Ciprofloxacin	0.125
Clindamycin	2
Colistin	> 256
Doxycycline	< 0.016
Gentamycine	24
Imipenem	0.125
Nitrofurantoin	1.5
Oxacillin	1.0
Rifampicin	0.016
Teicoplanin	1.5
Tobramycin	48
Cotrimoxazole	0.19
Vancomycin	12
Minimum inhibitory concentrations (MIC, in µg/mL) were determined using Etest on Mueller Hinton (MH) agar plates incubated at 37°C under anaerobic conditions in accordance with EUCAST guidelines.	

Virulence and antimicrobial resistance determinants

No virulence factor, and in particular, no botulinum neurotoxin (*bont*) gene was detected (ABRicate/VFDB) in the genome of strain Marseille-QA1073; despite the strain’s affiliation to the *C. botulinum* group (its closest relative, *C. argentinense*, corresponds to the former *C. botulinum* toxinotype G). The genome carries a chromosomal *vanF*-type glycopeptide-resistance operon

(*vanF/vanH/vanX/vanS/vanR/vanY*; identified by ARG-ANNOT and CARD and confirmed by the KEGG module M00651, the latter also independently recovered by Protologger) and a *cfrC* gene (23S rRNA methyltransferase associated with the PhLOPSA phenotype). No plasmid replicon was detected (PlasmidFinder), indicating that these determinants are chromosomally encoded. Phenotypically, the strain showed an elevated vancomycin MIC (12 mg/L) while remaining susceptible to teicoplanin (1.5 mg/L), a dissociated glycopeptide phenotype typical of *vanB/vanF*-type resistance, in which only vancomycin induces the operon. The *cfrC* determinant was reflected only by a borderline linezolid MIC (4 mg/L), clindamycin remaining low (2 mg/L), suggesting limited expression. Resistance to aminoglycosides (amikacin, gentamicin, tobramycin) is consistent with the intrinsic resistance of anaerobes rather than an acquired determinant.

Identification of strains

The protein spectra of strain Marseille-QA1073 could not be identified by MALDI-TOF MS (score < 1.7), despite comparison with a comprehensive database. Whole-genome sequencing was therefore performed, and the 16S rRNA gene sequence was extracted to determine the taxonomic position of the strain. BLASTn comparison showed that strain Marseille-QA1073 was most closely related to *Clostridium faecium* N37 (99.1%) and *Clostridium argentinense* ATCC 27322 (98.9%) (Fig. 2); as these similarities exceed the 98.7% threshold, the 16S rRNA gene did not discriminate the strain from its closest relatives. Its status was resolved by whole-genome comparison (below); at this stage, the strain could only be assigned to the genus *Clostridium* (family *Clostridiaceae*, phylum *Bacillota*). Genome accession JBSVQM000000000.

Genomic comparison

The dereplicated genome of strain Marseille-QA1073 spans 4,450,951 bp with a G + C content of 28.55 mol%. Following a genome dereplication step to address an initial contamination level, exceeding 5%, the final assembly achieved a high completeness of 98.87% and a low contamination rate of 2.38% (CheckM2). Compared to its closest phylogenetic relatives within the *Clostridium sensu stricto* group (Table S1), its genome metrics fall well within the expected genomic boundaries for this genus, confirming the high quality of the draft genome assembly. A total of 4,142 genes were predicted, comprising 4,041 protein-coding sequences and 101 RNA genes (12 rRNA and 89 tRNA) (Tables 4 and 5).

Table 4

Nucleotide content and genome statistical features of strain Marseille-QA1073^T

Attribute	Value
Genome Size (bp)	4,450,951
G + C content (%)	28.55
Contig	58
N50 (bp)	142,074
Total genes	4,142
rRNA /tRNA genes	12/89
Protein-coding genes	4,041
Completeness / contamination	98.87% / 2.38%
Accession number	JBSVQM000000000
Genome metrics including size, G + C content, and gene counts were determined from draft genome sequence. Completeness and contamination values were estimates using CheckM2.	

Table 5

Summary of 16S rRNA gene similarity, ANIb and dDDH values between strain Marseille-QA1073^T and its closest type strains

Comparison	16S rRNA (%)	dDDH d_4 (%) (GGDC)	ANI (%)
<i>C. argentinense</i> ATCC 27322	98.90	55.30	94.30
<i>C. faecium</i> N37	99.10	34.10	87.40
<i>C. senegalense</i> JC122	95.30	22.50	74.00
<i>C. combesii</i> DSM 20696	93.10	27.30	72.10
<i>C. botulinum</i> ATCC 25763	93.20	27.40	72.00
<i>C. subterminale</i> JCM 1417	95.20	22.20	71.50
<i>C. lundense</i> DSM 17049	94.00	21.70	70.10
<i>C. pasculi</i> DSM 10365	95.00	22.40	69.50
16S rRNA gene similarity was computed from MAFFT alignment (full-length reference sequences) ; ANIb average nucleotide identity (BLAST-based, JSpecies ; Goris <i>et al.</i> , 2007) ; dDDH, digital DNA-DNA hybridization (TYGS/GBDP formula d_4). Species-delineation thresholds : 98.65% (16S), 95% (ANIb) and 70% (dDDH). Note that the 16S rRNA gene does not discriminate the strain from <i>C. argentinense</i> and that from <i>C. faecium</i> (~ 99%), whereas ANIb and dDDH clearly do. GGDC mean Genome-to-Genome Distance Calculator. Formula d_4 (formula recommended for comparison).			

Strain Marseille-QA1073 shared an average nucleotide identity (ANIb) of 94.3% with *C. argentinense* and 87.4% with *C. faecium*, both well below the 95% species-delineation threshold (Fig. 3). Consistently, digital DNA-DNA hybridization (dDDH) values were 55.3% with *C. argentinense* and 34.1% with *C. faecium*, both well below the 70% threshold (Fig. 4).

Functional annotation of the predicted genes of strain Marseille-QA1073 using the eggNOG-mapper pipeline revealed a comprehensive distribution across the Clusters of Orthologous Groups (COG) functional categories. The most heavily represented functional categories included are transcription (**K**, 10.6%), energy production and conversion (**C**, 7.6%), and amino acid transport and metabolism (**E**, 7.5%) highlighting a dominance of core metabolic and regulatory functions. Furthermore, a substantial proportion of coding sequences (CDS) are dedicated to translation, ribosomal structure, and biogenesis (**J**, 7.0%), inorganic ion transport and metabolism (**P**, 6.0%), and replication, recombination and repair (**L**, 5.9%), underscoring stable and active translational machinery. (Fig. 5; Table 6). Finally, the notable presence of genes involved in signal transduction mechanism (**T**, 5.4%) suggests a strong capacity for active environmental adaptation. Overall, this functional reflects a robust and balanced metabolic potential, characteristic of bacteria equipped with active cellular machinery and a well-developed transcriptional regulation network.

Table 6
Distribution of functional classes of predicted gene according to the Database of Clusters Orthologous Groups (COG) for strain Marseille-QA1073^T

COG Category	Main biological function	Marseille-QA1073 ^T (%)
K	Transcription	10.60
C	Energy production and conversion	7.60
E	Amino acid metabolism and transport	7.50
J	Translation, ribosomal structure	7.00
P	Inorganic transport and metabolism	6.00
L	Replication, recombination, and repair	5.90
T	Signal transduction mechanisms	5.40
M	Cell wall/ membrane, envelop	4.90
V	Defense mechanisms	4.20
H	Coenzyme, transport, and metabolism	3.60
G	Carbohydrates transport and metabolism	3.50
F	Nucleotide transport and metabolism	3.30
Unclassified	Not assigned to any COG category	36.40
Total		100.00

Circular chromosome maps generated for the strain Marseille-QA1073^T provide an overview of its genomic features. This map depicts the structural organization, including the positioning of protein-coding sequences on strand, RNA genes, as well as variations in G + C content and G + C skew profiles (**Figure S1**).

Together, these genomic-based indices demonstrate that strain Marseille-QA1073^T represents a novel species within the genus *Clostridium*. On this basis, strain Marseille-QA1073^T is proposed as a novel species within the genus *Clostridium*, family *Clostridiaceae*, and phylum *Bacillota*. The genomic sequence has been deposited in GenBank under accession number JBSVQM000000000.

Predicted ecology by Protologger

In silico prediction of the ecological distribution of strain Marseille-QA1073^T showed a predominantly edaphic profile: the strain detected most frequently in rhizosphere (38.0%) and soil (17.5%) metagenomes, followed by aquatic biomes (marine sediment, 9.3%, freshwater 8.1%), whereas human-associated biomes were marginal (< 1%); no metagenome-assembled genome matched the strain (**Table S2**).

Discussion

Strain Marseille-QA1073 was isolated from human breast milk and characterised using a taxonogenomic approach. The 16S rRNA gene did not discriminate the strain from its closest relatives, sharing approximately 99% similarity with both *C. argentinense* and *C. faecium*, above the 98.7% threshold generally used for species delineation. This limited resolution of the 16S rRNA gene within *Clostridium* sensu stricto is well documented and illustrates why single-marker identification is insufficient at this level. Species delineation therefore relied on whole-genome indices : the average nucleotide identity (94.3% with *C. argentinense*, its closest relative) and the digital DNA-DNA hybridization value (55.3%) both fell well below the 95% and 70% thresholds, unambiguously defining strain Marseille-QA1073 as a novel species, for which the name *Clostridium niakharensense* sp. nov. is proposed.

Although the closest relative of strain Marseille-QA1073, *C. argentinense*, corresponds to the former *C. botulinum* toxinotype G, a lineage associated with botulinum neurotoxin production, no botulinum neurotoxin (bont) gene, and more generally no virulence factor, was detected in the genome (VFDB). This absence is reassuring regarding the safety of the strain and indicates that its relatedness to the *C. botulinum* group is phylogenetic rather than functional.

In this respect, culture-based recovery of *Clostridium* sensu stricto from human milk is exceptional. Most reports of “*Clostridium*” in milk rely on molecular methods and correspond to cluster IV/XIVa taxa now reassigned to other, often commensal genera (e.g. *Faecalibacterium*, *Roseburia*, *Blautia*), rather than to the genus *Clostridium* sensu stricto (Collabo et al., 2009; Jost et al., 2013; Togo et al., 2019). Genuine, cultured isolates remain largely restricted to *C. perfringens*, *C. tertium* and *C. sporogenes*, mostly in milk-bank or contamination settings (Liu et al., 2023), while aerotolerant Firmicutes dominate the cultivable fraction. Some clostridia, notably *C. perfringens* and *C. butyricum*, have further been associated with necrotizing enterocolitis in preterm infants, mostly as intestinal rather than milk-borne isolates (Cassir et al., 2015). Notably, prolonged anaerobic storage of milk has been shown to reveal the viability of low-abundance, slow-growing Clostridiales otherwise missed by standar culture (Schwab et al., 2019), consistent with the prolonged anaerobic enrichment required here to recover strain Marseille-QA1073. This member of the *C. argentinense* lineage, devoid of any virulence or toxin gene, thus expands the range of cultivable clostridia recovered from human milk; whether it contributes beneficially, as often observed for milk-associated clostridial taxa, remains to be established by further characterization.

The genome carried a chromosomal *vanF*-type glycopeptide-resistance operon (independently supported by the KEGG M00651 module recovery by Protologger) and a *cfrC* gene, without any plasmid replicon, indicating chromosomally encoded determinants. The *vanF* operon was phenotypically expressed: the strain showed an elevated vancomycin MIC (12 mg/L) contrasting with susceptibility to teicoplanin (1.5 mg/L), a dissociated glycopeptide phenotype characteristic of *vanB/vanF*-type resistance in which only vancomycin induces the operon. The *cfrC* determinant translated only into a borderline linezolid MIC (4 mg/L), clindamycin remaining low, suggesting limited expression. Resistance

to aminoglycosides was consistent with the intrinsic resistance of anaerobic bacteria. The presence of a functional, chromosomally located glycopeptide-resistance operon in an organism of human origin is noteworthy, although its clinical relevance remains to be established.

The predominantly edaphic distribution predicted for strain Marseille-QA1073 rather than contradicting its recovery from breast milk, would be consistent with the environmental lifestyle of most *Clostridium* sensu stricto species, which are widespread, spore-forming inhabitants of soil and sediments; its isolation from milk might therefore reflect environmental exposure of the host rather a milk-specific adaptation. This tentative interpretation is further supported by the presence of a complete flagellar gene set (*FlhA/B*, *Flg*, *Fli*, *MotA/B*), consistent with the motility observed microscopically. As these predictions rest on a partial 16S rRNA gene sequence, they remain indicative only and would require dedicated functional and ecological studies to confirm whether this taxon is a resident or a transient member of the milk microbiota.

The recovery of this fastidious, slow-growing anaerobe required a prolonged anaerobic enrichment within a culturomics framework, underlining the value of culturomics for accessing the cultivable diversity of the breast-milk microbiota. The description of *Clostridium niakharensense* sp. nov. adds to the growing inventory of novel taxa isolated from human milk and contributes to a better characterization of this under-explored microbial niche.

Conclusion

The comprehensive study of the phenotypic, genomic, and phylogenetic characteristics of strain Marseille-QA1073 has enabled its precise identification and taxonomic placement, clearly distinguishing it from all previously described species within the genus *Clostridium*. Genomic analysis demonstrated that while the 16S rRNA gene sequence shows similarity to closely related species, the dDDH and ANI values relative to its nearest phylogenetic neighbors remain significantly below the accepted species delimitation thresholds (ANIb 94.3%, dDDH 55.3% with *C. argentinense*). The strain carries a chromosomal vanF-type operon but no virulence factor or botulinum neurotoxin gene. Its isolation from human breast milk by culturomics illustrates the value of this approach for under-characterised niches. These findings support the proposal of a novel species within the genus *Clostridium*, for which the name *Clostridium niakharensense* sp. nov. is proposed. The type strain is Marseille-QA1073 (= CSUR QA1073 = CECT 31321).

Declarations

Description of *Clostridium niakharensense* sp. nov.

Clostridium niakharensense (nia.kha.ren'se. N.L. neut. adj. *niakharensense*, pertaining to Niakhar, Fatick region, Senegal, where the breast milk samples were collected).

Cells are Gram-stain-positive, endospore-forming, curved fusiform (sprindle-to lemon-shaped) rods with a mean length of $7.3 \pm 0.8 \mu\text{m}$ (range 6.4-8.3 μm), motile, occurring singly or in small clusters. Anaerobic; weak growth under microaerophilic conditions, none aerobically. Optimal growth occurs at 37 °C (range 20-45 °C), at pH 6.0-8.5; grows on standard low salt media, inhibited at $\text{NaCl} \geq 7.5 \%$ and catalase-and oxidase-negative. In API ZYM, activity for esterase (C4), leucine arylamidase, valine arylamidase, cystine arylamidase, α -chymotrypsin, acid phosphatase, naphthol-AS-BI-phosphohydrolase, and β -glucosidase; in API 20 A, positive for gelatin and esculin hydrolysis. Major fatty acid is: (C_{16:0}, 35.8 %) and (acid iso-C_{5:0}, 33.5 %). Carries a chromosomal *vanF*-type operon (vancomycin MIC 12 mg/L, teicoplanin 1.5 mg/L); no virulence factor or (*bont*) gene detected. Genome 4,450,951 bp, 28.55 % G+C.

The type strain, Marseille-QA1073 (= CSUR QA1073 = CECT 31321), was isolated from human breast milk of a mother breastfeeding a healthy child in Niakhar, Fatick region, in Senegal. The complete genome has been deposited in GenBank under accession number JBSVQM000000000.

Acknowledgments:

The authors would like to thank Claudia Andreu for sequencing and Rayan Semmar for depositing complete genome in GenBank.

FUNDING

This work was supported by a grant from the French Government managed by the National Research Agency under the “Investissements d’avenir (Investments for the Future)” programme with the reference ANR-10-IAHU-03 (Méditerranée Infection), by the Contrat Plan Etat-Région and the European funding FEDER IHUPERF.

Conflicts of Interest:

Sara Bellali received personal fees through a collaborative contract with Hitachi High-Tech Corporation. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Authors contributions:

All authors contributed to the conception, design, and execution of the study.

M.M., M.C., C.S. were responsible for the Conceptualisation of this study. Data curation was handled by **S.S., O.N., C.S., H.B** and **G.D._ S.S.**, performed bacterial isolation, and phenotypic characterization and **A.I.A.**, assisted with the biochemical analyses_ **S.S., A.M.**, and **M.B.**, conducted the formal analysis including phylogenetic, reconstruction, comparative genomics and statistical evaluations_ **S.B.**, performed the micrograph imaging_ **S.A.**, handled the strain deposition in the second culture collection_

N.O., performed the cellular fatty acid analyses_ **S.M.**, acquired the MALDI-TOF spectra and registered the strain in the database_ **M.M.**, secured funding for the study_ **O.N.**, performed the primary investigation including sample collection_ **M.M.**, **M.T.A.**, and **M.C.** designed the methodology_ **M.M.**, **C.S.**, and **M.C.** managed the project administration and oversaw the supervision _ **M.M.**, **C.S.**, **G.D.**, **H.B.** and **C.T.** provided the ressources_ **M.M.**, **M.C.**, **C.S.**, **M.T.A.** performed the validation of the results. **S.S.**, and **A.M.** writing-original draft_ **M.M.** and **S.S.** prepared the visualization of the results, and all authors contributed to the writing -review and editing.

Data sharing statement

The strain has been deposited in the bioBank of the Méditerranée Infection (IHU) within the CSUR (Collection des Souches de l'Unité des Rickettsies; <https://csur.eu/>) under the accession number QA1073. It was also deposited in the Colección Española de Cultivos Tipo (CECT <https://www.uv.es/cect>) under the identifier CECT 31321. The complete genome of the strain was deposited in GenBank under the assembly accession number **JBSVQM000000000**.

Ethical approval

Samples were collected as part of a study on breast milk microbiota and malnutrition, following approval by the the National Ethics and Committee for Health Research (CNERS) of the Senegalese Ministry of Health and Public Hygiene (MSHP; reference no. 00000233/MSAS/CNERS/SP on 11 September 2022).

References

1. Andrews S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data. Babraham *Bioinformatics*. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
2. Bolger AM, Lohse M, Usadel B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–20. <https://doi.org/10.1093/bioinformatics/btu170>.
3. Brüggemann H, Baumer S, Fricke, WF, Wiezer A, Liesegang H, Decker I, Gottschalk G. (2003). The genome sequence of *clostridium tetani*, the causative agent of tetanus disease. *Proc Natl Acad Sci USA* 100:1316–21. <http://doi.org/10.1073/pnas.0335853100>.
4. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. (2021). EggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol Biol Evol*. 9;38(12):5825-5829. <http://doi.org/10.1093/molbev/msab293>.
5. Cassir N, Benamar S, Khalil JB, Croce O, Saint-Faust M, Jacquot A, Million M, Raoult D, La Scola B. (1025). *Clostridium butyricum* Strains and Dysbiosis Linked to Necrotizing Enterocolitis in Preterm Neonates. *Clin Infect Dis*. 1;61(7):1107-15. <https://doi.org/10.1093/cid/civ468>.
6. Cerf O, Bergère J. (1968). La numération des spores de *Clostridium* et son application au lait et aux produits laitiers. 1. Méthodes et techniques générales de numération. *Lait* 48:501–19. <https://doi.org/10.1051/lait:1968475-47615>.

7. Chalita, M., Kim, Y. O., Park, S., Oh, H. S., Cho, J. H., Moon, J., Baek, N. et al., (2024). EzBioCloud: a genome-driven database and platform for microbiome identification and discovery. *Int J Syst Evol Microbiol* 74(6):006421. <https://doi.org/10.1099/ijsem.0.006421>.
8. Chklovski, A., Parks, D. H., Woodcroft, B. J., & Tyson, G. W. (2023). CheckM2: a fast, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nature Methods*, 20(8), 1203–1212. <https://doi.org/10.1038/s41592-023-01940-w>
9. Collado MC, Delgado S, Maldonado A, Rodríguez JM. (2009). Assessment of the bacterial diversity of breast milk of healthy women by quantitative real-time PCR. *Lett Appl Microbiol*. 48(5):523-8. <https://doi.org/10.1111/j.1472-765X.2009.02567.x>.
10. Collins MD, Lawson PA, Willems A, Cordora JJ, Fernandez-Garayzabal J, Garcia P, et al., (1994). The phylogeny of the genus *Clostridium*: proposal of five new genera and eleven new species combinations. *Int J Syst Bacteriol* 44(4):812–26. <https://doi.org/10.1099/00207713-44-4-812>.
11. Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Néron B, et al., (2018). CRISPRCasFinder, an update of CRISPRfinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res*. 46: W246–W251. <https://doi.org/10.1093/nar/gky425>.
12. Cruz-Morales Po, Orellana CA, Moutafis G, Moonen G, Rincon G, Nielsen LK, Marcellin E. (2019). Revisiting the Evolution and Taxonomy of Clostridia, a Phylogenomic Update. *Genome Biol and Evol*. 11(7):2035–44. <https://doi.org/10.1093/gbe/evz096>
13. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. (2018). NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34(15):2666-2669. <https://doi.org/10.1093/bioinformatics/bty149>.
14. Dione N, Sankar SA, Lagier JC, Khelaifia S, Michele C, Armstrong N, Richez M, Abrahão J, Raoult D, Fournier PE. (2016). Genome sequence and description of *Anaerosalibacter massiliensis* sp. nov. *New Microbes New Infect* 10:66–76. <https://doi.org/10.1016/j.nmni.2016.01.002>.
15. Galperin MY, Makarova KS, Wolf YI, Koonin EV. (2015). Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Research*; 43(D1): D261–D269. <https://doi.org/10.1093/nar/gku1223>.
16. Gaur P, Hada V, Rath RS, Mohanty A, Singh P, Rukadikar A. (2023). Interpretation of Antimicrobial Susceptibility Testing Using European Committee on Antimicrobial Susceptibility Testing (EUCAST) and Clinical and Laboratory Standards Institute (CLSI) Breakpoints: Analysis of Agreement. *Cureus* 15(3):e36977. <https://doi.org/10.7759/cureus.36977>.
17. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P and Tiedje JM (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol*. 57(Pt 1):81-91 <http://ijs.sgmjournals.org/content/57/1/81.long>.
18. Gilroy R, Ravi A, Getino M, Pursley I, Horton DL, Alikhan NF, Baker D, Gharbi K, Hall N, Watson M. et al., (2021). Extensive microbial diversity within the chicken gut microbiome revealed by metagenomics and culture. *PeerJ* 9:e10941. <https://doi.org/10.7717/peerj.10941>.

19. Göker M, Christensen H, Fingerle V, Kostovski M, Margos G, Moore ERB, Oren A, Sheila P, Reischl U, Vázquez-Boland JA. (2025). List of Recommended Names for bacteria of medical importance: report of the Ad Hoc Committee on Mitigating Changes in Prokaryotic Nomenclature. *Int J Syst Evol Microbiol* 75:006943. <https://doi.org/10.1099/ijsem.0.006943>.
20. Goodsir J, Wilson G. (1842). History of a Case in Which a Fluid Periodically Ejected from the Stomach Contained Vegetable Organisms of an Undescribed Form. *Edinb Med Surg J.* 57(151):430–43.
21. Grant JR, Enns E, Marinier E, Mandal A, Herman EK, Chen C, Graham M, Van Domselaar G, Stothard P. (2023). Proksee: in-depth characterization and visualization of bacterial genomes *Nucleic Acids Res*, 51: W484-W492 <https://doi.org/10.1093/nar/gkad326>
22. Grosse-Herrenthey A, Maier T, Gessler F, Schaumann R, Böhnelt H, Kostrzewa M, Krüger M. (2008). Challenging the problem of clostridial identification with matrix-assisted laser desorption and ionization-time-of-flight mass spectrometry (MALDI-TOF MS). *Anaerobe* 14(4):242–9. <https://doi.org/10.1016/j.anaerobe.06.002>
23. Gupta RS, Gao B. (2009). Phylogenomic analyses of clostridia and identification of new protein signatures specific to the genus *Clostridium* sensu stricto (grappe I). *Int J Syst Evol Microbiol* 59:285–94. <https://doi.org/10.1099/ijss.0.001792-0>.
24. Hauschild A. (2018). *Clostridium Botulinum*: Ecology and Control in Foods. CRC Press. <https://doi.org/10.1201/9781315139623>.
25. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P. (2017). Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Mol Biol Evol.* 1;34(8):2115-2122. <https://doi.org/10.1093/molbev/msx148>.
26. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I. et al., (2019). EggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 8;47(D1): D309-D314. <https://doi.org/10.1093/nar/gky1085>. 30418610.
27. Johnson JL, Francis BS. (1975). Taxonomy of the Clostridia: ribosomal ribonucleic acid homologies among the species. *J Gen Microbiol* 88(2):229–44. <https://doi.org/10.1099/00221287-88-2-229>.
28. Jost T, Lacroix C, Braegger C, Chassard C. (2013). Assessment of bacterial diversity in breast milk using culture-dependent and culture-independent approaches. *Br J Nutr.* 110(7):1253-62. <https://doi.org/10.1017/S0007114513000597>.
29. Katoh, Rozewicki, Yamada (2019) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform* 20:1160-1166. <https://doi.org/10.1093/bib/bbxA108>
30. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. (2001). Predicting transmembrane protein topology with a hidden Markov mode I: application to complete genomes. *J Mol Biol.* 305(3):567–80. <https://doi.org/10.1006/jmbi.2000.4315>.

31. Lagier JC, Dubourg G, Million M, Cadoret F, Bilen M, Fenollar F, et al. (2018). Culturing human microbiota and culturomics. *Nat Rev Microbiol*. 16:540-550. <https://doi.org/10.1038/s41579-018-0041-0>.
32. Lagkouvardos I, Joseph D, Kapfhammer M, Giritli S, Horn M, Haller D, Clavel T. (2016). IMNGS: a comprehensive open resource of processed 16S rRNA microbial profiles for ecology and diversity studies. *Sci Rep* 6:33721.
33. Lawson PA (2016). The taxonomy of the genus *Clostridium*: current status and future perspectives. *Microbiol China* 43:1070-1074
34. Lawson PA, Citron DM, Tyrrell KL, Finegold SM. (2016). Reclassification of *Clostridium difficile* as *Clostridioides difficile* (Hall and O'Toole 1935) Prévot 1938. *Anaerobe* 40:95–9. <https://doi.org/10.1016/j.anaerobe.2016.06.008>.
35. Lawson PA, Rainey FA. (2016). Proposal to restrict the genus *Clostridium* Prazmowski to *Clostridium butyricum* and related species. *Int J Syst Evol Microbiol* 66(2):1009–16. <https://doi.org/10.1099/ijsem.0.000824>.
36. Lemoine F, Correia D, Lefort V, Doppelt-Azeroual O, Mareuil F, Cohen-Boulakia S, Gascuel O (2019). NGPhylogeny.fr: new generation phylogenetic services for non-specialists. *Nucleic Acids Res* 2019.
37. Letunic I, Bork P. (2024). Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res*. Jul 5;52(W1): W78-W82. <https://doi.org/10.1093/nar/gkae268>.
38. Liu K, Liu Z, Chen X, Wang R, Wang D. (2023). Detection of *Clostridium perfringens* in donor milk at a human breast milk bank: a case report. *BMC Infect Dis*. 17;23(1):810. <https://doi.org/10.1186/s12879-023-08822-8>.
39. Meier-Kolthoff JP, Göker M. (2019). TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy *Nat Commun* 10:2182. <https://doi.org/10.1038/s41467-019-10210-3>.
40. Prazmowski A. (1880). Untersuchung über die Entwicklungsgeschichte und Fermentwirkung einiger Bakterien-Arten Inaugural Dissertation. Leipzig (Germany).
41. Rainey F, Hollen GJ, Small S. (2009). Genus I. *Clostridium*. In: Bergey's Manual of systematic Bacteriology. 2nd ed. Vol. 3. New York: Springer; p. 738-828.
42. Richter M, and Rosselló-Móra R (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci USA* 106(45):19126-31 <http://www.pnas.org/content/early/2009/10/22/0906412106.abstract>
43. Richter M, Rosselló-Móra R, Glöckner FO, Peplies J. (2015). JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* 32(6):929–931
44. Rossi-Tamisier M, Benamar S, Raoult D, Fournier PE. (2015). Cautionary tale of using 16S rRNA gene sequence similarity values in identification of human-associated bacterial species. *Int J Syst Evol Microbiol* 65(Pt 6):1929–34. <https://doi.org/10.1099/ijss.0.000161>.

45. Sasser M. (2006). Bacterial Identification by Gas Chromatography Analysis of Fatty Acids Methyl Esther (GC-FAME). MIDI, Technicaml Note #101
46. Schwab C, Voney E, Ramirez Garcia A, Vischer M, Lacroix C. (2019). Characterization of the Cultivable Microbiota in Fresh and Stored Mature Human Breast Milk. *Front Microbiol.* 20;10:2666. <https://doi.org/10.3389/fmicb.2019.02666>.
47. Seemann T. (2014). Prokka: rapid annotation of the prokaryotic genome. *Bioinformatics* 30(14):2068–9. <https://doi.org/10.1093/bioinformatics/BTU153>.
48. Seemann, T. (2016). ABRicate: Mass screening of contigs for antibiotic resistance genes. <https://github.com/tseemann/abricate>
49. Skerman VBD, McGowan V, Sneath PHA. (1980). Approved lists of bacterial names. *Int J Syst Bacteriol* 30(1):225–420. <https://doi.org/10.1099/00207713-30-1-225>
50. Suen JC, Hatheway CL, Steigerwalt AG, Brenner DJ. (1988). *Clostridium argentinense* sp. nov.: A Genetically Homogeneous Group Composed of All Strains of *Clostridium botulinum* Toxin Type G and Some Nontoxigenic Strains Previously Identified as *Clostridium subterminale* or *Clostridium hastiforme*. *Int J Syst Bacteriol* 38(4):375–81. <https://doi.org/10.1099/00207713-38-4-375>.
51. Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, Winther O, Brunak S. et al., (2022). SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat Biotechnol* 40(7):1023–5. <https://doi.org/10.1038/s41587-021-01156-3>.
52. Togo A, Dufour JC, Lagier JC, Dubourg G, Raoult D, Million M. (2019). Repertoire of human breast and milk microbiota: a systematic review. *Future Microbiol.* 14:623-641. <https://doi.org/10.2217/fmb-2018-0317>.
53. Wick RR, Judd LM, Holt KE. (2017). Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLOS Computational Biology*;13(6):e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>
54. Wiegel J, Tanner R, Rainey FA. (2006). An Introduction to the Family Clostridiaceae. In: Dworkin M, Falkow S, Rosenberg E, Schleifer KH, Stackebrandt E (eds) *The Prokaryotes*. Springer, New York, NY pp 654–78. https://doi.org/10.1007/0-387-30744-3_20.
55. Yutin N, Galperin MY. A (2013). genomic update on clostridial phylogeny: Gram-negative spore formers and other misplaced clostridia. *Environ Microbiol* 15(10):2631–41. <https://doi.org/10.1111/1462-2920.12173>.

Figures

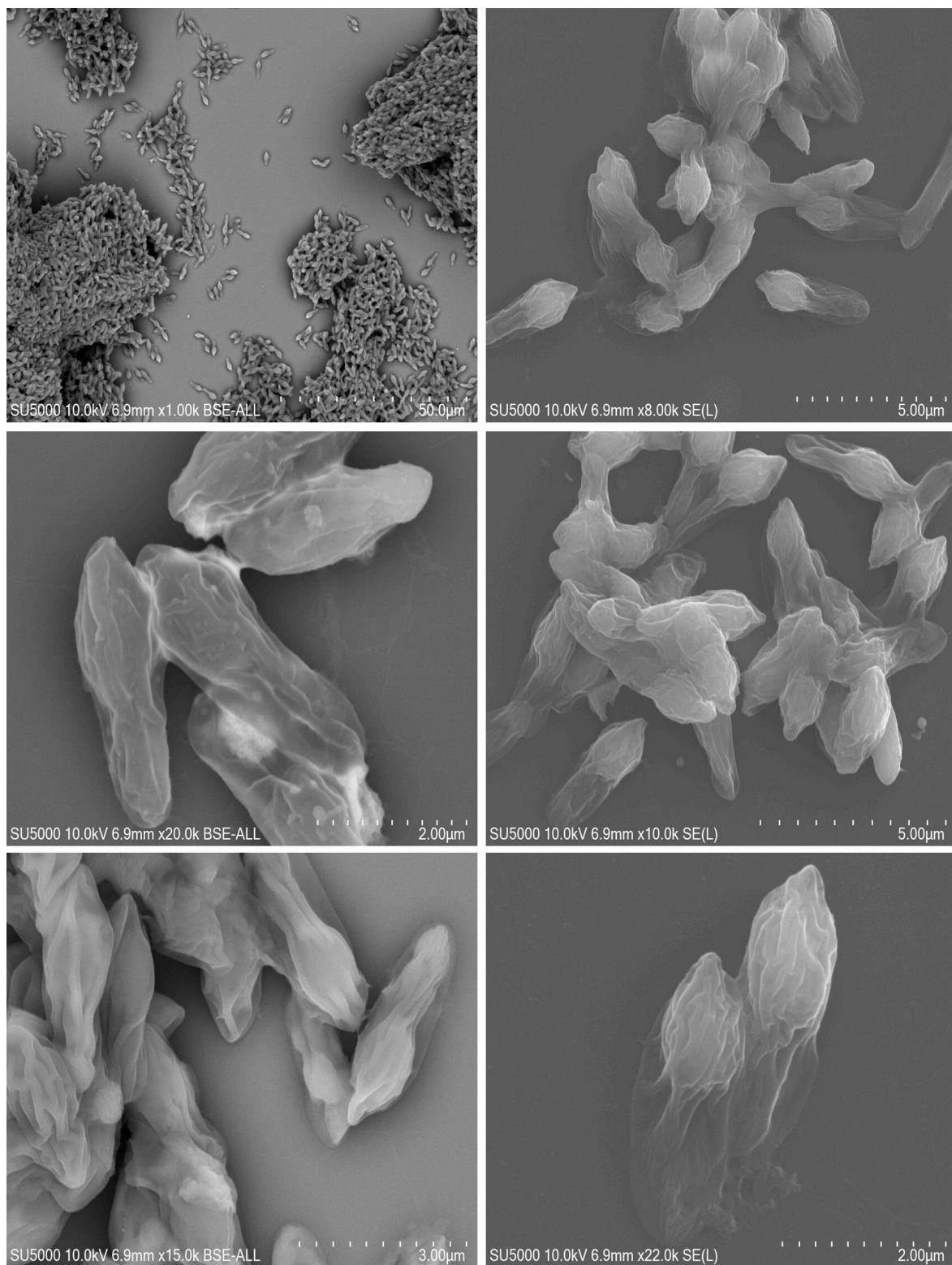


Figure 1

Scanning electron micrographs of flagellated fusiform cells of strain Marseille-QA1073^T obtained using a Hitachi SU5000 SEM. Parametric details are shown in the figures.

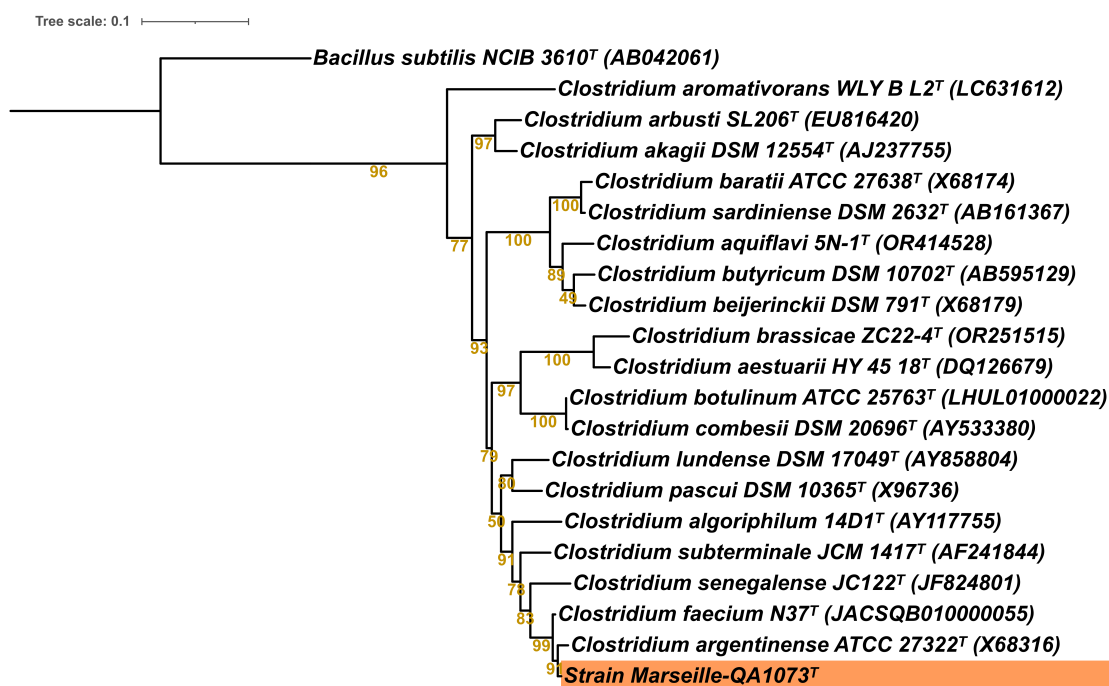


Figure 2

Phylogenetic tree based on the 16S rRNA gene sequences highlighting the position of strain Marseille-QA1073T among the most closely related validly published type strains. Accession numbers for all types of strains are in brackets.

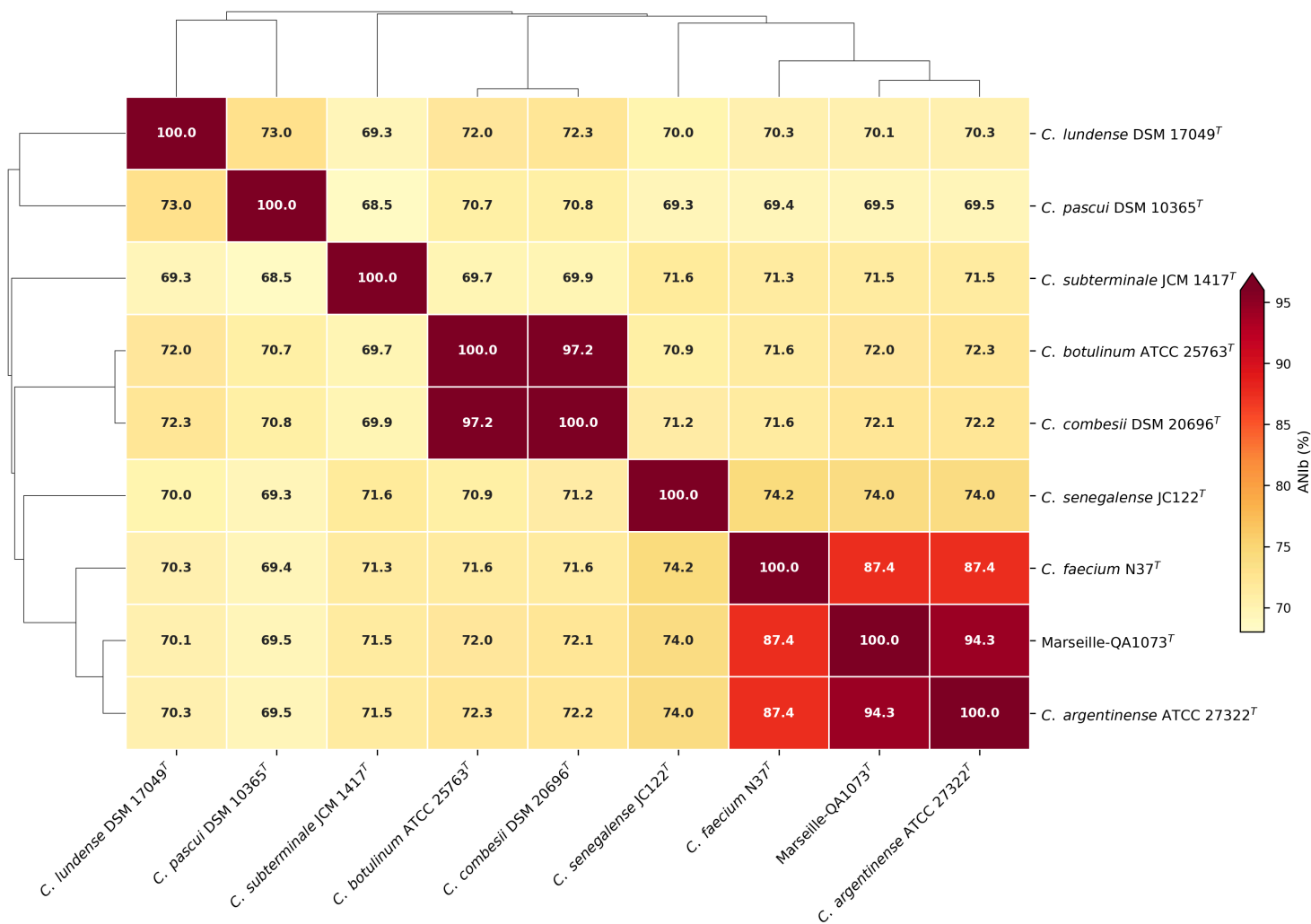


Figure 3

Average nucleotide identity (ANiB)_pourcentage_identity of strain Marseille-QA1073^T with closely strains.

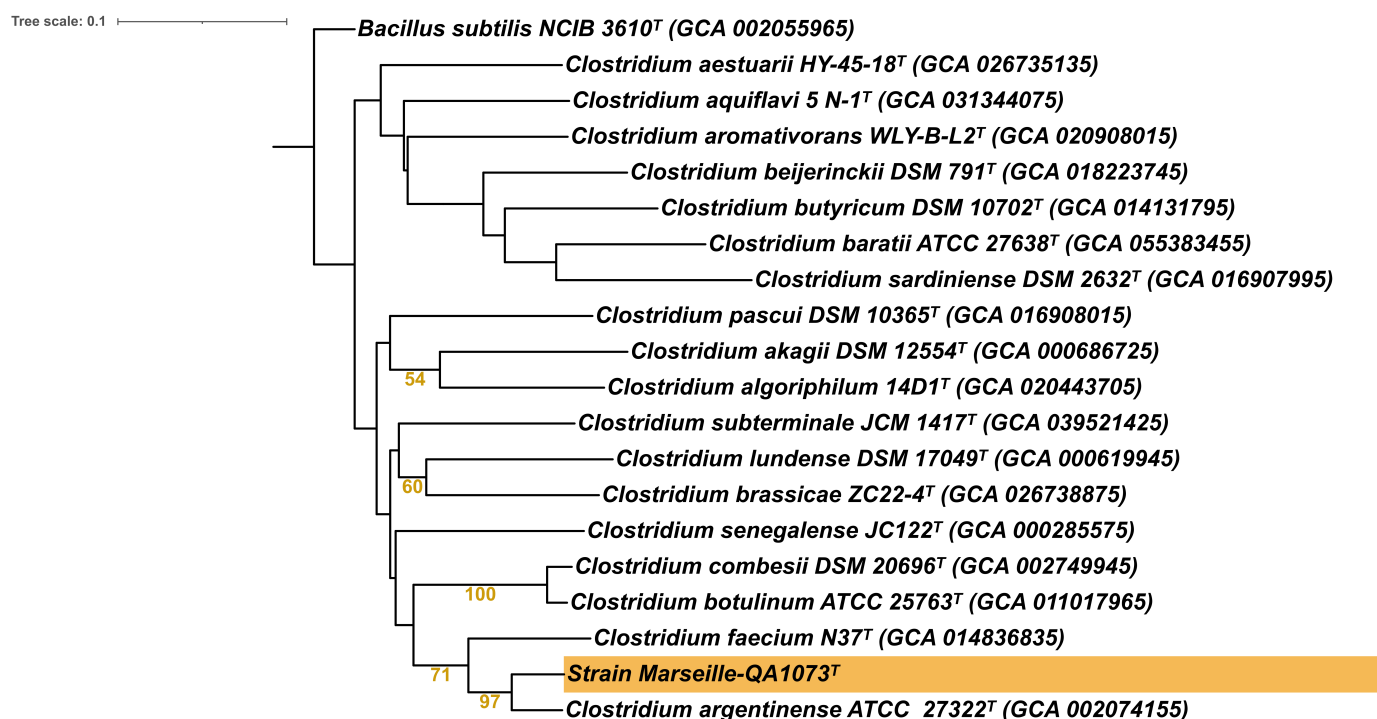


Figure 4

Maximum-likelihood phylogenomic tree based on genome sequences, showing strain Marseille-QA1073^T with its closely related species. (GenBank accession numbers are indicated in brackets).

Functional distribution of coding sequences among COG categories in
Clostridium niakharens Marseille-QA1073^T

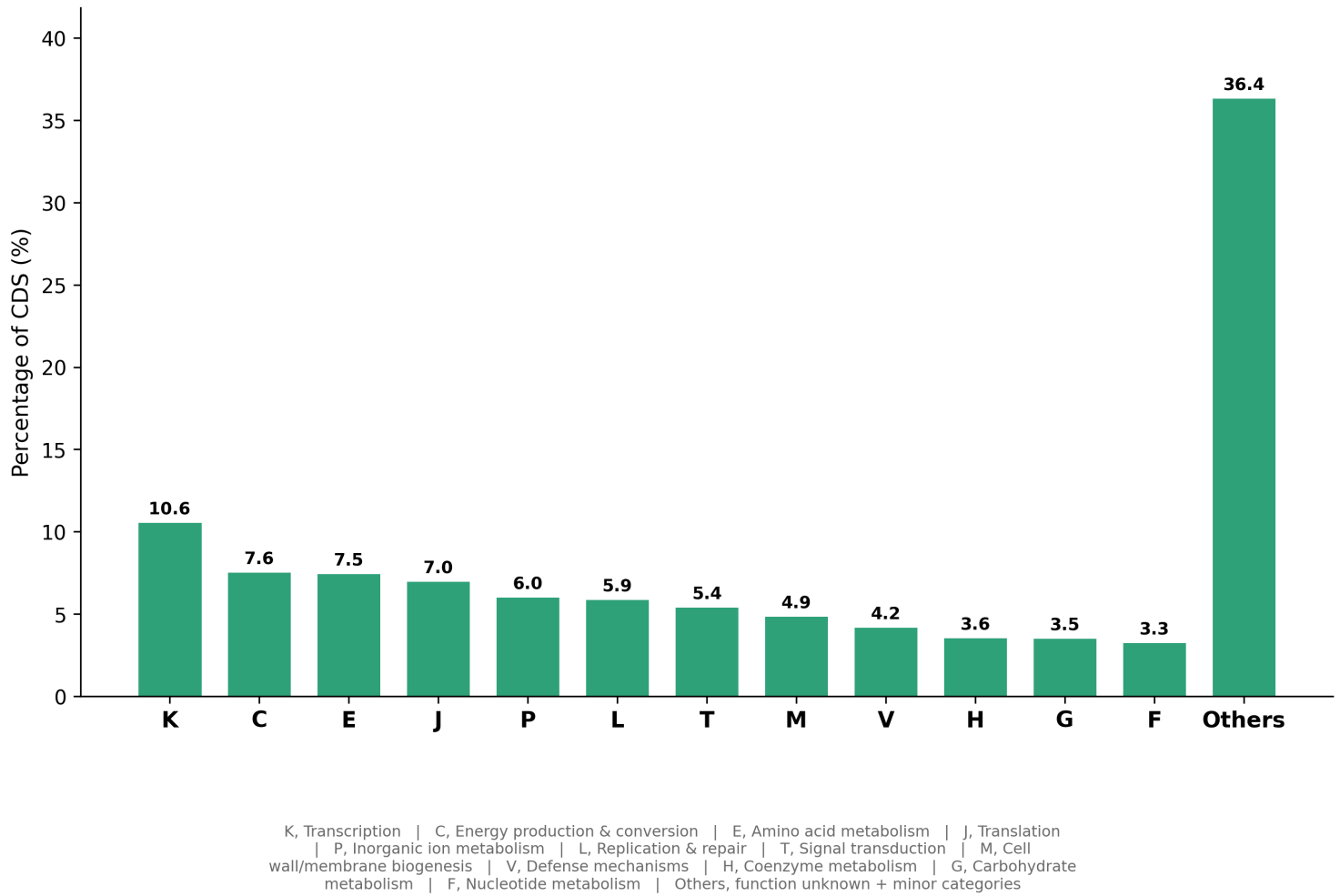


Figure 5

Functional distribution of coding sequences among COG categories in strain Marseille-QA1073^T (eggNOG).

Percentages are expressed relative to the 3,455 CDS assigned to at least one COG category; genes annotated with several categories are counted once per category. “others” includes category [S] function unknown) and minor categories category.

Environmental distribution across metagenomic biomes (IMNGS)

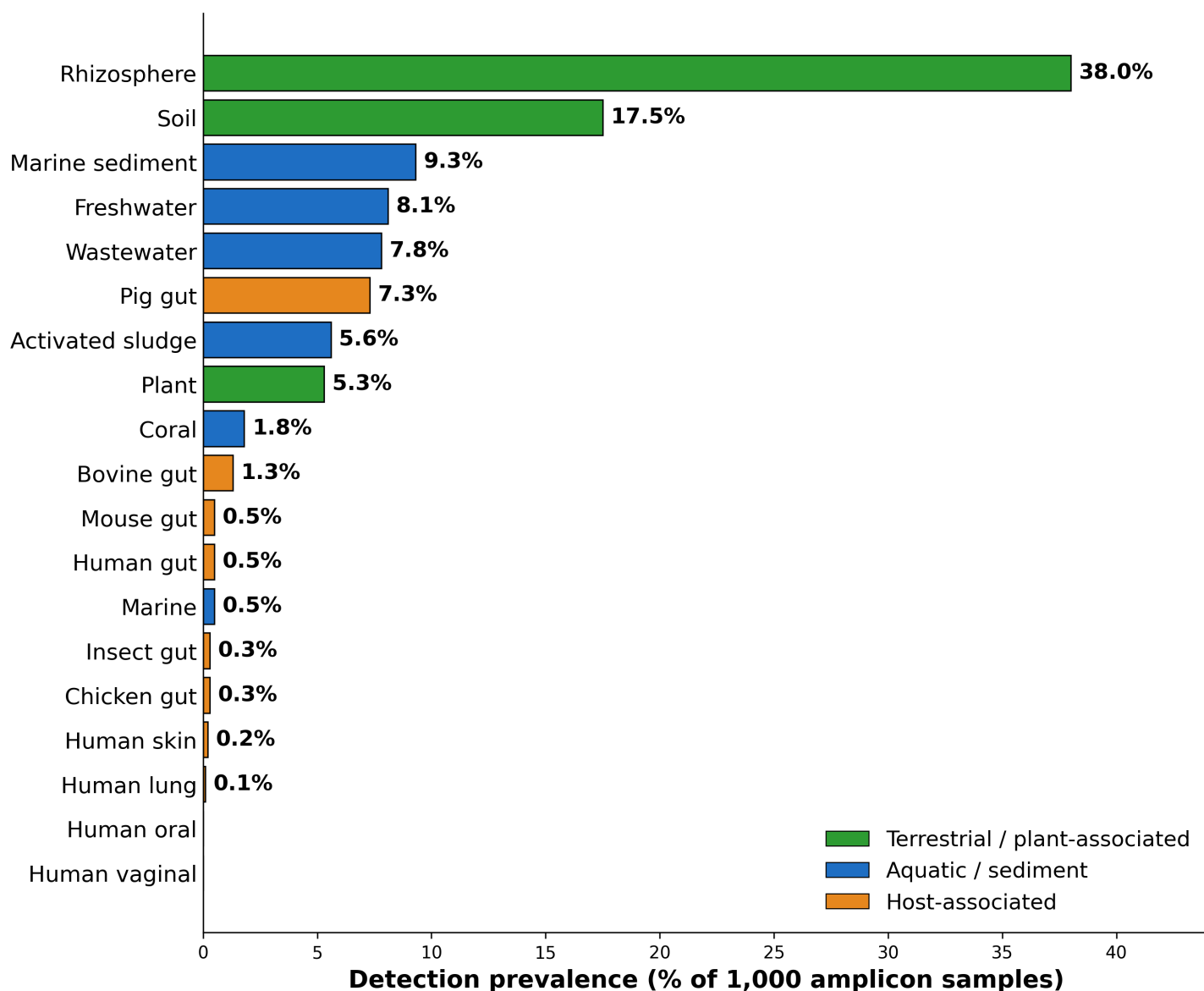


Figure 6

Environmental distribution of strain Marseille-QA1073^T across metagenomes biomes, inferred from the IMNGS/Protologger analysis of its 16S rRNA gene.

Bars represent the detection prevalence of the strain in each biome (percentage of 1,000 amplicons-sequencing samples screened per habitat), colour-coded by habitat class (terrestrial/plant-associated, aquatic/sediment, host-associated); these predictions rest on a partial 16S signature and are therefore indicative only.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableS1Genomefeaturesofstrain.docx](#)
- [TableS2Detectionprevalenceofstrain.docx](#)
- [FigS1prokseeClostridiumniakharensen.png](#)
- [Supplementarydata.docx](#)