Systematic and Applied Microbiology 44 (2021) 126277

Contents lists available at ScienceDirect

Systematic and Applied Microbiology

journal homepage: www.elsevier.com/locate/syapm

Comparative genome analysis of the genus *Hydrotalea* and proposal of the novel species *Hydrotalea lipotrueae* sp. nov., isolated from a groundwater aquifer in the south of Mallorca Island, Spain

Juan F. Gago [a,b,*], Tomeu Viver [a], Mercedes Urdiain [a], Silvia Pastor [c], Peter Kämpfer [d], Pedro A. Robledo [e], Elaine Ferreira [b], Ramon Rosselló-Móra [a,*]

[a] Marine Microbiology Group, Department of Animal and Microbial Biodiversity, Mediterranean Institute for Advanced Studies (IMEDEA, CSIC-UIB), Esporles, Spain
[b] The Deep Blue Sea Enterprise S.L., Barcelona, Spain
[c] Lipotrue S.L., Barcelona, Spain
[d] Institute of Applied Microbiology (IFZ), Justus Liebig Universität Giessen, Giessen, Germany
[e] Delegation of Geological and Mining Institute of Spain (IGME-CSIC) in Balearic Islands, Palma de Mallorca, Spain

ARTICLE INFO

Article history:
Received 12 July 2021
Revised 7 October 2021
Accepted 18 October 2021

Keywords:
Hydrotalea
Metagenomes
Geothermal anomalies
Taxonomy
Genomics

ABSTRACT

From a collection of > 140 strains isolated from groundwater with thermal anomalies for the purpose of obtaining good candidates with applications in the cosmetic industry, two strains were selected because of their taxonomic novelty. Among the isolates, strains TMF_100^T and TFM_099 stood out for their potential biotechnological relevance, and a comparative analysis of 16S rRNA gene sequences indicated that these strains represented a new species of the genus *Hydrotalea*. In addition, from the public genomic databases, metagenome-assembled genomes (MAGs) and single-cell amplified genomes (SAGs) could be retrieved that affiliated with this genus. These MAGs and SAGs had been obtained from different environmental samples, such as acid mine drainage or marine sediments. In addition to the description of the new species, the ecological relevance of the members of this genus was demonstrated by means of denitrification, CRISPR-Cas system diversity and heavy metal resistance, as well as their wide geographical distribution and environmental versatility. Supported by the taxonomic study, together with physiological and morphological differences and ecological features, we concluded that strain TMF_100^T represented a novel species within the genus *Hydrotalea*, for which we propose the name *Hydrotalea lipotrueae* sp. nov.
© 2021 The Author(s). Published by Elsevier GmbH. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Introduction

Hydrotalea Kämpfer et al. 2011 is a genus of the phylum *Bacteroidetes* classified into the family *Chitinophagaceae* with the type species *H. flava* [17]. Currently, the genus includes the two species *H. flava* [17] and *H. sandarakina* [1]. The members of this genus have a common phenotype, since they are orange- or yellow-pigmented without flexirubins. Cells are non-motile rods, Gram-stain-negative and aerobic with a salinity range of up to 1% [1]. All members described within this genus have been isolated from aquatic environments and one of them is slightly thermophilic. Recent studies have reported the potential implication of this genus in acid environments with high concentrations of metals, such as areas of mining activity [25,33]. These culture-independent metagenome surveys of marine sediments and hot springs showed the presence of metagenome assembled genomes (MAGs) that affiliated with the genus *Hydrotalea*, which exemplified the wide distribution of the family *Chitinophagaceae* [33]. The only species of this family that has been isolated from groundwater is *Heliimonas saccharivorans* L2-4 ^{T} [23] from a subterranean system similar to the one sampled in the current study.

Few insights into the ecological relevance of the *Hydrotalea* genus have been reported, especially in environments where pH or the presence of heavy metals determines the structure of microbial communities [25,33]. The origins of the two *Hydrotalea* species hitherto described represent aquatic environments, since *H. flava* was isolated from industry distilled water and *H. sandarakina* from hot spring water. However, some studies have reported that the *Hydrotalea* genus forms part of the microbiota associated with

* Corresponding authors at: Marine Microbiology Group, Department of Animal and Microbial Biodiversity, Mediterranean Institute for Advanced Studies (IMEDEA, CSIC-UIB), E-07190, Esporles, Spain.
E-mail addresses: jfgago.imedea@gmail.com (J.F. Gago), ramon@imedea.uib-csic.es (R. Rosselló-Móra).

https://doi.org/10.1016/j.syapm.2021.126277
0723-2020/© 2021 The Author(s). Published by Elsevier GmbH.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

insects [31], the gut bacterial community in freshwater herbivorous fish species [46] or ruminants [26]. The functional implication of *Hydrotalea* in their specific communities has also been described, such as for example the codification for antibiotic resistance genes in marine sediments [36]; or for the genes involved in organic carbon remediation and pyrite oxidation that indicated *Hydrotalea* was a member of the communities occurring in biological heap leaching systems used for the decomposition of ores [51]. Overall, there is evidence for the wide distribution of this taxon in different ecosystems, such as terrestrial or aquatic environments. Finally, their gene content and metabolic potential could make these organisms good candidates with potential applications in the biotechnological industry.

The current study reports the identification of a new member of the genus *Hydrotalea* obtained as a pure culture from a collection of > 140 strains isolated from several groundwater samples with thermal anomalies occurring in the south of the island of Mallorca, Spain. The results contributed to revealing the presence of *Hydrotalea* in these environments, since cultivable strains and metagenome assembled genomes (MAGs) were isolated from the groundwater samples. In addition, the study led to the description of a new species of the genus *Hydrotalea* represented by strain TMF_100^T for which we propose the name *Hydrotalea lipotrueae* sp. nov. The ecological relevance and environmental versatility of the members of the genus *Hydrotalea* were also shown.

Materials and methods

Sampling and bacterial culture

Two bacterial strains, TMF_100^T and TFM_099, were isolated after direct plating from water collected in November 2016 from an aquifer in the south of Mallorca Island (39°28′44.1″N 2°53′09.6″E), during part of a bioprospecting study associated with groundwater samples for understanding their potential uses in the cosmetic industry. Another five additional samples were obtained from the groundwater in different areas of Mallorca Island and their general features are detailed in Supplementary Fig. S1 and Supplementary Spreadsheet S1. On the same day of collection, samples were serially diluted using sterile buffer PBS1x (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4 and 1.8 mM KH_2PO_4) spread on different media (R2A, nutrient and M6 agar) and incubated at 30 °C, 40 °C and 50 °C until colonies were visible after a minimum period of two weeks and a maximum of three months. Subsequently, colonies were subcultured on their original medium in order to obtain pure cultures. Identification was carried out using the tandem approach of MALDI-TOF MS and 16S rRNA gene analysis, as described in previous studies [3,14,35,48,49]. From the six samples, >140 microorganisms were isolated, of which two of them (TMF_100^T and TFM_099) affiliated with the genus *Hydrotalea* using the 16S rRNA gene (see below). Pure cultures were stored in cryoprotectant solution containing 13% (w/v) glycerol at −80 °C. Random amplified polymorphic DNA (RAPD; [43]) was used to reveal clonality between the strains. The RAPD primers 241–2120 (5′-CTCAATGGCAGCGGCTATGG-3′) and 241 + 9 (5′-GTT TCGCTCGATGCGCTACC-3′) were applied for this purpose.

Phenotypic and chemotaxonomic characterization

The cellular morphology was examined using differential interference contrast (DIC, Nomarsky) microscopy (Zeiss Axio Imager) after fixing the culture with 3% formaldehyde. Gliding motility was analyzed using either the hanging drop method or a modified protocol described by [7]. The growth temperature range (4, 10, 15, 20, 25, 30, 35, 40, 45, 50 and 55 °C), pH range (4–11 with increments of 1 pH unit) and growth with added NaCl up to 6% (w/v) were undertaken in R2A broth after observing turbidity in the liquid medium for up to 7 days. The pH of the medium was adjusted using the following buffer systems: pH 5.0–8.0, 0.1 M citric acid/0.2 M Na_2HPO_4; pH 8.0–9.0, 0.1 M Tris/0.1 M HCl; pH 9.5–11.0, 0.05 M $NaHCO_3$/0.1 M NaOH [14]. The presence of flexirubin-type pigment in the culture was detected using 20% (w/v) KOH solution as described by [6]. Hydrolysis of casein, starch, tyrosine, Tweens 20 and 80 were determined according to the method described by [11]. Catalase activity was detected by bubble production in 3% (v/v) aqueous hydrogen peroxide solution and oxidase activity was determined using swabs impregnated with N,N-dimethyl-p-phenylenediamine dihydrochloride (Scharlab, ref: 06–120-050). Anaerobic growth on R2A agar was performed in an anaerobic jar using GENbox (BioMérieux, Lion). Other biochemical and physiological characteristics were carried out with API ZYM and API 20 NE (BioMérieux), according to the manufacturer's instructions, incubated at 29 °C for 4 h and 32 °C for up to 6 days, respectively. For the latter API test, results related to sole carbon assimilation were discarded, since further analysis for the same purpose was performed according to the method published by [18]. Acid production from carbohydrates was determined using API 50 CH, with API 50 CHB/E medium, incubated at 32 °C and the results were read for up to 6 days. Fatty acids were extracted from cell lysates of strain TMF_100^T grown for 48 h at 28 °C on R2A agar. Analysis of these lipid compounds was carried out according to the method described by Kämpfer and Kroppenstedt [16] through the separation of fatty acid methyl esters using gas chromatography (5898A, Hewlett Packard). Peaks were automatically integrated and fatty acid names and percentages were determined with Sherlock MIDI version 2.1 (TSBA version 4.1).

Genome sequence analysis

Sequencing the genome of strain TMF_100^T was performed using Illumina NovaSeq (2 × 150 bp, paired end reads) at the Macrogen Company, South Korea. From raw reads, the subsequent steps, which involved filtering, assembling and gene prediction, were performed as described before [14]. The genome sequence of strain TMF_100^T was deposited in the NCBI database under the accession number JAESDL000000000. The genomes of both available type strains *H. flava* CCUG 51397^T (LUHG01000000) and *H. sandarakina* DSM 23241^T (QKZV01000000), including the new isolate TMF_100^T, together with the closest relative MAGs and SAGs (single-amplified genomes; see below) affiliated with the genus *Hydrotalea*. All the genomes are available in the European Nucleotide Archive (ENA; https://www.ebi.ac.uk/ena/browser/home), and they were annotated using the protocol described by Roth E. Conrad (https://github.com/rotheconrad) for the SwissProt and TrEMBL databases. Furthermore, comparative annotation was performed between the results from the last step and those given by rapid annotation using subsystem technology (RAST; [4]) and the Kyoto Encyclopedia of Genes and Genomes (KEGG; [19]). From the latter database, a list of Kegg orthologs (KO) was used to reconstruct the metabolic pathways. CRISPR spacer prediction was carried out for all of them using the bioinformatic tool CRISPRCasfinder, available as an online version at https://crispr-cas.i2bc.paris-saclay.fr [10]. Venn diagrams were performed by using the online tool from the Van de Peer Lab (http://bioinformatics.psb.ugent.be/software/details/Venn-Diagrams). The average nucleotide identity (ANI) between all the genomes was calculated using the BLAST (ANIb) algorithm through the JSpecies web server (v3.7.3; [40]). Average amino acid identity (AAI) was also calculated according to Konstantinidis and Tiedje [21] using the script *ani.rb* from the Enveomics Collection [41]. From this collection,

the *HMM.essential.rb* script was used to estimate completeness and contamination for each genome analyzed in this study.

Metagenome analysis

Samples T2 and T6 were sequenced using Illumina HiSeq (2 × 100 bp, paired end reads) and Illumina Nextera DNA XT (2 × 150 bp, paired end reads), respectively. Metagenomic raw reads were trimmed using Prinseq-lite tool v.0.20.4 by applying the following parameters: min_length: 50, trim_qual_right: 30, trim_qual_type: mean and trim_qual_window 20. To compare the novel strain TMF_100^T with similar genomes in other environments where *Hydrotalea* genomes had been retrieved, a total of ten metagenomic datasets from acid mine water, marine sediments and hot springs were used for this purpose. The origin of these metagenomes was Brazil (4) [25,33], the Gulf of Khambhat (4) [36], the Arabian Sea (1) [36] and Yellowstone (1) [53], which were downloaded from ENA and JGI (Joint Genome Institute; https://genome.jgi.doe.gov/portal/). Details concerning the dataset are summarized in Supplementary Spreadsheet S2.

Recruitment of metagenomic reads against the reference genomes of genus *Hydrotalea* (*H. flava* CCUG 51397^T and *H. sandarakina* DSM 23241^T) and strain TMF_100^T was carried out using BLASTn [54], selecting reads with ≥ 95% similarity and an alignment length ≥ 70%. In order to obtain the relative abundance of the analyzed genomes (% of reads), the number of mapped reads was divided by the total number of reads in each metagenome and divided again by the size (bp) of the total length of the genome. From recruitment plots, average and sequencing breadth were calculated using the Enveomics Scripts Collection [41]. Metagenomes of samples T2 and T6 have been deposited under study number PRJEB46206.

Metagenome-assembled genomes (MAGs) and single amplified genomes (SAGs)

A total set of ten MAGs and six SAGs (single amplified genomes) identified as *Hydrotalea* species were available in the European Nucleotide Archive (ENA), (https://www.ebi.ac.uk/ena/browser/home) database [25,33,34,36], and they were included in the present study in order to evaluate the ecological distribution of the members of this genus (Fig. 1). In addition, eight MAGs originated from a marine sediment metagenome in the Gulf of Khambhat and they were available from within the project PRJNA598413 with the following accession numbers: WVWQ01000001, WVXH01000001, WVYD01000001, WVZD01000001, WVZT01000001, WWAN01000001, WWBA01000001 and WWBX01000001. A single MAG with accession number DTJW01000000 assembled from a hot spring metagenome sample from Yellowstone National Park in October 2012 (detailed information available under project accession number PRJNA480137), and one MAG from the Sossego Mine (June 2014) with the accession number SCKY01000001 from

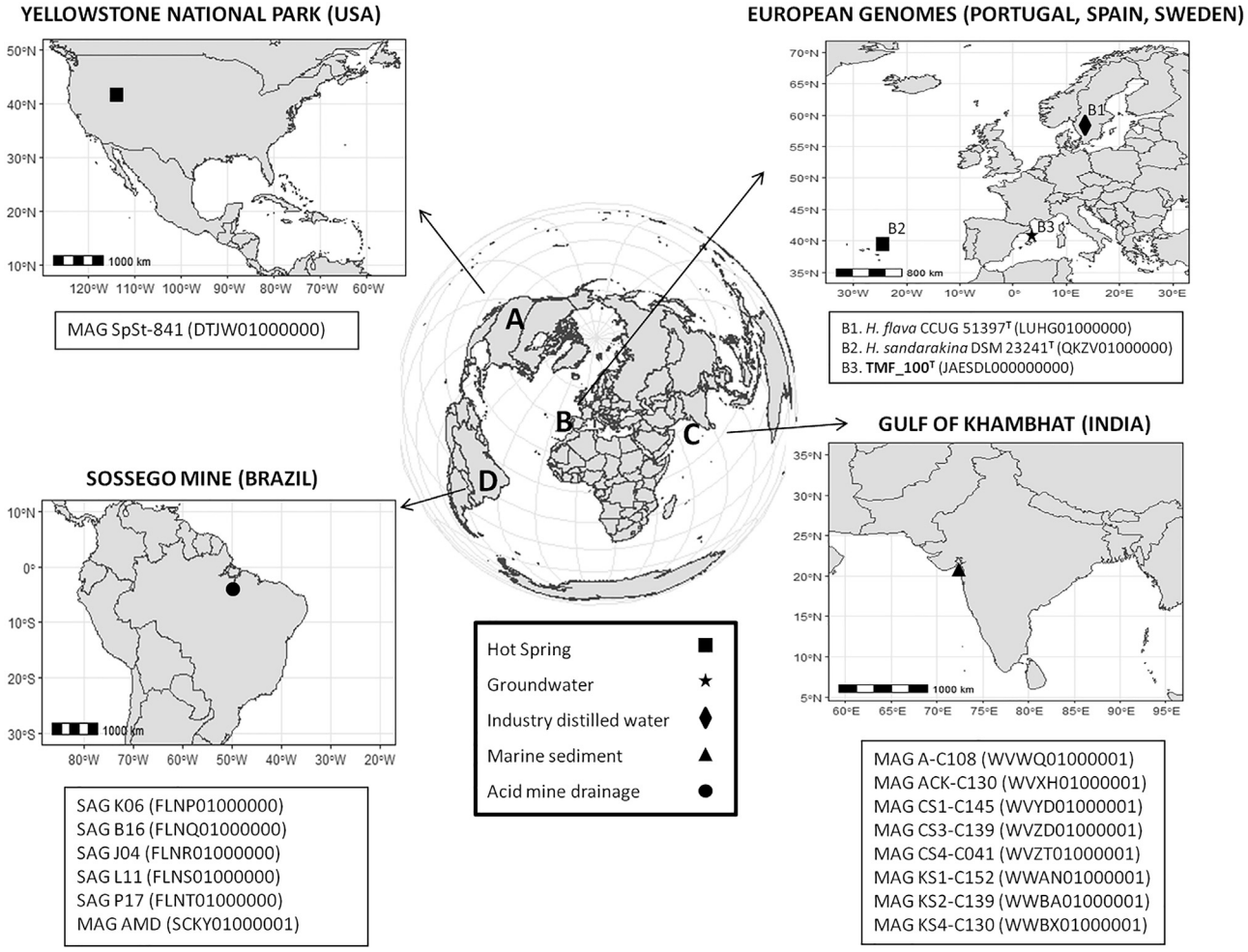

Fig. 1. Geographical distribution of the original samples from which one or more genome sequences were obtained, including the type strains *H. flava* CCUG 51397^T and *H. sandarakina* DSM 23241^T. The name of the genomes retrieved from each location is given in the corresponding boxes, together with their accession numbers in brackets.

within the project PRJNA513256 were also included. Finally, six SAGs were incorporated that had also been obtained from the same acid mine drainage site from Brazil submitted within the project PRJNA513256 under accession numbers FLNO01000000, FLNP01000000, FLNQ01000000, FLNR01000000, FLNS01000000 and FLNT01000000. The map with the geographical distribution of the MAGs and SAGs, including strain TMF_100^T, was created using the packages ggplot2, sf, rnaturalearth and rnaturalearthdata in R Studio v 1.1.463 (http://www.r-project.org, TeamR RStudio, 2016).

Phylogenetic reconstruction based on 16S rRNA, core and housekeeping genes

A phylogenetic tree was constructed based on 16S rRNA gene sequences retrieved from a set of genomes analyzed in the present study, which included the strain to characterize, as well as their closest relatives and the MAGs and SAGs whose 16S gene sequence was present in the genome. According to this, the small subunit ribosomal was extracted using the barrnap v0.9 tool (https://github.com/tseemann/barrnap) and then imported into the latest updated LTP_12_2020 database [30] available at https://imedea.uib-csic.es/mmg/ltp, and aligned by using the SINA v1.3.1 aligner [38] implemented in the ARB program package [29]. Final alignments were manually improved following the reference alignment in ARB Editor. Sequences were used to reconstruct *de novo* trees based on the RAxML [44] method using GTRGAMMA correction.

The predicted protein sequences of a partial set of genomes included in the study, comprising in this case the described species, the closest relatives and MAGs, were used in order to perform phylogenetic reconstruction based on single-copy core genes and a set of 101 essential genes. Predicted protein sequences were compared using an all-versus-all BLAST (v2.2.31), identifying the shared reciprocal best matches according to the method described previously [14]. After aligning all the protein sequences shared between them using MUSCLE v3.8.31 [12], these were concatenated and used to build a phylogenetic tree based on the RAxML [44] algorithm with PROTGAMMA correction [22]. A phylogenetic analysis based on housekeeping genes was also performed from the same genomes as those mentioned above and, after applying the methodology previously published [14], a reconstructed tree with aligned protein sequences was carried out again using the RAxML algorithm with PROTGAMMA correction.

Results and discussion

As part of a wider study related to the bioprospecting of potentially valuable bacterial resources from subsurface habitats, a total of six groundwater samples were collected for this purpose in November 2016. The direct plating approach rendered a collection of ~ 140 strains, two of which affiliated taxonomically with the genus *Hydrotalea*. Both of these two strains, TMF_100^T and TFM_099, were from the same sample (T2) obtained from the Lluc-major area of Mallorca at a temperature of 28.7 °C. TMF_100^T and TFM_099 were isolated from the original sample with no previous enrichment and were shown to be coexisting clonal varieties in accordance with the RAPD analysis that had identical banding patterns for both strains. Therefore, as both isolates were considered to be the same strain, only isolate TMF_100^T was selected for further analysis (Supplementary Fig. S2).

Phylogenetic analyses

All genomes, MAGs and SAGs identified as *Hydrotalea* present in the ENA repository were compared. The two almost complete genome sequences of the two type strains *H. flava* CCUG 51397^T isolated from water samples in Sweden between July 2005 and December 2006 [17,24] and *H. sandarakina* DSM 23241^T obtained

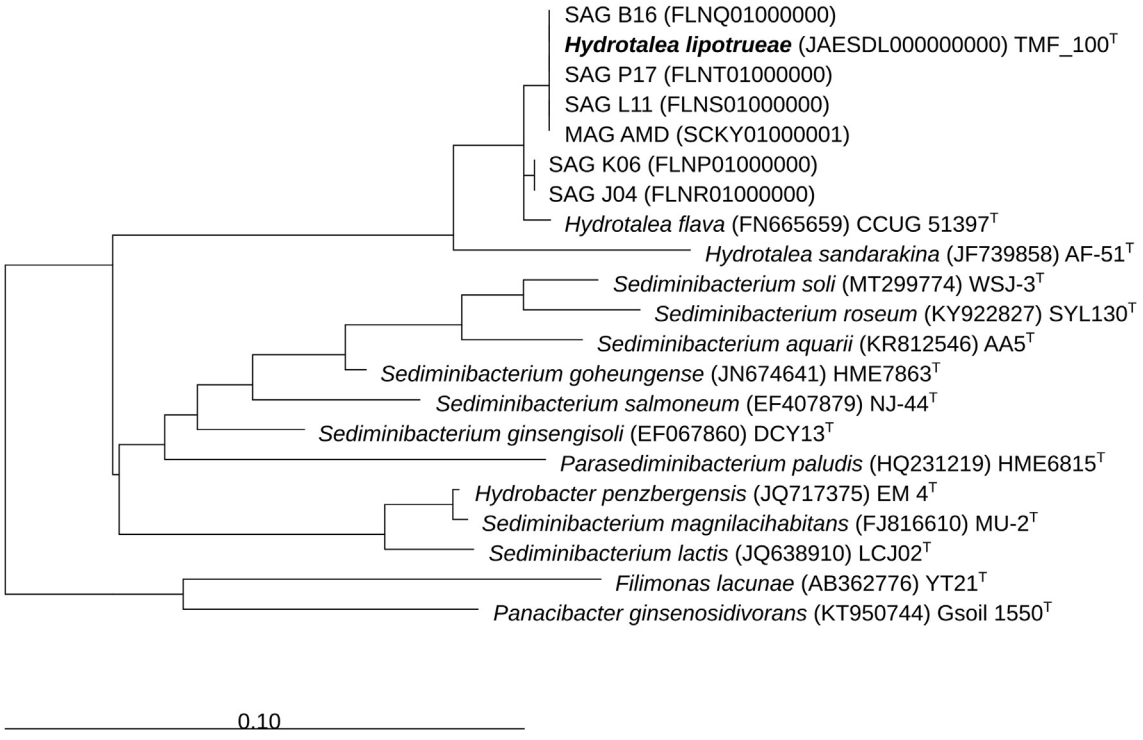

Fig. 2. Phylogenetic reconstruction based on the 16S rRNA gene sequence analysis of all *Hydrotalea* species available in the LTP_12_2020, as well as the available SAGs and MAGs affiliated with this genus. The tree was reconstructed using the maximum likelihood algorithm and no filter was applied (except the termini to remove the gene flanking sequences) in order to take into account all variable positions. Bar indicates 10% sequence divergence.

from hot spring water on the island of São Miguel, Azores [1], represented the only cultivated members of the genus. On the other hand, ten MAGs (A-C108, ACK-C130, CS1-C145, CS3-C139, CS4-C041, KS1-C152, KS2-C139, KS4-C130, AMD and SpSt-841) and six SAGs (A04, K06, B16, J04, L11 and P17) reconstructed from microbial metagenomes of different environments were found, and they affiliated with members of the genus *Hydrotalea* (Fig. 1). Eight MAGs were assembled from a marine sediment metagenome of the Gulf of Khambhat within the study carried out by the University of Maharaja Krishnakumarsinhji Bhavnagar [34; 36] that, according to the description in the ENA database, all came from the same site and had the same data (April 2017). One MAG was assembled from a hot spring metagenome sample collected in Yellowstone National Park in October 2012 by researchers from Hong Kong University [53]. Another MAG was obtained from water samples from Sossego Mine (June 2014; [25]) as part of the project led by the Federal University of Lavras (UFLA; Brazil). Finally, six SAGs from the same acid mine drainage site in Brazil were obtained from a previous study also carried out by UFLA [33] using samples collected in July 2014 for single-cell genomics investigation. Therefore, given the possibility to investigate the biogeography and distribution of this bacterial taxa, these genomes were included in the current study.

16S rRNA gene sequences were detected in five out of the six SAGs and only one out of the ten MAGs. This could be due to the fact that the initial screening and selection of SAGs was based on amplification of the bacterial small subunit rRNA, as already reported [33]. The lack of 16S rRNA gene sequences in MAGs is not uncommon and could be associated with issues related to the assembly of these genes into metagenomes [2]. As a result of the 16S rRNA gene sequence phylogenetic reconstruction, two clades could be observed. The first clade, which embraced most of the sequences involved in this study, grouped strain TMF_100^T, MAG AMD and SAGs B16, L11 and P17. The second clade included SAGs K06 and J04, and was independent from the first clade and type strain *H. flava* CCUG 51397^T (Fig. 2). *H. sandarakina* AF-51^T affiliated separately from all *Hydrotalea* sequences. The 16S rRNA gene sequence identity between all members of the first clade was 100%, and showed an identity of 99.3% and 96.0% with the type strains *H. flava* CCUG 51397^T and *H. sandarakina* AF-51^T, respectively. The second clade was formed exclusively by SAGs K06 and J04, with a reciprocal identical 16S rRNA gene sequence sharing 99.5% identity with the members of the first clade. *H. flava* CCUG 51397^T was once again the closest relative species with 99.6%, followed by *H. sandarakina* AF-51^T with 96.2% (Supplementary Spreadsheet S3).

Given that some MAGs and SAGs did not include the 16S rRNA gene sequence, the genomic uniqueness and discreteness were assessed using a phylogenomic approach. The low completeness of the SAGs hampered their inclusion in the analysis; however, the ten MAGs were considered due to their completeness (Table 1). It was corroborated based on a maximum likelihood reconstruction using 1,080 core genome genes (Fig. 3) or 73 essential housekeeping genes (Supplementary Fig. S3) that the new isolate formed an independent lineage together with the MAGs and some of the SAGs studied. The lack of a 16S rRNA gene in some MAGs and the low completeness of some SAGs was solved by the use of the type strain genomes together with MAG AMD that could be included in both approaches (16S rRNA gene and genome trees), which helped to resolve the branching pattern of the clades. The MAGs from Brazil (AMD) and the Gulf of Khambhat (A-C108, ACK-C130, CS1-C145, CS3-C139, CS4-C041, KS1-C152, KS2-C139 and KS4-C130) affiliated with strain TMF_100^T, whereas MAG SpSt-841 from Yellowstone was associated with *H. sandarakina* AF-51^T (Fig. 2 and Supplementary Fig. S3).

Table 1
Main features of the genomes, MAGs and SAGs affiliated to *Hydrotalea*, including the type strains *Hydrotalea flava* CCUG 51397^T and *Hydrotalea sandarakina* DSM 23241^T. Bold type letter corresponds to strain TMF_100^T described in this study.

Location	Genome	Accession number	Completeness (%)	Contamination (%)	Contigs	Size (bp)	N50 (bp)	L50	% GC	CDS	rRNA	tRNA
Mallorca – Spain (Groundwater)	***Hydrotalea lipotrueae* TMF_100^T**	JAESDL000000000	97.1	0	70	3,465,771	111,497	10	37.0	3,297	3 (23S/5S/16S)	39
Gulf of Khambhat – India (Marine sediments)	MAG KS2-C139	WVBA01000001	81.4	1	558	3,163,410	10,556	97	37.0	3,539	1 (5S)	34
	MAG KS4-C130	WVBX01000001	81.4	1	613	3,193,437	10,329	98	37.0	3,584	1 (5S)	34
	MAG KS1-C152	WVAN01000001	81.4	1	582	3,201,758	10,586	97	37.0	3,597	1 (5S)	35
	MAG CS4-C041	WVZT01000001	81.4	1	562	3,190,995	10,586	97	37.0	3,568	1 (5S)	35
	MAG CS1-C145	WVYD01000001	81.4	1	614	3,185,564	10,329	98	37.0	3,581	1 (5S)	34
	MAG CS3-C139	WVZD01000001	81.4	1	586	3,179,459	10,556	97	37.0	3,563	1 (5S)	34
	MAG ACK-C130	WVXH01000001	81.4	1	613	3,193,437	10,329	98	37.0	3,584	1 (5S)	34
	MAG A-C108	WVWQ01000001	81.4	1	568	3,169,773	10,556	97	37.0	3,545	1 (5S)	35
Carajás – Brazil (Acid mine drainage)	MAG AMD	SCKY01000001	97.1	0	113	3,931,870	147,328	9	37.0	3,941	5 (1 23S/3 5S/1 16S)	62
	SAG L11	FLNS01000000	42.2	0	92	1,105,265	22,041	16	36.9	1,148	3 (23S/5S/16S)	13
	SAG J04	FLNR01000000	47.1	0	116	1,579,673	21,646	23	36.8	1,650	3 (23S/5S/16S)	18
	SAG B16	FLNQ01000000	50	10.8	125	1,752,129	24,033	23	36.8	1,765	3 (23S/5S/16S)	23
	SAG K06	FLNP01000000	52	0	90	1,106,849	23,214	15	37.1	1,153	3 (23S/5S/16S)	17
	SAG P17	FLNT01000000	51	0	131	2,082,446	26,556	25	36.8	2,094	3 (23S/5S/16S)	25
	SAG A04	FLNO01000000	34.3	0	68	710,749	18,064	13	36.5	749	2 (23S/5S)	6
Yellowstone – USA (Hot spring water)	MAG SpSt-841	DTJW01000000	97.1	2	103	3,149,885	52,395	20	35.5	2,930	2 (5S)	35
Sweden (Industry distilled water)	*Hydrotalea flava* CCUG 51397^T	LUHG01000001	97.1	0	116	3,413,796	83,650	11	37.2	3,328	3 (23S/5S/16S)	39
São Miguel, Azores – Portugal (Hot spring water)	*Hydrotalea sandarakina* DSM 23241^T	QKZV01000001	97.1	0	34	3,268,305	230,815	5	35.7	2,953	5 (1 23S/2 5S/2 16S)	43

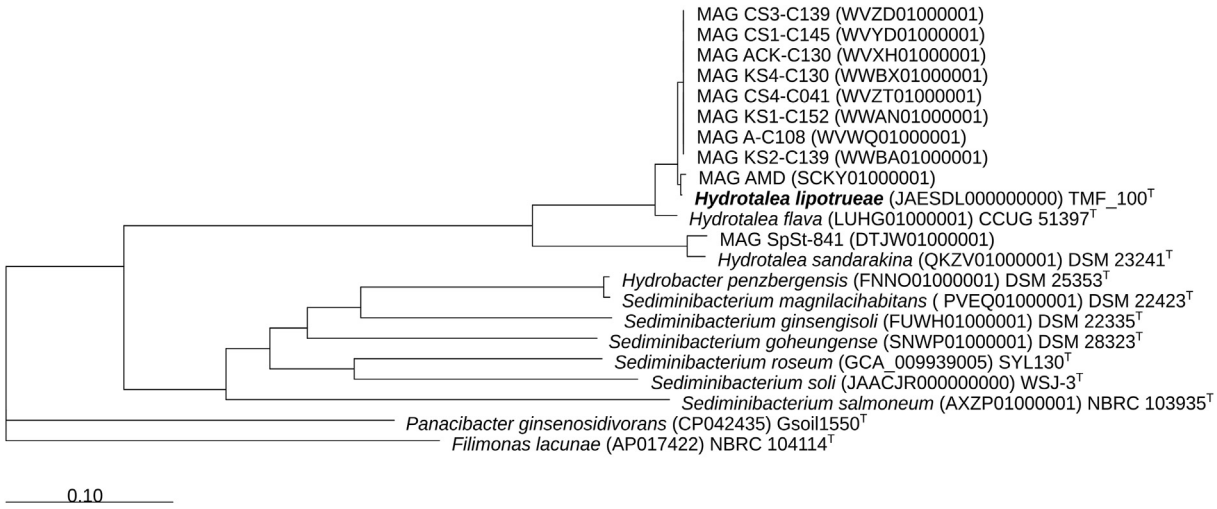

Fig. 3. Phylogenetic reconstruction based on 1,080 concatenated core orthologous gene sequences of all available *Hydrotalea* genomes, MAGs and SAGs, together with the genome of the new strain TMF_100^T. The tree was reconstructed using the maximum likelihood algorithm. Bar indicates 10% sequence divergence. The accession number of each sequence is given in brackets.

ANI [40] and AAI [21] values confirmed that the isolate described in this study represented a novel species (Table 2). Strain TMF_100^T showed an ANI value of 93.60% and an AAI value of 93.94% with its closest relative *H. flava* CCUG 51397^T, which was clearly below the species threshold level [39]. Interestingly, it was also found that the six SAGs (A04, K06, B16, J04, L11 and P17) together with the MAG from acid mine drainage (AMD) and those from marine sediment in the Gulf of Khambhat (A-C108, ACK-C130, CS1-C145, CS3-C139, CS4-C041, KS1-C152, KS2-C139 and KS4-C130) shared ANI values between 98.24% and 99.16% with strain TMF_100^T, which reinforced the idea that they all belonged to the same species, as observed in the phylogenetic analyses. In addition, the MAGs originating from marine sediments (A-C108, ACK-C130, CS1-C145, CS3-C139, CS4-C041, KS1-C152, KS2-C139 and KS4-C130) shared ∼ 100% ANI among themselves, and ANI and AAI values of < 95% with *H. flava* CCUG 51397^T and *H. sandarakina* AF-51^T. MAG SpSt-841 showed ANI and AAI values of 94.18% and 96.56% with *H. sandarakina* AF-51^T that, in accordance with the 16S rRNA gene and genome trees, could represent a new as yet uncultivated species of the genus.

Morphological, phenotypic and chemotaxonomic characterization of the novel species

As the members of the clades could be considered to be a new genomospecies within the genus *Hydrotalea*, phenotypic analyses were performed in order to formally name it as a new species. Consequently, colonies of TMF_100^T were round with a regular shape, flat and a smooth surface on R2A, with a diameter of 0.5–1.0 mm and orange pigmentation, as possessed by other member species of *Hydrotalea* (Supplementary Fig. S4; [1,17]). Unlike the *Hydrotalea* species described to date, the novel strain was motile by gliding. The presence of flexirubin-type pigment was not detected. Cell morphology was rod-shaped, between 1.0–5.0 μm in length and 0.3–0.5 μm in width, similar to the other members of this genus. Strain TMF_100^T did not grow at temperatures below 15 °C but grew up to 45 °C, with optimum growth at 30 °C. Growth was visible within the pH range from 4 to 9 and a saline concentration of up to 1%. Based on biochemical and enzymatic analysis, isolate TMF_100^T was positive for hydrolysis of casein, starch, catalase, oxidase, Tween 20, alkaline phosphatase, esterase (C4), esterase lipase (C8), leucine arylamidase, valine arylamidase, cysteine arylamidase, trypsin, α-chymotrypsin, acid phosphatase, naphthol-AS-BI-phosphohydrolase, α-galactosidase, β-galactosidase, α-glucosidase, n-acetyl-β-glucosaminidase, α-mannosidase, glucose fermentation (weakly positive after 48 h), aesculin hydrolase and 4-nitrophenyl-β-D-galactopyranoside hydrolysis. In addition, results were negative for hydrolysis of tyrosine, Tween 80, lipase (C14), β-glucuronidase, α-fucosidase, nitrate reduction, indole production, arginine dihydrolase, urease and gelatinase. In order to obtain the best knowledge of single carbon source assimilation in strain TMF_100^T, the method described by [18] was used. Accordingly, D-fructose, D-galactose, gluconate, D-glucose, D-maltose and D-saccharose were assimilated as sole carbon sources (weak positive results after 72 h incubation); but not acetate, propionate, *N*-acetylgalactosamine, *N*-acetylglucosamine, L-arabinose, D-cellobiose, D-mannose, glycerol, D-mannitol, maltitol, α-D-melibiose, L-rhamnose, D-ribose, salicin, D-xylose, adonitol, *i*-inositol, D-sorbitol, putrescine, *cis*-aconitate, *trans*-aconitate, 4-aminobutyrate, adipate, azelate, fumarate, glutarate, DL-3-hydroxybutyrate, itaconate, DL-lactate, 2-oxoglutarate, pyruvate, suberate, citrate, mesaconate, L-alanine, β-alanine, L-ornithine, L-phenylalanine, L-serine, L-aspartate, L-histidine, L-leucine, L-proline, L-tryptophan, 3-hydroxybenzoate, 4-hydroxybenzoate and phenylacetate. The results of API 50 CH showed a very similar pattern in comparison to the other members described in this genus, in particular *H. flava* CCUG 51397^T (Table 3). Acid formation was detected from D-xylose, D-galactose, D-glucose, D-fructose, D-mannose, L-rhamnose, methyl-αD-mannopyranoside, methyl-α D-glucopyranoside, N-acetylglucosamine, amygdalin, arbutin, aesculin ferric citrate, salicin, D-cellobiose, D-maltose, D-lactose, D-melibiose, D-saccharose, trehalose, melezitose, raffinose, starch, gentiobiose, and turanose. The fatty acid profile (Table 4) showed branched iso-$C_{15:0}$ and iso-$C_{17:0}$ 3-OH as the major components present among the members of the genus *Hydrotalea*. Despite the fact that the profiles were very similar between themselves, certain differences could be found in *H. sandarakina* AF-51^T with a relative abundance of iso-$C_{16:0}$.

The diagnostic phenotype of TMF_100^T is shown in Table 3. Based on the physiological comparison between the different members of this genus, several traits, such as cell morphology, pigmentation, fatty acid profiles or acid-production, revealed a low

Table 2
Pairwise genomic comparison of the genomes affiliated to *Hydrotalea* based on amino acid identity (ANI: lower triangle; values in brackets represent the percentage of aligned nucleotides) and protein translated genes (AAI: upper triangle; values in brackets represent the percentage of shared proteins). The accession number for each genome is given in brackets.

	Genome size Mbp	G + C mol%	Completeness	Contamination	*Hydrotalea lipotrueae* TMF_100^T	SAG A04	SAG K06	SAG B16	SAG J04	SAG L11	SAG P17	MAG AMD	MAG A-C108	MAG ACK-C130	MAG CS1-C145	MAG CS3-C139	MAG CS4-C041	MAG KS1-C152	MAG KS2-C139	MAG KS4-C130	*Hydrotalea flava* CCUG 51397^T	MAG SpSt-841	*Hydrotalea sandarakina* DSM 23241^T
H. lipotrueae TMF_100^T(JAESDL000000000)	3.46	37	97.1	0	*	95.40 [90.88]	97.09 [91.89]	97.45 [88.92]	95.78 [87.49]	96.83 [90.61]	97.86 [88.77]	98.48 [94.04]	99.00 [90.50]	99.04 [90.47]	99.02 [90.63]	99.00 [90.63]	99.02 [90.50]	99.02 [90.60]	99.02 [90.63]	99.04 [90.47]	93.94 [85.14]	80.68 [84.39]	80.65 [85.56]
SAG A04 (FLNO01000000)	0.7	36.5	34.3	0	98.94 [88.62]	*	73.26 [54.56]	86.33 [69.69]	87.26 [67.10]	74.37 [54.98]	82.29 [71.22]	98.21 [93.87]	94.27 [88.88]	94.21 [89.03]	94.24 [88.74]	94.22 [88.88]	94.27 [88.88]	94.16 [89.03]	94.24 [88.74]	94.21 [89.03]	94.02 [90.45]	77.50 [83.05]	78.00 [83.47]
SAG K06 (FLNP01000000)	1.1	37.1	52	0	98.50 [92.13]	98.88 [32.17]	*	81.62 [57.51]	84.16 [61.66]	76.90 [47.39]	86.13 [70.41]	97.88 [92.62]	95.92 [89.31]	95.92 [89.31]	95.94 [89.21]	95.92 [89.31]	95.92 [89.31]	95.92 [89.31]	95.94 [89.21]	95.92 [89.39]	92.73 [87.93]	79.72 [81.47]	79.91 [82.02]
SAG B16 (FLNQ01000000)	1.75	36.8	50	10.8	99.00 [89.88]	98.33 [55.91]	98.04 [42.03]	*	84.99 [55.79]	89.68 [66.54]	92.90 [66.10]	99.85 [95.36]	96.62 [87.06]	96.55 [87.17]	96.64 [87.11]	96.58 [87.11]	96.60 [87.17]	96.57 [87.17]	96.58 [87.17]	96.55 [87.17]	93.20 [85.49]	78.21 [74.71]	77.99 [76.09]
SAG J04 (FLNR01000000)	1.58	36.8	47.1	0	98.24 [87.83]	99.13 [54.49]	99.16 [45.59]	97.60 [44.93]	*	80.98 [56.59]	89.03 [65.32]	97.50 [90.67]	95.38 [84.11]	95.38 [84.11]	95.38 [84.18]	95.38 [84.18]	95.41 [84.05]	95.38 [84.18]	95.38 [84.18]	95.38 [84.11]	92.56 [83.52]	78.53 [75.91]	78.49 [77.56]
SAG L11 (FLNS01000000)	1.1	36.9	42.2	0	98.93 [90.00]	97.48 [33.76]	98.04 [31.07]	99.21 [55.31]	97.34 [40.78]	*	89.20 [67.94]	99.61 [95.91]	95.42 [87.64]	95.37 [87.73]	95.35 [87.64	95.42 [87.64]	95.42 [87.64]	95.42 [87.64]	95.42 [87.54]	95.37 [87.73]	91.11 [82.80]	75.68 [77.13]	76.35 [78.34]
SAG P17 (FLNT01000000)	2.08	36.8	51	0	98.87 [89.51]	98.68 [53.26]	98.23 [56.51]	99.44 [59.32]	98.31 [56.44]	99.51 [58.49]	*	99.92 [96.32]	97.33 [84.43]	97.30 [84.48]	97.37 [84.43]	97.33 [84.43]	97.33 [84.48]	97.31 [84.48]	97.30 [84.53]	97.30 [84.49]	93.54 [83.35]	78.82 [74.63]	78.89 [75,59]
MAG AMD (SCKY01000001)	3.93	37	97.1	0	98.86 [93.45]	99.32 [95.23]	98.98 [93.21]	99.98 [97.94]	98.85 [91.71]	99.98 [96.92]	99.98 [97.77]	*	98.22 [82.84]	98.21 [81.99]	98.19 [82.34]	98.22 [82.53]	98.19 [82.41]	98.17 [82.05]	98.20 [83.18]	98.21 [81.99]	94.42 [87.80]	80.44 [85.13]	80.50 [86.22]
MAG A-C108 (WVWQ01000001)	3.17	36.9	81.4	1	99.16 [92.78]	98.26 [77.68]	98.20 [80.17]	98.68 [80.14]	97.86 [77.62]	98.27 [77.97]	98.42 [78.01]	98.66 [90.78]	*	100.00 [99.58]	100.00 [99.18]	100.00 [99.79]	100.00 [99.12]	100.00 [99.60]	100.00 [99.39]	100.00 [99.57]	94.55 [83.00]	80.41 [83.39]	80.31 [83.72]
MAG ACK-C130 (WVXH01000001)	3.19	36.8	81.4	1	99.16 [92.11]	98.26 [77.68]	98.20 [80.17]	98.68 [80.14]	97.86 [77.62]	98.27 [77.97]	98.42 [78.01]	98.66 [90.13]	100.00 [98.48]	*	100.00 [98.53]	100.00 [99.55]	100.00 [98.74]	99.98 [98.89]	99.98 [99.45]	100.00 [100]	94.51 [83.07]	80.41 [83.39]	80.30 [83.72]
MAG CS1-C145 (WVYD01000001)	3.18	36.8	81.4	1	99.16 [92.34]	98.26 [77.68]	98.20 [80.17]	98.68 [80.14]	97.86 [77.66]	98.27 [77.97]	98.43 [78.04]	98.65 [90.36]	100.00 [98.31]	99.99 [97.86]	*	100.00 [98.95]	100.00 [98.53]	100.00 [98.71]	100.00 [99.66]	100.00 [98.53]	94.53 [83.10]	80.41 [83.39]	80.30 [83.79]
MAG CS3-C139 (WVZD01000001)	3.18	36.9	81.4	1	99.16 [92.52]	98.26 [77.68]	98.20 [80.17]	98.68 [80.14]	97.86 [77.66]	98.27 [77.97]	98.43 [78.04]	98.65 [90.53]	100.00 [98.60]	99.99 [98.45]	100.00 [97.92]	*	100.00 [98.83]	100.00 [99.48]	100.00 [99.54]	100.00 [99.55]	94.53 [83.07]	80.41 [83.39]	80.30 [83.79]
MAG CS4-C041 (WVZT01000001)	3.19	36.9	81.4	1	99.16 [92.17]	98.26 [77.68]	98.20 [80.17]	98.68 [80.14]	97.86 [77.62]	98.27 [77.97]	98.42 [78.01]	98.66 [90.18]	100.00 [98.28]	100.00 [97.67]	99.99 [97.79]	99.99 [98.05]	*	99.98 [99.31]	99.98 [99.36]	100.00 [98.74]	94.50 [83.10]	80.41 [83.39]	80.28 [83.75]
MAG KS1-C152 (WWAN01000001)	3.2	36.9	81.4	1	99.16 [91.87]	98.26 [77.68]	98.20 [80.17]	98.68 [80.14]	97.86 [77.66]	98.27 [77.97]	98.43 [78.04]	98.65 [89.90]	100.00 [98.48]	100.00 [97.94]	100.00 [97.92]	100.00 [98.37]	100.00 [98.47]	*	100.00 [99.66]	99.98 [98.89]	94.49 [83.16]	80.41 [83.39]	80.27 [83.82]
MAG KS2-C139 (WWBA01000001)	3.16	36.9	81.4	1	99.16 [92.99]	98.26 [77.68]	98.20 [80.17]	98.68 [80.14]	97.86 [77.66]	98.27 [77.97]	98.43 [78.04]	98.65 [90.99]	99.99 [98.42]	99.99 [98.45]	100.00 [97.83]	100.00 [98.01]	99.99 [98.41]	100.00 [98.51]	*	99.98 [99.45]	94.53 [83.10]	80.41 [83.39]	80.30 [83.79]
MAG KS4-C130 (WWBX01000001)	3.19	36.8	81.4	1	99.16 [92.11]	98.26 [77.68]	98.20 [80.17]	98.68 [80.14]	97.86 [77.62]	98.27 [77.97]	98.42 [78.01]	98.66 [90.13]	100.00 [98.48]	100.00 [98.53]	100.00 [97.63]	100.00 [98.03]	100.00 [97.67]	100.00 [97.94]	99.99 [98.45]	*	94.51 [83.07]	80.41 [83.39]	80.30 [83.72]
H. flava CCUG 51397^T (LUHG01000001)	3.41	37.2	97.1	0	93.60 [85.96]	94.97 [89.63]	93.78 [85.97]	94.29 [84.81]	94.09 [82.42]	93.33 [78.63]	94.00 [82.56]	93.93 [88.48]	94.23 [84.25]	94.23 [83.64]	94.23 [83.85]	94.23 [84.01]	94.03 [77.10]	94.23 [83.43]	94.23 [84.44]	94.23 [83.64]	*	80.43 [84.56]	80.27 [86.15]
MAG SpSt-841 (DTJW01000000)	3.15	35.5	97.1	2	76.70 [76.50]	76.47 [70.61]	77.22 [72.28]	76.51 [66.74]	77.00 [67.03]	76.79 [61.63]	76.68 [65.58]	76.73 [76.81]	76.62 [67.92]	76.62 [67.92]	76.59 [68.67]	76.59 [68.80]	76.62 [67.92]	76.62 [67.92]	76.62 [67.92]	76.62 [67.92]	76.44 [75.77]	*	96.56 [90.12]
H. sandarakina DSM 23241^T (QKZV01000001)	3.27	35.7	97.1	0	76.83 [75.14]	76.84 [71.74]	77.59 [74.76]	76.63 [68.27]	77.11 [68.85]	76.96 [64.65]	76.95 [67.19]	76.84 [75.29]	76.69 [70.53]	76.70 [70.02]	76.70 [70.20]	76.70 [70.34]	76.63 [66.54]	76.70 [69.85]	76.70 [70.69]	76.70 [70.02]	76.54 [74.66]	94.18 [91.67]	*

heterogeneous phenotypic repertoire. The difference in temperature range for growth, motility, enzymatic activity or carbon-source assimilation and acid-formation tests suggested some differential features between strain TMF_100^T and its closest relative species *H. flava* CCUG 51397^T.

Genome characterization and comparative analysis

The genome size of TMF_100^T was 3.46 Mbp with a G + C content of 37% (Table 1), which was well in accordance with the other type strains of the genus *Hydrotalea. H. sandarakina* DSM 23241^T presented the smallest genome (3.27 Mbp) and the lowest G + C content (35.7%). Genome pairwise comparisons between type strains of all *Hydrotalea* species showed that the total core-genome was comprised of 2,304 orthology groups (OGs). From these, a total of 445, 427 and 405 were specific for TMF_100^T, *H. flava* CCUG 51397^T and *H. sandarakina* DSM 23241^T, respectively (Supplementary Fig. S5A). After analyzing the specific proteins of strain TMF_100^T, the presence was found of bacteriocin, components involved in membrane transport, heavy metal resistance or CRISPR-associated proteins, among other functions (Supplementary Spreadsheet S4). According to the KEGG annotation, strain TMF_100^T encoded for the main pathways involved in central carbohydrate metabolism, such as glycolysis (Embden-Meyerhof pathway), the TCA cycle (Krebs cycle) or pentose phosphate pathways. Moreover, genes involved in gliding motility were identified and were mainly represented by Gld and Spr proteins [3], in accordance with the positive motility test performed in the laboratory. Conspicuously, *H. flava* CCUG 51397^T encoded for gliding genes, while only a few of them were found in *H. sandarakina* AF-51^T and, despite their taxonomic characterizations, they were considered negative for the motility tests. However, the motility of these strains may need to be stimulated under laboratory conditions in order to reactivate the apparent lack of motility [42].

From the different sequence genomes reconstructed from metagenomic samples obtained in distinct years (2012, 2014 and 2017), as well as environments (acid mine, marine sediment and hot spring) and location (USA, Brazil and India) (Fig. 1), those with the best quality for the analyses of the gene content were used. The G + C content of the MAGs ranged from 35.5% to 37.8% with a genome size between 3.15 Mbp and 3.93 Mbp, whose completeness and contamination were ≥ 81.4% and ≤ 2%, respectively (Table 1). On the other hand, the six SAGs from acid mine drainage in Brazil (recovered at the same location as the MAGs but one month later), showed a G + C content in accordance with the rest of the genomes but with a greater variability in the genome size due to their incompleteness that ranged between 0.7 Mbp to 2.08 Mbp (Table 1). The genomic pairwise comparison between all of them, including strain TMF_100^T and the closest relative *H. flava* CCUG 51397^T, showed a core-genome constituted by 2,369 OGs (Supplementary Fig. S5B).

Interestingly, the strains shared gene clusters encoding for heavy metal resistance as part of their core genome. This was especially true for copper resistance determinants that were described for all SAGs and the MAG AMD that had been obtained from acid mine drainage in Brazil, and they were hypothesized to act as a protective mechanism against the high concentration of this element that could induce damage of the biological system [25,33]. The Cu resistance genes, and others also encoding for proteins conferring resistance to heavy metals, which act by pumping them out of the cytoplasm or periplasm, were present in MAGs originating from marine sediments in the Golf of Khambhat and also in our strain TMF_100^T obtained from subsurface freshwater (Supplementary Spreadsheet S5). The CzcCBA efflux protein complex or copper translocating P-type ATPases could be detected in most of the genomes together with other components, such as multicopper

Table 3
Diagnostic phenotypic characteristics of *H. lipotrueae* TMF_100^T with the closely related *Hydrotalea* species.

Characteristics	*H. lipotrueae* TMF_100^T	*H. flava* CCUG 51397^T	*H. sandarakina* (AF-51^T/AF-50)
Isolation source	Groundwater	Industry distilled water	Hot spring water
Pigmentation	Orange	Orange-yellow	Orange
Growth temperature range (°C)	15–45	20 to 37	25–52.5
Optimum growth temperature (°C)	30	25 to 30	45
Motility	+	-	-
Hydrolysis of:			
Gelatin	-	-	+
Enzyme activity:			
β-glucuronidase	-	+	+
α-fucosidase	-	+	+
Assimilation of:			
Gluconate	+	(-)	+
D-cellobiose	-	(+)	+
D-mannose	-	(+)	+
Glycerol	-	(-)	+
L-rhamnose	-	(+)	-
D-xylose	-	(+)	+
DL-lactate	-	(-)	+
L-ornithine	-	(-)	+
L-aspartate	-	(+)	+
L-proline	-	(-)	+
Acid production from:			
D-ribose	-	(-)	+
D-galactose	+	(+)	-
Glycogen	-	(+)	+
Potassium 5-ketogluconate	-	(-)	+
Major fatty acids	iso-$C_{15:0}$, iso-$C_{17:0}$ 3-OH	iso-$C_{15:0}$, iso-$C_{17:0}$ 3-OH	iso-$C_{15:0}$, iso-$C_{17:0}$ 3-OH, iso-$C_{16:0}$
DNA G + C content (%)	37	37	35.2–35.3
ANI/AAI *H. flava*	93.60 / 93.94	100 / 100	76.54 / 80.27
ANI/AAI *H. sandarakina*	76.83 / 80.65	76.54 / 80.27	100 / 100
Reference			[1]
Reference ()		[1]	
Reference [ ]		[17]	

In the API ZYM test, all the strains were positive for alkaline phosphatase, esterase (C4), esterase lipase (C8), leucine arylamidase, valine arylamidase, cysteine arylamidase, trypsin, α-chymotrypsin, acid phosphatase, naphthol-AS-BI-phosphohydrolase, α-galactosidase, β-galactosidase, α-glucosidase, n-acetyl-β-glucosaminidase and α-mannosidase; but negative for lipase (C14). All strains were positive for casein, starch and aesculin.

oxidase or protein CutC involved in copper homeostasis [20]. Apart from Cu-resistance genes, components related to arsenic resistance (i.e. arsenate reductase or Acr3 transporter; [13]) or mercury (i.e. mercury ion reductase or mercuric transport protein; [8]) were also detected, confirming that these microorganisms have developed molecular detoxification mechanisms to survive heavy metal-rich environments.

One of the hypotheses proposed by [33] is for the key role of *Hydrotalea* species in the denitrification process and therefore their potential implication in the regulation of pH in acid mine environments. In the current study, acidity was not a characteristic of the environments from where TMF_100^T was isolated (Supplementary Spreadsheet S1), despite the fact that the biogeochemical nitrogen cycle has an important relevance in aquatic ecosystems [52]. Denitrification and dissimilatory nitrate reduction (DNRA) are key processes involved in nitrate reduction. In both cases, the first step consists of the conversion of nitrate to nitrite catalysed by nitrate reductase, such as NarGHI or NapAB. In the present case, these enzymes were not detected in strain TMF_100^T in accordance with a negative result for nitrate reduction in the phenotypic test. On the other hand, the whole-genome analysis indicated that this new strain encoded a nitrite reductase (NirBD) from DNRA and other enzymes involved in the denitrification process (Supplementary Fig. S6). Regarding the latter, four enzymes were implied in the complete denitrification process, where nitrate and nitrite are reduced to molecular nitrogen: nitrate reductase (Nar), nitrite reductase (Nir), nitric oxide reductase (Nor) and nitrous oxide reductase (Nos) [37]. However, nitrite reductase (NirK or NirS) and nitrous oxide reductase NosZ constitute the main key enzymes for the ecology of denitrifying bacteria [52]. The gene *nosZ* was detected in almost all the genomes, including strain TMF_100^T, MAGs (Brazil, Gulf of Khambhat and Yellowstone) and the type strains *Hydrotalea flava* CCUG 51397^T and *Hydrotalea sandarakina* DSM 23241^T. Due to the low completeness of the SAGs, this enzyme could only be detected in SAG B16. Moreover, in these genomes, nitric oxide reductase NorBC was also detected that, together with nitrous oxide reductase NosZ, was implied in the last steps of the denitrification process that transforms nitric oxide to nitrogen. The nitrite reductase gene, encoding for the NirK-like enzyme, was detected in most of the genomes; in particular, a copper-containing nitrite reductase (EC 1.7.2.1). According to [28], this enzyme could be considered as a NirK, implied in the reduction of nitrite to nitric oxide, which has been detected in prokaryotic and fungal communities. The gene encoding for this enzyme was present in strain TMF_100^T, the MAGs from Brazil and the Gulf of Khambhat, the type strain *Hydrotalea flava* CCUG 51397^T and some SAGs, such as A04, J04 and P17. Based on these observations, it seems that this bacterial taxon could play an important role in the denitrification process, in accordance with [33], especially those genomes affiliated to the new species proposed in this study and also the type strain *Hydrotalea flava* CCUG 51397^T.

The presence of CRISPR-Cas systems was also investigated due to their special relevance, since this mechanism was emphasized in the description for the SAGs isolated from acid mine drainage in Brazil [33]. CRISPR-Cas systems are constituted by a CRISPR array, encoding individual spacers separated by repeats, and CRISPR-associated proteins [50]. The latter, Cas proteins, are highly diverse and thus CRISPR-Cas systems are classified into 2 classes, 6 types and 33 subtypes [32]. Accordingly, all genomes involved in

Table 4
Cellular fatty acids of strains TMF_100^T, *Hydrotalea flava* CCUG 51397^T and *Hydrotalea sandarakina* AF-51^T grown on the same medium (R2A) and at the same temperature (28 °C). Profiles indicated with + symbols were determined in this study. Bold type letters correspond to the most relevant traits determined.

Fatty acids	TMF_100^T	*Hydrotalea flava* CCUG 51397^T	*Hydrotalea sandarakina* AF-51^T
Saturated			
$C_{9:0}$	0.1	-	-
$C_{10:0}$	0.3	-	-
$C_{15:0}$	3.1	-	-
$C_{16:0}$	1.5	3.1 ± 0.2	2.4 ± 0.1
Unsaturated			
$C_{17:1}\omega6c$ [i]	-	tr	1.6 ± 0.1
iso $C_{17:1}\omega9c$	4.9	-	-
Branched chain			
iso-$C_{14:0i}$	-	0.7 ± 0.1	1.8 ± 0.1
iso-$C_{13:0}$	3.8	3.8 ± 0.3	2.0 ± 0.1
iso-$C_{15:1}$ G	6.5	2.5 ± 0.3	-
iso-$C_{15:0}$	**30.9**	**24.9 ± 0.5**	**26.2 ± 0.2**
iso-$C_{16:0}$	5.3	4.3 ± 0.4	**8.4 ± 0.4**
iso-$C_{17:0}$	1.7	5.6 ± 0.4	3.1 ± 0.2
iso-$C_{15:0}$ 3-OH	2.9	3.8 ± 0.4	2.8 ± 0.1
iso-$C_{16:0}$ 3-OH	5.1	5.3 ± 0.5	4.2 ± 0.2
iso-$C_{16:1}$H	1.4	0.7 ± 0.1	1.7 ± 0.1
iso-$C_{17:0}$ 3-OH	**18.0**	**20.9 ± 1.0**	**10.5 ± 0.4**
anteiso-$C_{15:0}$	4.3	-	-
Hydroxy			
$C_{15:0}$ 2-OH [i]	-	0.6 ± 0.1	0.6 ± 0.1
$C_{16:0}$ 3-OH	1.4	2.5 ± 0.3	1.5 ± 0.1
$C_{17:0}$ 2-OH	2.0	1.7 ± 0.4	1.3 ± 0.1
$C_{17:0}$ 3-OH [i]	-	0.6 ± 0.1	tr
Summed features*			
3	3.0	3.2 ± 0.3	6.9 ± 0.2
4	-	-	0.8 ± 0.2
9	-	3.0 ± 0.2	6.6 ± 0.3
Unknown 12.560 [i]	-	1.1 ± 0.2	1.6 ± 0.1
Unknown 13.565	2.1	3.0 ± 0.4	4.9 ± 0.3
Unknown 16.582	1.4	1.3 ± 0.2	0.7 ± 0.1
Reference	+	[1]	[1]

[i]Not named in the database used in this study.
Tr: trace
- : not detected
*Summed feature 3 comprised $C_{16:1}\omega7c$ and/or iso-$C_{15:0}$ 2-OH in the database used in this study, and $C_{16:1}\omega7c$ and $C_{16:1}\omega6c$ in Albuquerque et al. [1]; summed feature 4 comprised iso-$C_{17:1}$ I and/or anteiso-$C_{17:1}$B; summed feature 9 comprised iso-$C_{17:1}\omega9c$ and/or 10-methyl $C_{16:0}$

this current study, except MAG SpSt-841 (DTJW01000000), encoded for a complete or partial CRISPR-Cas system. The strain TMF_100^T presented two types of CRISPR-Cas systems. One of them, placed in contig 9, was clearly a homologue of type II-C (Fig. 4A); while contig 20 harbored the second type of CRISPR-Cas that apparently corresponded to type I-B (Fig. 4B). However, the latter did not present the same organization of module genes as described by [32] for this system. The canonical type I-B operon comprises, according to Makarova et al. [32], a specific set of Cas proteins after the CRISPR-array, which is constituted by Cas 2, Cas1, Cas 4 and Cas 3, followed by the effector complex (Cas 5, Cas 7 and Cas 8) and Cas 6. Within the CRISPR-Cas system, the protein effector complex carries out the recognition and destruction of invader nucleic acid [50]. In the case of strain TMF_100^T, the operon structure was identical but different in the effector complex, since the order of the module genes was Cas 7, Cas 8 and Cas 5. Moreover, in a further analysis using the CRISPRCas Finder tool [10], the Cas proteins that constituted this complex in TMF_100^T belonged to type I-C, although, despite this fact, the remaining Cas proteins presented the same operon structure as type I-B. To our knowledge, this could be a new type only found to date in strain TMF_100^T.

For the rest of the genomes, the complete set of *cas* genes encoding for type II-C was detected in the eight MAGs from the Gulf of Khambhat (A-C108, ACK-C130, CS1-C145, CS3-C139, CS4-C041, KS1-C152, KS2-C139 and KS4-C130), as well as for MAG AMD isolated from an acid mine sample in Brazil. The latter was also characterized by a disrupted synteny, since the Cas-system was distributed in two different contigs (Fig. 4A). SAGs B16, J04, L11 and P17 encoded for the type II-C system in accordance with the results published by [33]. On the other hand, a canonical type I-B operon was present in MAG AMD and was partially complete in SAGs L11, K06, P17, J04 and B16. This CRISPR-Cas system could not be identified for MAGs from the Gulf of Khambhat, although the repeat element in the CRISPR array was identical to those genomes for which type I-B was present. In this case, the CRISPR array was located on the edge of the contig and the adjacent genes could not be identified. Altogether, at least seven consensus repeat sequences could be detected and each organism exhibited between one to sixty-three different spacer regions (Supplementary Spreadsheet S6). Interestingly, most of the repeat sequences were shared between them, highlighting those present in strain TMF_100^T that were also detected in MAGs and/or SAG B16 and *H. flava* CCUG 51397^T. Moreover, CRISPR spacers were different and unique for each genome with the exception of SAGs B16, L11 and P17, as described by [33]. Taking into account that a segment of invading DNA is incorporated as a spacer within the CRISPR system [45], it could be hypothesized that these organisms had been infected by different viruses and, therefore, different viral populations would have inhabited the original environment from where the samples were collected.

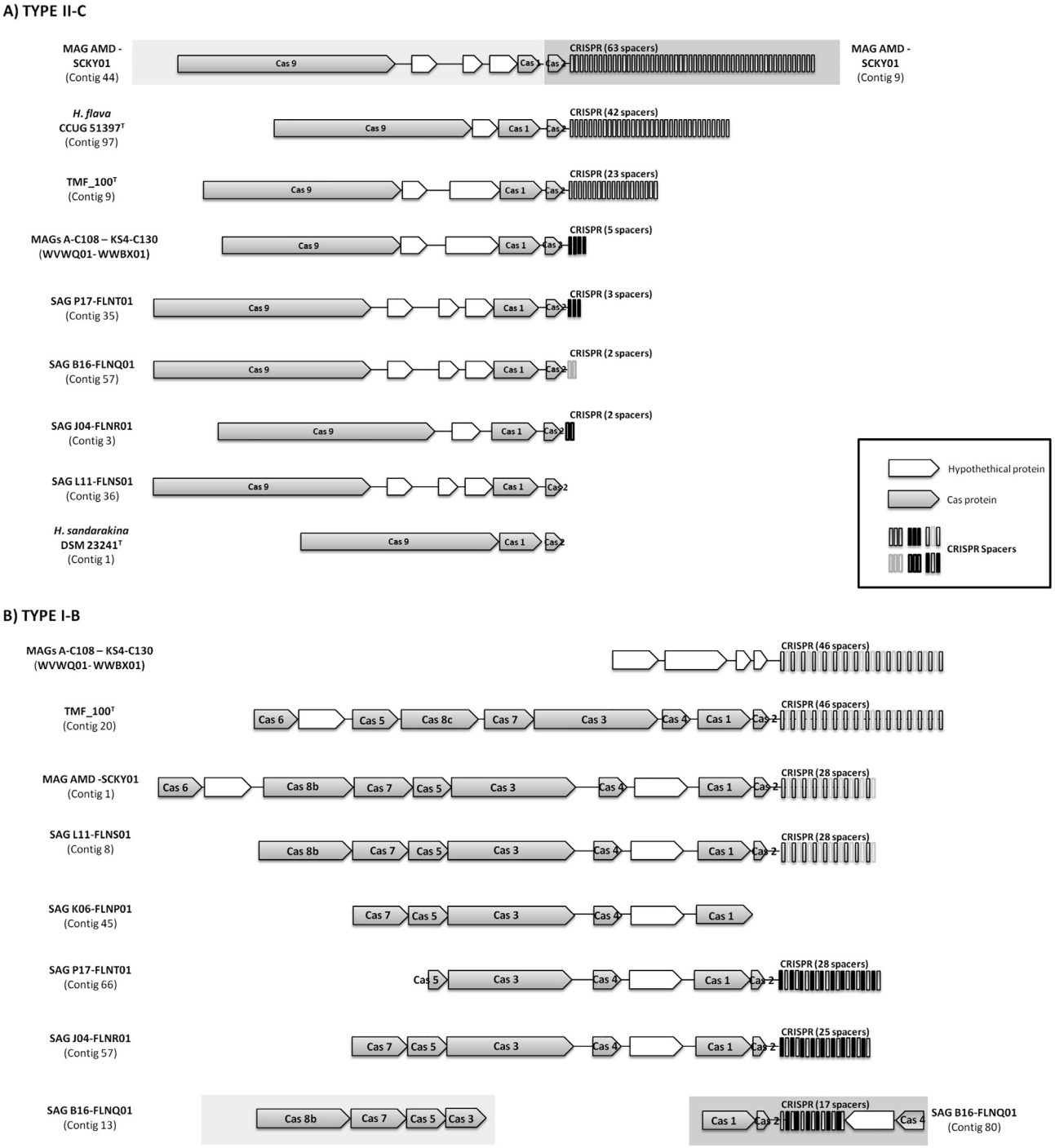

Fig. 4. CRISPR-Cas systems detected in the new isolate *H. lipotrueae* TMF_100^T and their comparison with the *Hydrotalea* genomes included in this study. A: Type II-C; B: Type I-B.

Ecological relevance

The metagenomes of the six collected subsurface samples were from where the isolates were used to recruit the reference genomes of *H. flava* CCUG 51397^T, *H. sandarakina* DSM 23241^T and strain TMF_100^T. The species *H. flava* CCUG 51397^T and strain TMF_100^T were present only in samples T2 and T6 (Fig. 5), and TMF_100^T was clearly the most abundant, especially in its sample of origin (T2). No sequences from the subsurface metagenomes were recruited using the *H. sandarakina* DSM 23241^T genome or MAG SpSt-841 as references, suggesting that these *Hydrotalea* species were not present in these samples. The relative genome abundance values ranged between 0.01% to 0.015% for TMF_100^T and 0.006% to 0.009% for *H. flava* CCUG 51397^T (Supplementary Spreadsheet S7). The recruitment plots of strain TMF_100^T indicated a low intra-specific diversity because the majority of reads mapped at a 100% identity. However, the recruitment in T6 indicated the presence of a different strain, as shown by the absence of coverage for some genomic regions (genomic islands). These genomic islands encoded for 430 CDS, of which 201 belonged to strain-specific genes (Supplementary Spreadsheet S8). The annotation of these genes showed components related to mobile genetic elements,

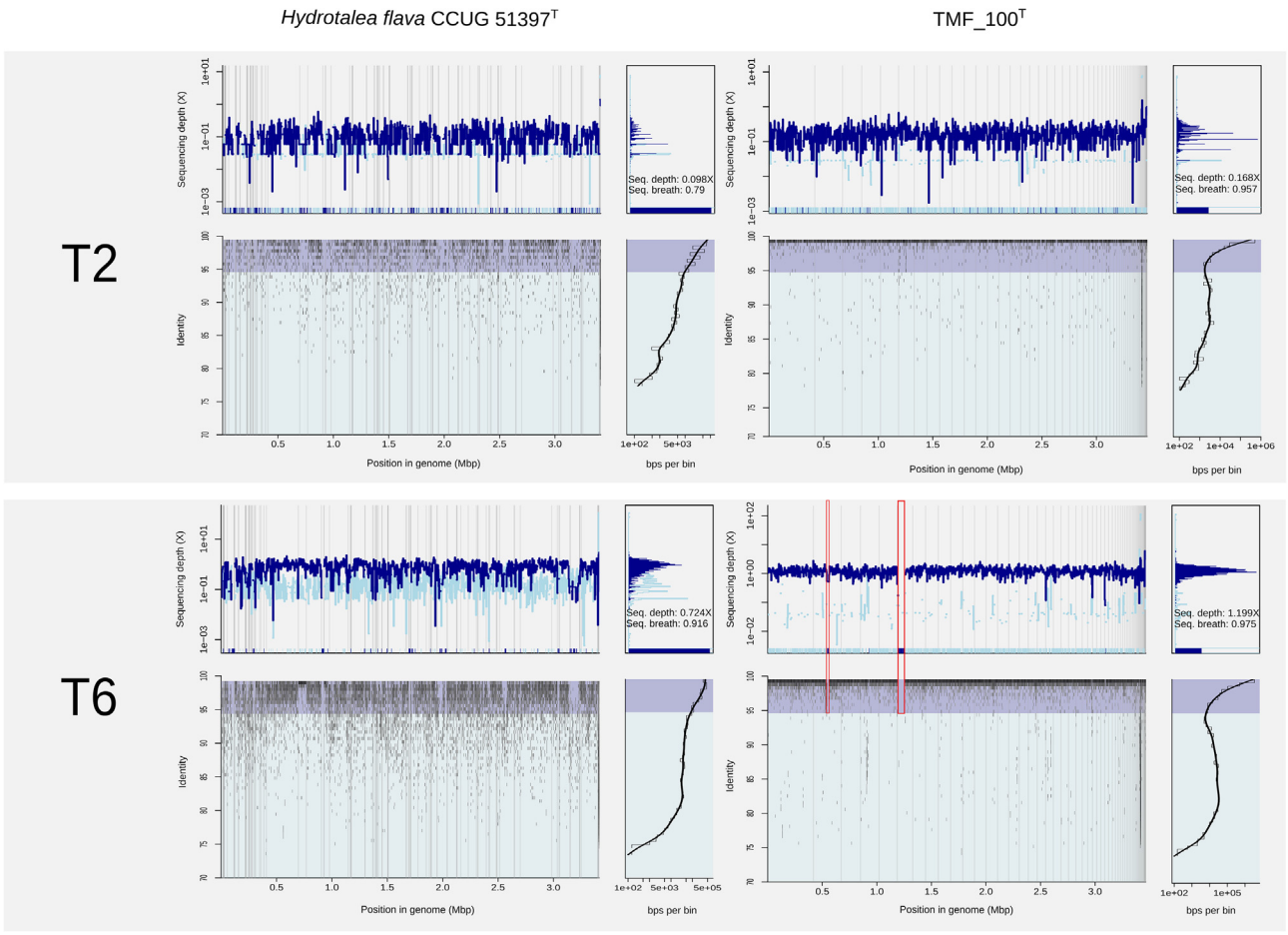

Fig. 5. Recruitment plot of *Hydrotalea flava* CCUG 51397^T (left) and strain TMF_100^T (right) against metagenomes T2 and T6. The sequencing depth or coverage at the top of each panel was estimated based on reads mapping at > 95% sequence identity. The boxes on the sequencing depth chart indicate genomic islands where no reads mapped to the contig.

such as conjugative transposases. Conspicuously, these hypervariable regions also encoded for transcriptional regulators involved in arsenic and mercury resistance (i.e. ArsR and MerR family proteins), the AraC family protein involved in carbon-source metabolism, pathogenesis or stress responses [9,15,27], bacteriocins with antimicrobial activity against other bacteria, and CRISPR-associated proteins. We could hypothesize that these genomic islands, which encode for functions that can be advantageous for the host, may be horizontally transferable due to the presence of conjugative transposases [5].

The metagenomic samples from where MAGs and SAGs were isolated (Brazil, Gulf of Khambhat and Yellowstone; Fig. 1) were also used for recruitment purposes, and *H. flava* CCUG 51397^T and strain TMF_100^T were present in metagenomes from acid mine drainage in Brazil and Gulf of Khambhat, but not in Yellowstone. Conversely, only *H. sandarakina* DSM 23241^T and MAG SpSt-841 (Supplementary Fig. S7) were detected in Yellowstone. As expected, MAG SpSt-841 was the most abundant with a relative abundance of 0.03% (Supplementary Spreadsheet S9). Interestingly, the TMF_100^T relative abundance ranged from 0.6% to 2.01% in the acid mine drainage from Brazil; whereas *H. flava* CCUG 51397^T values ranged from 0.4% to 1.29% (Supplementary Spreadsheet S9). Both genomes were also detected, but at lower abundances in the metagenomes from the Gulf of Khambhat, and only in sample GOCS4. In this case, the relative abundance was 0.074% and 0.046%, respectively. The recruitment plot of TMF_100^T in GOCS4 showed that most of the reads mapped at 100% identity. According to this, we could assume that the same population was present in both locations (Brazil and Spain) as part of the resident microbial community, but with a low intra-specific diversity. This study highlighted the ecological relevance of members of the genus *Hydrotalea* especially due to their global distribution and environmental versatility characterized by the ability to inhabit ecosystems with different temperatures, pH levels or salinities; and they could also be involved in symbiotic relationships, as has already been hypothesized [25]. This genus seems to have relevant features, such as denitrification and heavy metal resistance, as well as CRISPR-Cas systems, relevant for their survival in variable ecosystems.

Conclusions

Strain TMF_100^T represents the first isolate of a widely distributed species of the genus *Hydrotalea* that shows high relevance in its environments of origin due to its relatively high abundance [25,33,36] and physiological versatility. This is supported by the fact that they inhabit environments as distinct as oligotrophic environments, industry distilled waters and subsurface groundwaters, as well as acid mine drainage and marine sediments. The taxonomic results based on phylogenetic, genomic, phenotypic and ecological analyses, clearly indicated that isolated strain TMF_100^T represented a new species within the genus *Hydrotalea*, for which we propose *Hydrotalea lipotrueae* sp. nov., named after the biotechnology company Lipotrue, S.L. The protologue describing the diagnostic traits of the new species is given in Table 5.

Table 5
Protologue of the new species description.

Species name	*Hydrotalea lipotrueae*
Specific epithet	*lipotrueae*
Species status	sp. nov.
Species etymology	(li.po.true'ae. N.L. gen. n. *lipotrueae* of the Lipotrue company
Description of the new taxon and diagnostic traits	Rod-shaped cells, 1.0–5.0 μm long and 0.3–0.5 μm wide, forming orange colonies after 3 days growth on R2A agar at 30 °C, shiny with regular shape, flat elevation, smooth surface and with a diameter of 0.5–1.0 mm. Cells are motile by gliding and exhibit growth in the ranges of 15–45 °C, up to 1% NaCl and pH 4–9. The organism is positive for starch, aesculin hydrolysis and Tween 20 but negative for catalase, oxidase, casein and gelatin hydrolysis. Nitrate is not reduced. Assimilation of D-fructose, D-galactose, gluconate, D-glucose, D-maltose and D-saccharose. Acid formation from D-xylose, D-galactose, D-glucose, D-fructose, D-mannose, L-rhamnose, methyl-αD-mannopyranoside, methyl-αD-glucopyranoside, N-acetylglucosamine, amygdalin, arbutin, aesculin ferric citrate, salicin, D-cellobiose, D-maltose, D-lactose, D-melibiose, D-saccharose, trehalose, melezitose, raffinose, starch, gentiobiose, and turanose. The major fatty acids were iso-$C_{15:0}$ and iso-$C_{17:0}$ 3-OH.
Country of origin	Spain
Region of origin	Mallorca Island
Date of isolation (dd/mm/yyyy)	11/12/2016
Source of isolation	Groundwater
Sampling date (dd/mm/yyyy)	14/11/2016
Latitude (xx°xx'xx" N/S)	39°28′44.1″N
Longitude (xx°xx'xx" E/W)	2°53′09.6″E
Altitude (meters above sea level)	Groundwater
16S rRNA gene accession nr.	-
Genome accession number [RefSeq; EMBL; ...]	JAESDL000000000
Genome status	Draft – 97.1% completeness
Genome size	3.46
G + C %	37
Number of strains in study	1
Source of isolation of non-type strain	Groundwater
Information related to the Nagoya Protocol	Not applicable
Designation of the type strain	TMF_100^T
Strain collection numbers	CECT 30288^T = CCUG 75179^T

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors want to acknowledge Aharon Oren for supervising the taxonomic etymologies of the new species and Roth E. Conrad for the useful pipeline for annotating amino acid sequences. This study was funded by funding investment from the cosmetic enterprise Lipotrue S.L., and by the Spanish Ministry of Science, Innovation and Universities projects CLG2015_66686-C3-1-P, PGC2018-096956-B-C41 and RTC-2017-6405-1 (to RRM) and was also supported with European Regional Development Fund (FEDER) funds.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.syapm.2021.126277.

References

[1] Albuquerque, L., Rainey, F.A., Nobre, M.F., da Costa, M.S. (2012) *Hydrotalea sandarakina* sp. nov., isolated from a hot spring runoff, and emended descriptions of the genus *Hydrotalea* and the species *Hydrotalea flava*. Int. J. Syst. Evol. Microbiol. 62, 1603–1608.
[2] Alneberg, J., Karlsson, C.M.G., Divne, A.M., Bergin, C., Homa, F., Lindh, M.V., Hugerth, L.W., Ettema, T.J.G., Bertilsson, S., Andersson, A.F., Pinhassi, J. (2018) Genomes from uncultivated prokaryotes: a comparison of metagenome-assembled and single-amplified genomes. Microbiome. 6, 173.
[3] Alejandre-Colomo, C., Viver, T., Urdiain, M., Francis, B., Harder, J., Kämpfer, P., Amann, R., Rosselló-Móra, R. (2020) Taxonomic study of nine new *Winogradskyella* species occurring in the shallow waters of Helgoland Roads, North Sea. Proposal of *Winogradskyella schleiferi* sp. nov., *Winogradskyella costae* sp. nov., *Winogradskyella helgolandensis* sp. nov., *Winogradskyella vidalii* sp. nov., *Winogradskyella forsetii* sp. nov., *Winogradskyella ludwigii* sp. nov., *Winogradskyella ursingii* sp. nov., *Winogradskyella wichelsiae* sp. nov., and *Candidatus* "Winogradskyella atlantica" sp. nov. Syst. Appl. Microbiol. 43, 126128.
[4] Aziz, R.K., Bartels, D., Best, A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., Zagnitko, O. (2008) TheRAST Server: rapid annotations using subsystems technology. BMC Genomics. 9, 1–15.
[5] Bellanger, X., Payot, S., Leblond-Bourget, N., Guédon, G. (2014) Conjugative and mobilizable genomic islands in bacteria: evolution and diversity. FEMS Microbiol. Rev. 38, 720–760.
[6] Bernardet, J.F., Nakagawa, Y., Holmes, B. (2002) Subcommittee on the taxonomy of Flavobacterium and Cytophaga-like bacteria of the International Committee on Systematics of Prokaryotes Proposed minimal standards for describing new taxa of the family Flavobacteriaceae and emended description of the family. Int. J. Syst. Evol. Microbiol. 52, 1049–1070.
[7] Bowman, J.P. (2000) Description of *Cellulophaga algicola* sp. nov., isolated from the surfaces of Antarctic algae, and reclassification of *Cytophaga uliginosa* (ZoBell and Upham 1944) Reichenbach 1989 as *Cellulophaga uliginosa* comb. nov. Int. J. Syst. Evol. Microbiol. 50, 1861–1868.
[8] Boyd, E.S., Barkay, T. (2012) The mercury resistance operon: from an origin in a geothermal environment to an efficient detoxification machine. Front. Microbiol. 3, 349.
[9] Brown, N.L., Stoyanov, J.V., Kidd, S.P., Hobman, J.L. (2003) The MerR family of transcriptional regulators. FEMS Microbiol. Rev. 27, 145–163.
[10] Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Néron B, Rocha EPC, Vergnaud G, Gautheret D, Pourcel C. (2018) CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Res. 46, 246-251.
[11] Cowan, S.T., Steel, K.J. Manual for the Identification of Medical Bacteria. Cambridge University Press, London.
[12] Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32 (5), 1792–1797.
[13] Firrincieli, A., Presentato, A., Favoino, G., Marabottini, R., Allevato, E., Stazi, S.R., Scarascia Mugnozza, G., Harfouche, A., Petruccioli, M., Turner, R.J., Zannoni, D., Cappelletti, M. (2019) Identification of resistance genes and response to arsenic in *Rhodococcus aetherivorans* BCP1. Front. Microbiol. 10, 888.

[14] Gago, J.F., Viver, T., Urdiain, M., Pastor, S., Kämpfer, P., Ferreira, E., Rosselló-Móra, R. (2021) Description of the three new *Alteromonas* species *Alteromonas antoniana* sp. nov., *Alteromonas lipotrueae* sp. nov. and *Alteromonas lipotrueiana* sp. nov. isolated from marine environments, and proposal of reclassification of the genus *Salinimonas* as *Alteromonas*. Syst. Appl. Microbiol. 44, 126226.

[15] Ibarra, J.A., Pérez-Rueda, E., Segovia, L., Puente, J.L. (2008) The DNA-binding domain as a functional indicator: the case of the AraC/XylS family of transcription factors. Genetica. 133 (1), 65–76.

[16] Kämpfer, P., Kroppenstedt, R.M. (1996) Numerical analysis of fatty acid patterns of coryneform bacteria and related taxa. Can. J. Microbiol. 42, 989–1005.

[17] Kämpfer, P., Lodders, N., Falsen, E. (2011) *Hydrotalea flava* gen. nov., sp. nov., a new member of the phylum *Bacteroidetes* and allocation of the genera *Chitinophaga*, *Sediminibacterium*, *Lacibacter*, *Flavihumibacter*, *Flavisolibacter*, *Niabella*, *Niastella*, *Segetibacter*, *Parasegetibacter*, *Terrimonas*, *Ferruginibacter*, *Filimonas* and *Hydrotalea* to the family *Chitinophagaceae* fam. nov. Int. J. Syst. Evol. Microbiol. 61, 518–523.

[18] Kämpfer, P., Steiof, M., Dott, W. (1991) Microbiological characterization of a fuel-oil contaminated site including numerical identification of heterotrophic water and soil bacteria. Microb. Ecol. 21, 227–251.

[19] Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., Tanabe, M. (2016) KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 44 (D1), D457–D462.

[20] Kim, E.-H., Nies, D.H., McEvoy, M.M., Rensing, C. (2011) Switch or funnel: how RND-type transport systems control periplasmic metal homeostasis. J. Bacteriol. 193 (10), 2381–2387.

[21] Konstantinidis, K.T., Tiedje, J.M. (2005) Towards a genome-based taxonomy for prokaryotes. J. Bacteriol. 187 (18), 6258–6264.

[22] Lanave, C., Preparata, G., Saccone, C., Serio, G. (1984) A new method for calculating evolutionary substitution rates. J. Mol. Evol. 20, 86–93.

[23] Leandro, T., França, L., Nobre, M.F., Rainey, F.A., da Costa, M.S. (2013) *Heliimonas saccharivorans* gen. nov., sp. nov., a member of the family *Chitinophagaceae* isolated from a mineral water aquifer, and emended description of *Filimonas lacunae*. Int. J. Syst. Evol. Microbiol. 63, 3793–3799.

[24] Leite, L.R., Medeiros, J.D., Fernandes, G.R., Araújo, F., Pylro, V.S., Salim, A.C., Volpini, Â., Oliveira, G., Cuadros-Orellana, S. (2016) Draft genome sequence of *Hydrotalea flava* strain CCUG 51397T. Genome Announc. 4, e00527–16.

[25] Lemos, L.N., Medeiros, J.D., Dini-Andreote, F., Fernandes, G.R., Varani, A.M., Oliveira, G., Pylro, V.S. (2019) Genomic signatures and co-occurrence patterns of the ultra-small Saccharimonadia (phylum CPR/Patescibacteria) suggest a symbiotic lifestyle. Mol. Ecol. 28, 4259–4271.

[26] Li, B., Zhang, K., Li, C., Wang, X., Chen, Y., Yang, Y. (2019) Characterization and comparison of microbiota in the gastrointestinal tracts of the goat (*Capra hircus*) during preweaning development. Front. Microbiol. 10, 2125.

[27] Li, Q., Li, C., Xie, L., Zhang, C., Feng, Y., Xie, J. (2017) Characterization of a putative ArsR transcriptional regulator encoded by Rv2642 from *Mycobacterium tuberculosis*. J. Biomol. Struct. Dyn. 35, 2031–2039.

[28] Long, A., Song, B., Fridey, K., Silva, A. (2015) Detection and diversity of copper containing nitrite reductase genes (nirK) in prokaryotic and fungal communities of agricultural soils. FEMS Microbiol. Ecol. 91, 1–9.

[29] Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai, T., Steppi, S., Jobb, G., Forster, W., Brettske, I., Gerber, S., Ginhart, A.W., Gross, O., Grumann, S., Hermann, S., Jost, R., König, A., Liss, T., Lussmann,T., May, M., Nonhoff, B., Reichel, B., Strehlow, R., Stamatakis, A., Stuckmann, N.,Vildig, A., Lenke, M., Ludwig, T., Bode, A., Schleifer, K.H. (2004) ARB; a software environment for sequence data. Nucleic Acids Res. 32, 1363–1371.

[30] Ludwig, W., Viver, T., Westram, R., Gago, J.F., Bustos-Caparros, E., Knittel, K., Amann, R., Rosselló-Móra, R. (2021) Release LTP_12_2020, featuring a new ARB alignment and an improved 16S rRNA tree for prokaryotic type strains. Syst. Appl. Microbiol. 44, 126218.

[31] Machado, V.E., Martins, P.M., Ferreira, H., Ferro, M., Bacci, M., Pinto, M.C. (2014) Bacterial groups associated with *Nyssomyia neivai* (Diptera: Psychodidae) sandflies. J. Vector Borne Dis. 51, 137–139.

[32] Makarova, Kira S., Wolf, Yuri I., Iranzo, Jaime, Shmakov, Sergey A., Alkhnbashi, Omer S., Brouns, Stan J.J., Charpentier, Emmanuelle, Cheng, David, Haft, Daniel H., Horvath, Philippe, Moineau, Sylvain, Mojica, Francisco J.M., Scott, David, Shah, Shiraz A., Siksnys, Virginijus, Terns, Michael P., Venclovas, Česlovas, White, Malcolm F., Yakunin, Alexander F., Yan, Winston, Zhang, Feng, Garrett, Roger A., Backofen, Rolf, van der Oost, John, Barrangou, Rodolphe, Koonin, Eugene V. (2020) Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. Nat. Rev. Microbiol. 18 (2), 67–83.

[33] Medeiros, J.D., Leite, L.R., Pylro, V.S., Oliveira, F.S., Almeida, V.M., Fernandes, G. R., Salim, A.C.M., Araújo, F.M.G., Volpini, A.C., Oliveira, G., Cuadros-Orellana, S. (2017) Single-cell sequencing unveils the lifestyle and CRISPR-based population history of *Hydrotalea* sp. in acid mine drainage. Mol. Ecol. 26, 5541–5551.

[34] Mootapally, C., Nathani, N.M., Poriya, P., Beleem, I., Dabhi, J.C., Gadhvi, I.R., Joshi, C.G. (2019) Antibiotic resistome biomarkers associated to the pelagic sediments of the gulfs of Kathiawar Peninsula and Arabian Sea. Sci. Rep. 9, 17281.

[35] Mora-Ruiz, M.R., Font-Verdera, F., Díaz-Gil, C., Urdiain, M., Rodríguez-Valdecantos, G., González, B., Orfila, A., Rosselló-Móra, R. (2015) Moderate halophilic bacteria colonizing the phylloplane of halophytes of the subfamily *Salicornioideae (Amaranthaceae)*. Syst. Appl. Microbiol. 38, 406–416.

[36] Nathani, Neelam M., Mootapally, Chandrashekar, Dave, Bharti P. (2019) Antibiotic resistance genes allied to the pelagic sediment microbiome in the Gulf of Khambhat and Arabian Sea. Sci. Total Environ. 653, 446–454.

[37] Olaya-Abril, A., Hidalgo-Carrillo, J., Luque-Almagro, V.M., Fuentes-Almagro, C., Urbano, F.J., Moreno-Vivián, C., Richardson, D.J., Roldán, M.D. (2018) Exploring the denitrification proteome of *Paracoccus denitrificans* PD1222. Front. Microbiol. 9, 1137.

[38] Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., Glockner, F. O. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. 35 (21), 7188–7196.

[39] Richter, M., Rosselló-Móra, R. (2009) Shifting the genomic gold standard for the prokaryotic species definition. Proc. Natl. Acad. Sci. U.S.A. 45, 19126–19131.

[40] Richter, M., Rosselló-Móra, R., Glöckner, F.O., Peplies, J. (2015) JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. Bioinformatics. 16, btv681.

[41] Rodriguez-R, L.M., Konstantinidis, K.T. (2016) The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ. 4, e1900v1, Preprints.

[42] ROSSELLO-MORA, R.A., GARCIA-VALDeS, E., LALUCAT, J. (1993) Taxonomic relationship between *Pseudomonas perfectomarina* ZoBell and *Pseudomonas stutzeri*. Int. J. Syst. Evol. Microbiol. 43 (4), 852–854.

[43] Sikorski, J., Rosselló-Móra, R., Lorenz, M.G. (1999) Analysis of genotypic diversity and relationships among *Pseudomonas stutzeri* strains by PCR-based genomic fingerprinting and multilocus enzyme electrophoresis. Syst. Appl. Microbiol. 22, 393–402.

[44] Stamatakis, A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 22, 2688–2690.

[45] Tajkarimi, M., Wexler, H.M. (2017) CRISPR-Cas Systems in *Bacteroides fragilis*, an important pathobiont in the human gut microbiome. Front. Microbiol. 8, 2234.

[46] Talwar, Chandni, Nagar, Shekhar, Lal, Rup, Negi, Ram Krishan (2018) Fish gut microbiome: current approaches and future perspectives. Indian J. Microbiol. 58 (4), 397–414.

[48] Viver, Tomeu, Cifuentes, Ana, Díaz, Sara, Rodríguez-Valdecantos, Gustavo, González, Bernardo, Antón, Josefa, Rosselló-Móra, Ramon (2015) Diversity of extremely halophilic cultivable prokaryotes in Mediterranean, Atlantic and Pacific solar salterns: Evidence that unexplored sites constitute sources of cultivable novelty. Syst. Appl. Microbiol. 38 (4), 266–275.

[49] Viver, Tomeu, Orellana, Luis, González-Torres, Pedro, Díaz, Sara, Urdiain, Mercedes, Farías, María Eugenia, Benes, Vladimir, Kaempfer, Peter, Shahinpei, Azadeh, Ali Amoozegar, Mohammad, Amann, Rudolf, Antón, Josefa, Konstantinidis, Konstantinos T., Rosselló-Móra, Ramon (2018) Genomic comparison between members of the *Salinibacteraceae* family, and description of a new species of *Salinibacter (Salinibacter altiplanensis* sp. nov.) isolated from high altitude hypersaline environments of the Argentinian Altiplano. Syst. Appl. Microbiol. 41 (3), 198–212.

[50] Wimmer, F., Beisel, C.L. (2020) CRISPR-Cas systems and the paradox of self-targeting spacers. Front. Microbiol. 10, 3078.

[51] Xiao, Yunhua, Liu, Xueduan, Liang, Yili, Niu, Jiaojiao, Zhang, Xian, Ma, Liyuan, Hao, Xiaodong, Gu, Yabin, Yin, Huaqun (2016) Insights into functional genes and taxonomical/phylogenetic diversity of microbial communities in biological heap leaching system and their correlation with functions. Appl. Microbiol. Biotechnol. 100 (22), 9745–9756.

[52] Zhang, Y., Song, C., Zhou, Z., Cao, X., Zhou, Y. (2019) Coupling between nitrification and denitrification as well as its effect on phosphorus release in sediments of Chinese shallow lakes. Water. 11, 1809.

[53] Zhou, Z., Liu, Y., Xu, W., Pan, J., Luo, Z.-H., Li, M. (2020) Genome- and community-level interaction insights into carbon utilization and element cycling functions of Hydrothermarchaeota in hydrothermal sediment. mSystems 5 (1). https://doi.org/10.1128/mSystems.00795-19.

[54] Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990) Basic local alignment search tool. J Mol Biol 215 (3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.