

**ORIGINAL ARTICLE**

# High-quality draft genome sequence of *Gaiella occulta* isolated from a 150 meter deep mineral water borehole and comparison with the genome sequences of other deep-branching lineages of the phylum *Actinobacteria*

Rita Severino<sup>1</sup> | Hugo J. C. Froufe<sup>2</sup> | Cristina Barroso<sup>1,2</sup> | Luciana Albuquerque<sup>1</sup> | Alexandre Lobo-da-Cunha<sup>3,4</sup> | Milton S. da Costa<sup>1</sup> | Conceição Egas<sup>1,2</sup> 

<sup>1</sup>Center for Neuroscience and Cell Biology, University of Coimbra, Coimbra, Portugal

<sup>2</sup>Next Generation Sequencing Unit, Biocant, Cantanhede, Portugal

<sup>3</sup>CIMAR/CIIMAR—Centro Interdisciplinar de Investigação Marinha e Ambiental, Universidade do Porto, Porto, Portugal

<sup>4</sup>Laboratório de Biologia Celular, Instituto de Ciências Biomédicas Abel Salazar, ICBAS, Universidade do Porto, Porto, Portugal

**Correspondence**

Conceição Egas, Center for Neuroscience and Cell Biology, University of Coimbra, 3004-504 Coimbra, Portugal.  
Email: cegas@biocant.pt

**Funding information**

HORIZON 2020, Grant/Award Number: project Metafluidics, ref. 685474; FEDER funds; Foundation for Science and Technology; GenomePT project; European Regional Development Fund

**Abstract**

*Gaiella occulta* strain F2-233<sup>T</sup> (=CECT 7815 = LMG 26412), isolated from a 150 meter deep mineral water aquifer, was deemed a candidate for high-quality draft genome sequencing because of the rare environment from which it was isolated. The draft genome sequence (QQZY00000000) of strain F2-233<sup>T</sup> is composed of approximately 3 Mb, predicted 3,119 protein-coding genes of which 2,545 were assigned putative functions. Genome analysis was done by comparison with the other deep-branching *Actinobacteria* neighbors *Rubrobacter radiotolerans*, *Solirubrobacter soli* and *Thermoleophilum album*. The genes for the tricarboxylic acid cycle, gluconeogenesis and pentose phosphate pathway, were identified in *G. occulta*, *R. radiotolerans*, *S. soli* and *T. album* genomes. Genes of the Embden–Meyerhof–Parnas pathway and nitrate reduction were identified in *G. occulta*, *R. radiotolerans* and *S. soli*, but not in the *T. album* genome. Alkane degradation is precluded by genome analysis in *G. occulta*. Genes involved in *myo*-inositol metabolism were found in both *S. soli* and *G. occulta* genomes. A Calvin–Benson–Bassham (CBB) cycle with a type I RuBisCO was identified in *G. occulta* genome, as well. However, experimental growth under several conditions was negative and CO<sub>2</sub> fixation could not be proven in *G. occulta*.

**KEYWORDS**

*Actinobacteria*, deep mineral water aquifer, *Gaiella occulta*, *Gaiellaceae*, genome

## 1 | INTRODUCTION

Bottled mineral waters originate from aquifers that are generally tapped from boreholes and piped to plants for bottling. According to European Union law mineral water cannot be disinfected to remove or decrease the number of microorganisms (Anonymous, 2009). However, the water can be treated to remove unstable

elements such as iron, manganese, sulfur, and arsenic or to (re) introduce carbon dioxide. Therefore, bottled mineral water contains a complex microbiota that, presumably, originates from the source. Quality control procedures are mandatory to monitor the presence/absence of indicator bacteria for fecal contamination, as well as surface water infiltration and pathogenic bacteria such as *Pseudomonas aeruginosa* (Anonymous, 2009). Heterotrophic plate

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *MicrobiologyOpen* published by John Wiley & Sons Ltd.

counts are also monitored to record alterations in the number of colony-forming units (CFUs).

The majority of studies on the microbial diversity of bottled water have been performed on still natural mineral waters using culture-dependent approaches (Guillot & Leclerc, 1993; Morais & da Costa, 1990; Vachée, Mossel, & Leclerc, 1997). Microbial abundances estimated by CFUs indicate that heterotrophic bacteria number is low at source (around 10 CFU/ml) but increases to about  $10^4$ – $10^5$  CFU/ml during storage at room temperature (Croville, Cantet, & Saby, 2011; Morais & da Costa, 1990; Warburton, 1993). More recently culture-dependent and culture-independent techniques were used together to determine the microbial diversity and abundances present at source of one mineral water, to assess microbial stability of the source over a 1 year period, and to examine the microbial dynamics after bottling throughout 6 months of storage of the mineral water in factory produced plastic bottles (França, López-Lopéz, Rosselló-Móra, & da Costa, 2015). In all cases, communities were largely dominated by *Bacteria* affiliated with the *Alphaproteobacteria*, *Betaproteobacteria*, and *Gammaproteobacteria*. Several isolates representing new species were also characterized and described from the aquifer and the bottled mineral water (Albuquerque et al., 2011; França, Albuquerque, & da Costa, 2015; Franca, Albuquerque, Sanchez, Farelaira, & da Costa, 2017; Leandro, França, Nobre, Rainey, & da Costa, 2013; Leandro et al., 2012). Among them, *Gaiella occulta* is the sole representative of the family *Gaiellaceae* of the order *Gaiellales* within the deep-branching lineages of the phylum *Actinobacteria* and was deemed a candidate for high-quality genome sequencing given the rare environment from which it was isolated. The phylum *Actinobacteria* comprises several deeply branching lineages that consist of species of the orders *Rubrobacterales*, *Solirubrobacterales*, *Thermoleophilales*, and *Gaiellales* (Foesel, Geppert, Rohde, & Overmann, 2016). The order *Gaiellales* and the family *Gaiellaceae* comprises only the species *G. occulta* strain F2-233<sup>T</sup> (Albuquerque et al., 2011).

## 1.1 | Organism information

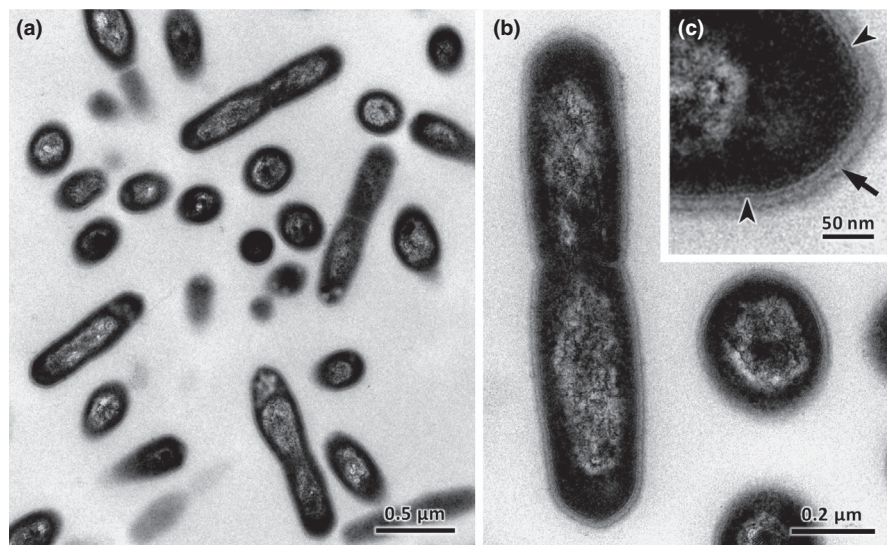
*Gaiella occulta* is a nonpigmented, Gram-negative staining, nonmotile, aerobic, and chemoorganotrophic, with an optimal growth temperature of 35–37°C, optimum pH for growth between 6.5 and 7.5, and was isolated from a 150 meter deep water aquifer in Portugal (Albuquerque et al., 2011; Foesel et al., 2016). The strain forms short rod-shaped cells 1.0–3.0 µm in length by 0.3–0.5 µm in width (by transmission electron microscopy) (Figure 1). Strain F2-233<sup>T</sup> can assimilate carbohydrates, organic acids, and amino acids. Nitrate is reduced to nitrite; long chain *n*-alkanes are not used as carbon and energy source.

16S rRNA gene sequence comparisons show *G. occulta* to belong to the family *Gaiellaceae* of the order *Gaiellales* which represents a deep branch lineage of the phylum *Actinobacteria*. The phylogenetic tree based on the 16S rRNA gene of strain F2-233<sup>T</sup> in the deep branching taxa of the *Actinobacteria* shows *G. occulta* F2-233<sup>T</sup> to form a distinct lineage separate from those of the orders *Rubrobacterales*, *Thermoleophilales*, and *Solirubrobacterales* (Figure 2).

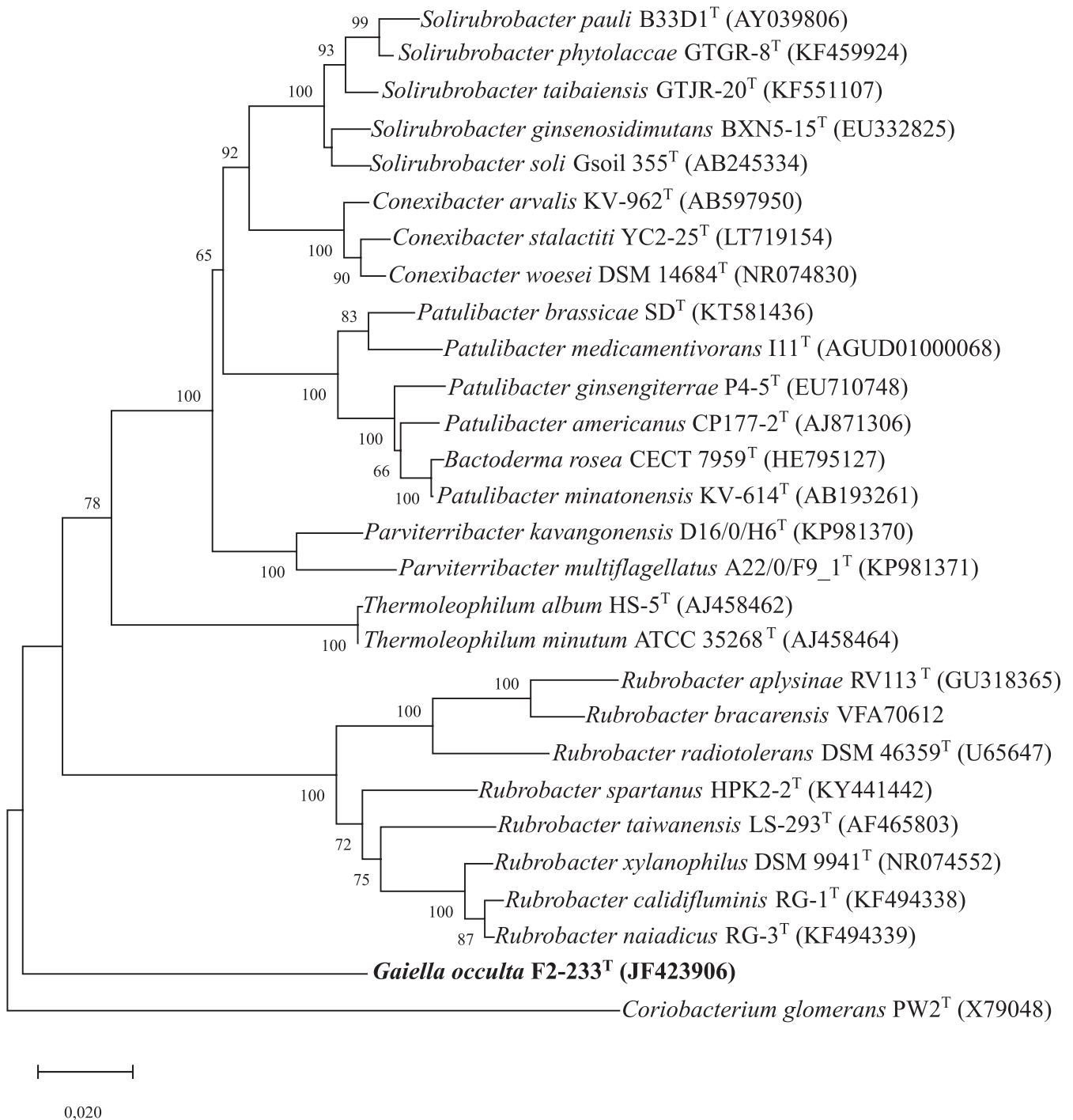
## 2 | MATERIALS AND METHODS

### 2.1 | Growth conditions and genomic DNA preparation

Strain F2-233<sup>T</sup> was grown in 1 L Erlenmeyer flasks containing 300 ml of R2A medium ([http://www.dsmz.de/microorganisms/medium/pdf/DSMZ\\_Medium830.pdf](http://www.dsmz.de/microorganisms/medium/pdf/DSMZ_Medium830.pdf)) at 37°C in a rotary water bath shaker until late exponential phase of growth for DNA extraction. To ascertain CO<sub>2</sub> fixation and the enzymatic activity of RuBisCo, *G. occulta* was grown under aerobic and anaerobic conditions in sealed 50 ml serum ampules at 37°C containing 40 ml of a minimal medium previously described, supplemented with 0.02% yeast extract (Hahnke, Moosmann, Erb, & Strous, 2014). The medium contained (per liter) 0.5 g (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.5 g MgSO<sub>4</sub>·7H<sub>2</sub>O, 0.1 g CaCl<sub>2</sub>·2H<sub>2</sub>O, 6 g HEPES, 0.12 g K<sub>2</sub>HPO<sub>4</sub>, 0.04 g KH<sub>2</sub>PO<sub>4</sub>, and 1 ml



**FIGURE 1** Transmission electron microscopy of *Gaiella occulta*. (a) Low amplification of a group of cells in longitudinal section and cross section. (b) Dividing cell with central septum. (c) Higher magnification showing a thin peptidoglycan layer (arrow) and outer leaflet of the cell membrane (arrow head)



**FIGURE 2** 16S rRNA gene sequence phylogenetic tree. The position of *Gaiella occulta* within the radiation of the deep-branching taxa of the *Actinobacteria* is shown. The scale bar represents two inferred substitutions per 100 nucleotides. The numbers at branching points of the neighbor-joining tree represent bootstrap values from 1,000 replications

trace element solution. Phosphate was added after sterilization. The trace element solution contained (per liter) 7.3 g Na<sub>2</sub>EDTA, 0.085 g CuCl<sub>2</sub>·2H<sub>2</sub>O, 0.72 g ZnSO<sub>4</sub>·7H<sub>2</sub>O, 2.5 g FeSO<sub>4</sub>·7H<sub>2</sub>O, 0.02 g MnCl<sub>2</sub>·4H<sub>2</sub>O, 0.242 g Na<sub>2</sub>MoO<sub>4</sub>·2H<sub>2</sub>O and 2 g NaHCO<sub>3</sub>. The headspace for anaerobic growth was flushed with N<sub>2</sub> for 5 min. Formate, acetate and pyruvate were tested as sole carbon sources and mixtures of formate and acetate, or formate and pyruvate were

used to improve growth. In order to test the utilization of hydrogen as electron acceptor, approximately 3 ml H<sub>2</sub> was also added to anaerobic and aerobic ampules after sterilization and prior to inoculation. Cell density was measured at 610 nm.

Total genomic DNA was extracted following the method of Nielsen et al. (Nielsen, Fritze, & Priest, 1995). Briefly, cells were lysed with a solution of lysozyme, guanidium thiocyanate, and sodium

**TABLE 1** Genome and annotation statistics for *Gaiella occulta* F2-233<sup>T</sup>

Attribute	Value	% of Total
Genome size (bp)	3,028,529	100
DNA coding (bp)	2,785,809	91.99
DNA G + C (bp)	2,169,941	71.65
DNA scaffolds	34	
Total genes	3,167	100
Protein coding genes	3,119	98.48
RNA genes	48	1.51
Genes with function prediction	2,545	81.60
Genes assigned to COGs	1,718	54.24
Genes with Pfam domains	2,502	79.00
CRISPR repeats	1	

*n*-lauryl sarcosine. DNA was extracted with chloroform:isoamyl alcohol (24:1, v:v), precipitated with isopropanol and washed with 70% ethanol, dried and resuspended in water. RNase was included in the extraction process. DNA purity was assessed in a 1% agarose gel electrophoresis. DNA was quantified by fluorescence in the Invitrogen Qubit<sup>®</sup> 2.0 fluorometer (Thermo Fisher Scientific, Carlsbad, CA).

## 2.2 | Genome sequencing and assembly

Genomic DNA was prepared with the Nextera XT DNA Library Preparation Kit (Illumina) and sequenced using the paired-end (PE) 2x300 bp V3 kit on the MiSeq<sup>®</sup> (Illumina, San Diego) at the Next Generation Sequencing Unit, Center for Neuroscience and Cell Biology (CNC/Biocant, Cantanhede, Portugal). Sequenced reads were quality filtered with Trimmomatic (Bolger, Lohse, & Usadel, 2014) and assembled with SPAdes version 3.9.0 (Bankevich et al., 2012).

## 2.3 | Genome annotation and analysis

The resulting contigs were annotated with PGP (Prokaryotic Genome Prediction) (Egas et al., 2014). Genome estimated completeness was verified with CheckM (version 1.0.7) based on lineage-specific marker sets (Parks, Imelfort, Skennerton, Hugenholtz, & Tyson, 2015). Contamination was tested by CheckM for protein-coding genes and RNAmmer version 1.2 (Lagesen et al., 2007) and Usearch61 (Edgar, 2010) (with the Greengenes database, version 13.8, ID threshold of 0.6 and an e-value of  $1e^{-6}$ ) for complete or partial 16S rRNA gene. The genome of strain F2-233<sup>T</sup> was compared with the available genomes of organisms of deeply branching orders of the *Actinobacteria*, namely *Rubrobacter radiotolerans* RSPS-4 (NZ\_CP007514.1, NZ\_CP007515.1, NZ\_CP007516.1 and NZ\_CP007517.1), *Solirubrobacter soli* DSM 22325<sup>T</sup> (NZ\_AUIK0000000.1), and *Thermoleophilum album* strain ATCC 35263<sup>T</sup> (NZ\_FNWJ0000000.1), with GET\_HOMOLOGUES using BLASTP and OrthoMCL (Contreras-Moreira & Vinuesa,

**TABLE 2** Number of genes associated with general COG functional categories

Code	Value	% of Total	Description
J	137	4.39	Translation, ribosomal structure and biogenesis
A	0	-	RNA processing and modification
K	150	4.81	Transcription
L	83	2.66	Replication, recombination and repair
B	1	0.03	Chromatin structure and dynamics
D	23	0.74	Cell cycle control, cell division, chromosome partitioning
V	19	0.61	Defense mechanisms
T	63	2.02	Signal transduction mechanisms
M	116	3.72	Cell wall/membrane biogenesis
N	36	1.15	Cell motility
U	43	1.38	Intracellular trafficking and secretion
O	65	2.08	Posttranslational modification, protein turnover, chaperones
C	170	5.45	Energy production and conversion
G	132	4.23	Carbohydrate transport and metabolism
E	263	8.43	Amino acid transport and metabolism
F	67	2.15	Nucleotide transport and metabolism
H	108	3.46	Coenzyme transport and metabolism
I	89	2.85	Lipid transport and metabolism
P	89	2.85	Inorganic ion transport and metabolism
Q	49	1.57	Secondary metabolites biosynthesis, transport and catabolism
R	240	7.69	General function prediction only
S	118	3.78	Function unknown
-	0	44.92	Not in COGs

2013). Orthologous genes were annotated against the Kyoto Encyclopedia of Genes and Genomes (KEGG) and assigned to metabolic pathways (sequence similarity cutoff e-value of  $1e^{-5}$ ) using KOBAS 2.0 (Xie et al., 2011).

One partial genome sequence of a *Gaiella* sp. (LSTI01000000) appearing in the database was amplified from a single cell recovered from soil. The sequence contains 995,360 bp with 1,036 genes coding for 915 proteins the vast majority of which are annotated as hypothetical proteins. The organism is closely related to *G. occulta*

(95%) based on the phylogenetic analysis of a truncated 16S rRNA gene with 892 bp. The G + C ratio of the DNA is 66.4%.

### 3 | RESULTS AND DISCUSSION

#### 3.1 | Genome properties

The F2-233<sup>T</sup> strain DNA sequence run generated 5,261,564 paired-end reads of which 3,362,091 high-quality reads remained after quality filtering. The average read length was of 197 bp. The *de novo* read assembly produced 34 contigs with an N50 size of 401,372 bp. The high-quality draft assembled genome sequence consisted of 3,028,529 bp, with a sequencing depth of coverage of 520-fold and a DNA G + C content of 71.65% (Table 1). The genome had a total of 3,167 genes, including 2,545 protein-coding genes, 45 tRNA genes and three rRNA genes (a single copy of the genes 23S, 16S and 5S) (Table 1). CheckM estimated the genome to be near-completion (97.84%) and the level of contamination to be extremely low (1.29%). No contamination was detected for 16S rRNA genes as tested by RNAmmer and Usearch61.

The draft genome sequence of *G. occulta* F2-233<sup>T</sup> (=CECT 7815<sup>T</sup> = LMG 26415<sup>T</sup>) has been deposited in the Short Read Archive (SRA) under the accession number SRR7537062 and the genome assembly under the accession number QQZY00000000.

#### 3.2 | Genome annotation

The draft genome comprised 2,545 genes with putative functions (~82% of total protein-coding genes) and 1,718 genes allocated to the Clusters of orthologous groups (COG) functional categories (55% of total protein-coding genes). The most abundant COG category was "Amino acid transport and metabolism" followed by "General function prediction only" and "Energy production and conversion" (Table 2).

#### 3.3 | Insights from the genome sequence

##### 3.3.1 | Central metabolism

Genes coding for the enzymes of the Embden–Meyerhof–Parnas pathway (EMP), namely ATP-dependent 6-phosphofructokinase (EC 2.7.1.11, Gocc\_2786, Gocc\_2787) and fructose-bisphosphate aldolase (EC 4.1.2.13, Gocc\_1033) were identified in the genome of *G. occulta* as well as in the genomes of *R. radiotolerans* and *S. soli*. However, a fructose-bisphosphate aldolase gene was not identified in the genome sequence of *T. album*, predicting that the organism cannot channel hexoses through this pathway.

The genes coding for 2-keto-3-deoxygluconate-6-phosphate aldolase (EC 4.1.2.14) and 6-phosphogluconate dehydratase (EC 4.2.1.12), characteristic of the Entner–Doudoroff pathway were not found in the four genomes, indicating that glycolysis does not proceed through this pathway. The gene coding for the enzyme fructose-1,6-bisphosphatase (EC 3.1.3.11, Gocc\_1034), required for

gluconeogenesis, is present in the genome of *G. occulta* and in the genomes of *R. radiotolerans*, *S. soli*, and *T. album*. All putative genes coding for the enzymes of the pentose phosphate pathway and tricarboxylic acid cycle were identified in the genomes of *G. occulta*, *R. radiotolerans* and *S. soli* and *T. album*.

*Gaiella occulta* and *S. soli* utilize myo-inositol as single carbon and energy source for growth (Albuquerque et al., 2011; Kim et al., 2007). Several genes coding for enzymes involved in the metabolism of myo-inositol, namely *iolABCDEG*, were identified in *G. occulta* and *S. soli* genomes (Yoshida et al., 2008). However, the gene *iolJ* that codes for 6-phospho-5-dehydro-2-deoxy-D-gluconate aldolase (EC 4.1.2.29) and leads to the formation of dihydroxyacetone-phosphate and malonate semialdehyde was not identified in the genomes of these two organisms. This gene was also not found in the genomes of other bacterial strains capable of myo-inositol catabolism such as *Clostridium perfringens*, *Legionella pneumophila*, and *Enterobacter aerogenes*. It is possible that the *iolJ* gene is replaced by a yet annotated aldolase gene (Berman & Magasanik, 1966; Kawsar, Ohtani, Okumura, Hayashi, & Shimizu, 2004; Manske, Schell, & Hilbi, 2016).

The genome predicts a Calvin–Benson–Bassham (CBB) cycle with a type I RuBisCO (EC 4.1.1.39, Gocc\_0241 and Gocc\_0242) but lacks sedoheptulose-bisphosphatase (EC 3.1.3.37) that leads to the synthesis of sedoheptulose-phosphate. This cycle is also predicted in the genome of *S. soli*. In *S. soli* DSM 22325<sup>T</sup> the fructose-1,6-bisphosphatase/sedoheptulose-1,7-bisphosphatase gene is predicted to have a dual function, and this may also be the case for *G. occulta*. Other genes potentially involved in CO<sub>2</sub> metabolism were identified in *G. occulta*, namely a formate dehydrogenase (EC 1.17.99.7, Gocc\_1455), a putative hydrogenase (Gocc\_1122 to Gocc\_1127) and three carbon monoxide dehydrogenases (CODH, EC 1.2.5.3, Gocc\_1071 to Gocc\_1073, Gocc\_1636 to Gocc\_1638 and Gocc\_2945 to Gocc\_2947) (Shi et al., 2015). Several CODH genes were also identified in *R. radiotolerans* and *S. soli*. Growth was not observed under anaerobic conditions. Growth was observed under aerobic conditions in media containing acetate and pyruvate or combinations of formate plus acetate and formate plus pyruvate. The addition of hydrogen did not enhance growth under any condition examined. In all cases growth under aerobiosis was low and it was impossible to obtain enough cell mass to perform enzymatic assays for RuBisCO activity. We, therefore, predict but do not confirm CO<sub>2</sub> fixation via the CBB cycle in this organism. A complete CBB cycle has also been identified in the genome of *Thermus* sp. NMX2.A1 although CO<sub>2</sub> fixation through this pathway has not been confirmed in this organism (Müller et al., 2016).

A complete electron transport chain is predicted in *G. occulta* genome, namely NADH dehydrogenase (EC 1.6.5.3, Gocc\_2267 to Gocc\_2280), succinate dehydrogenase (EC 1.3.5.1, Gocc\_1693 to Gocc\_1695), cytochrome *bc1* (EC 1.10.2.-, Gocc\_3013 to Gocc\_3015), and cytochrome *c* oxidase (EC 1.9.3.1, Gocc\_0001, Gocc\_0002, Gocc\_0004). Homologs of *G. occulta* NADH dehydrogenase, succinate dehydrogenase, cytochrome *bc1* and cytochrome *c* oxidase genes were identified in the genomes of *S. soli* and *T. album*.

Both *G. occulta*, *S. soli*, and *T. album* succinate dehydrogenase complex seems to lack the gene *sdhD*, which codes for the anchor subunit D, although the absence of this gene is not unusual (Horsefield, Iwata, & Byrne, 2004). In the genome of *R. radiotolerans*, we could not identify homologs for the following genes, NADH dehydrogenase *nuoE*, *nuoF* and *nuoG*, succinate dehydrogenase *sdhC*, and cytochrome *bc1* cytochrome-*c* subunit. The ATPase of the type strains examined here are of the common bacterial F-type.

*n*-Alkanes with 13–20 carbons in length are utilized for growth of *T. album* (Zarilla & Perry, 1984). Genes involved in alkane degradation were identified in the *T. album* genome, namely an AlkG2-type rubredoxin (WP\_093117361.1), a putative flavin-containing monooxygenase (EC 1.14.13.8, WP\_093115507.1), and NAD(P)/FAD-dependent oxidoreductase (WP\_093116100.1), which shares 47.8% amino acid sequence identity with *Acinetobacter* sp. *almA* (A5H9N6), involved in the degradation of long-chain *n*-alkanes (Rojo, 2009; Smits, Witholt, & van Beilen, 2003; Van Beilen et al., 2002). Homologs of *T. album* alkane-degradation genes were also identified in *S. soli*. Genes for the metabolism of alkanes were not identified in the genome of *R. radiotolerans*, where alkane utilization was not examined, or in *G. occulta*, that does not utilize alkanes as growth sources (Albuquerque et al., 2011).

Genes coding for the uptake and reduction of nitrate were identified in the genomes of *G. occulta*, *R. radiotolerans*, and *S. soli*, namely the MFS-type nitrate/nitrite transporter (*narK/nasA*, Gocc\_2854) and the respiratory *narGHJ* nitrate reductase complex (EC 1.7.5.1, Gocc\_2855 to Gocc\_2859), although the reduction of nitrate to nitrite was not observed in *S. soli* (Kim et al., 2007). Genes coding for the two-component system transduction pathway *NarX/NarL* identified in *G. occulta* (Gocc\_1932 and Gocc\_1933) were not identified in *S. soli* genome, but nitrate reductase expression does not seem to be altered by the absence of *narX* (Laub & Goulian, 2007; Sohaskey & Wayne, 2003). The *sox* genes, as well as other genes involved in sulfite oxidation/sulfate reduction, namely adenylylsulfate reductase (EC 1.8.99.2), dissimilatory sulfite reductase (EC 1.8.99.5) or sulfite dehydrogenase (cytochrome) (EC 1.8.2.1), were not identified in the *G. occulta* genome sequence or the other deeply branching *Actinobacteria* whose genome has been sequenced, precluding the utilization of reduced sulfur compounds as electron donors.

*Gaiella occulta* uses fructose, glucose, mannose, ribose, xylose, and *myo*-inositol as single carbon source for growth (Albuquerque et al., 2011). The genes for ABC transport systems for ribose (Gocc\_0079 to Gocc\_0081; Gocc\_2298 to Gocc\_2300; Gocc\_3068, Gocc\_3069 and Gocc\_3071), rhamnose (Gocc\_0123 to Gocc\_0125; Gocc\_0231 to Gocc\_0234), D-xylose (Gocc\_1293 to Gocc\_1295), and raffinose/stachyose/melibiose (Gocc\_1613 to Gocc\_1616) were identified in *G. occulta* genome. Four unspecific sugar transport systems (Gocc\_0123 to Gocc\_0125, Gocc\_0222 to Gocc\_0224, Gocc\_2321 to Gocc\_2323 and Gocc\_3042 to Gocc\_3044), that may be involved in the transport of glucose, fructose, mannose, and *myo*-inositol, were also identified. No PTS-type transporters were identified in *G. occulta* genome. By in large, *S. soli* and *R. radiotolerans* have similar ABC transporters for sugars (Egas et al., 2014).

### 3.3.2 | Stress response

*Gaiella occulta* genome sequence has the key enzymes for the main DNA repair mechanisms, except for the mismatch repair pathway. Genes *mutS* and *mutL* were not encountered, as in many *Actinobacteria* and *Archaea* (Castañeda-García et al., 2017). A gene coding for the endonuclease *NucS* (Gocc\_1770) was identified, suggesting *G. occulta* may use the noncanonical mismatch repair pathway described recently for *Mycobacterium smegmatis* and *Streptomyces coelicolor* (Castañeda-García et al., 2017). *Thermoleophilum album* and *S. soli* may also use this alternative mismatch repair pathway as they also lack homologs of *mutS* and *mutL* and have a *nucS* homolog.

*Rubrobacter xylanophilus* and *R. radiotolerans* accumulate the compatible solutes mannosylglycerate, trehalose, and low levels of di-*myo*-inositol-phosphate generally involved in osmotic adaptation in (hyper)thermophilic organisms that in the *Rubrobacter* spp. are constitutively accumulated (Empadinhas et al., 2007; Nobre, Alarico, Fernandes, Empadinhas, & da Costa, 2008). Genes coding for enzymes involved in the synthesis of mannosylglycerate or di-*myo*-inositol-phosphate were not identified in *G. occulta*. In *R. radiotolerans*, trehalose synthesis can proceed via four pathways namely *TpS/TpP*, *TreS*, *TreT*, and *TreY/TreZ* (Egas et al., 2014; Nobre et al., 2008). In *G. occulta* only the genes coding for *TpS/TpP* and *TreS* were detected in the genome (*tpS*, EC 2.4.1.15, Gocc\_2203, *tpP*, EC 3.1.3.12, Gocc\_2154, *treS*, EC 5.4.99.16, Gocc\_0097). The genes *tpS* and *tpP* were also identified in the genomes of *S. soli* and *T. album*.

Several genes involved in reactive oxygen species (ROS) detoxification in *R. radiotolerans* were identified in the genome of *G. occulta*, namely, two genes encoding peroxiredoxins (EC 1.11.1.15, Gocc\_0321, Gocc\_2238), which reduce hydrogen peroxide to water, genes coding for thioredoxin *TrxA* (Gocc\_1414), thioredoxin reductase *TrxB* (EC 1.8.1.9, Gocc\_2333), and glutaredoxin *GrxC* (Gocc\_0876), involved in redox balance, three genes coding for LysR transcriptional regulators (Gocc\_1704, Gocc\_1792 and Gocc\_0289) and one ABC-type  $Mn^{2+}/Zn^{2+}$  transport system (Gocc\_1485, Gocc\_1486) (Egas et al., 2014). In *T. album* genome we could not identify *lysR* transcriptional regulators or ABC-type  $Mn^{2+}/Zn^{2+}$  transport system genes.

## 4 | CONCLUSIONS

The genome of *G. occulta*, the sole representative of the order *Gaiellales*, was sequenced, analyzed and compared to the existing genomes of the most closely related actinobacterial deep-branching species *R. radiotolerans*, *S. soli*, and *T. album*. The main objective was to compare the results of the genome sequence analysis with the phenotypic characteristics of these organisms. Despite the phylogenetic distances between these strains obtained by 16S rRNA gene sequence analysis and different phenotypic characteristics, genome sequence analysis showed that many characteristics were shared among these organisms. With the exception of *T. album*, the genome analysis of the three other strains were similar with respect to the metabolism of hexoses and the central metabolism in general. We

conclude that *T. album* cannot use sugars for lack of transporters and the lack of a fructose-bisphosphate aldolase gene for metabolism of sugars. The genome of *G. occulta* and *S. soli* appear to predict CO<sub>2</sub> fixation via the CBB cycle, however, CO<sub>2</sub> fixation could not be proven. The type strain of *T. album* has genes that predict the hydrolysis of *n*-alkanes and is known to be able to use only these substrates for growth. *S. soli* also possesses genes for the degradation of *n*-alkanes, but growth on these substrates were not examined in this organism. *Gaiella occulta* does not possess the homologs found in *T. album* and does not grow on *n*-alkanes. The results obtained in this study indicate that all organisms appear to be strict chemoorganotrophs and, for the most part, corroborate the phenotypes of these strains.

## ACKNOWLEDGEMENTS

We would like to thank Fred Rainey, University of Alaska Anchorage USA, for the phylogenetic analysis. This work was supported by HORIZON 2020, under the project Metafluidics, ref. 685474. This work was also supported by FEDER funds through the Operational Programme Competitiveness Factors—COMPETE 2020 and national funds by FCT—Foundation for Science and Technology under the strategic project UID/NEU/04539/2013. H. J. C. F. was supported by the GenomePT project (POCI-01-0145-FEDER-022184), with funds from COMPETE 2020—Operational Programme for Competitiveness and Internationalisation (POCI), Lisboa, Portugal Regional Operational Programme (Lisboa2020), Algarve Portugal Regional Operational Programme (CRESC Algarve2020), under the PORTUGAL 2020 Partnership Agreement, through the European Regional Development Fund (ERDF), and by Fundação para a Ciência e a Tecnologia (FCT).

## CONFLICT OF INTERESTS

None declared.

## AUTHORS CONTRIBUTION

L. A. grew the organism, and extracted DNA. C. B. sequenced *G. occulta* genome. H. J. C. F. assembled, annotated and compared the genome sequence to other genome sequences available. A. L.-d.-C. performed Transmission Electron Microscopy. R. S. conducted anaerobic growth experiments and analyzed metabolic pathways. R. S., C. E. and M. S. d. C. wrote the manuscript.

## ETHICS STATEMENT

None required.

## DATA ACCESSIBILITY

The genome sequence of *G. occulta* F2-233<sup>T</sup> is publicly available in the SRA under the accession number SRR7537062 and the genome assembly under the accession number QQZY00000000.

## ORCID

Conceição Egas  <https://orcid.org/0000-0002-2307-5414>

## REFERENCES

- Albuquerque, L., França, L., Rainey, F. A., Schumann, P., Nobre, M. F., & da Costa, M. S. (2011). *Gaiella occulta* gen. nov., sp. nov., a novel representative of a deep branching phylogenetic lineage within the class *Actinobacteria* and proposal of *Gaiellaceae* fam. nov. and *Gaiellales* ord. nov. *Systematic and Applied Microbiology*, 34(8), 595–599. <https://doi.org/10.1016/j.syapm.2011.07.001>
- Anonymous. (2009). Directive 2009/54/EC of the European Parliament and of the Council of 18 June 2009 on the exploitation and marketing of natural mineral water. *Official Journal of the European Union*, L164, 45–58.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., ... Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, 19(5), 455–477. <https://doi.org/10.1089/cmb.2012.0021>
- Berman, T., & Magasanik, B. (1966). The pathway of myo-inositol degradation in *Aerobacter aerogenes*. *The Journal of Biological Chemistry*, 241, 800–806.
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Castañeda-García, A., Prieto, A. I., Rodríguez-Beltrán, J., Alonso, N., Cantillon, D., Costas, C., ... Blázquez, J. (2017). A non-canonical mismatch repair pathway in prokaryotes. *Nature Communications*, 8, 14246. <https://doi.org/10.1038/ncomms14246>
- Contreras-Moreira, B., & Vinuesa, P. (2013). GET\_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Applied and Environmental Microbiology*, 79(24), 7696–7701. <https://doi.org/10.1128/AEM.02411-13>
- Croville, J. L., Cantet, J., & Saby, S. (2011). Water treatments. In N. Dege (Ed.), *Technology of bottled water* (3rd ed., pp. 141–174). Wiley-Blackwell.
- Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26(19), 2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>
- Egas, C., Barroso, C., Froufe, H. J. C., Pacheco, J., Albuquerque, L., & da Costa, M. S. (2014). Complete genome sequence of the Radiation-Resistant bacterium *Rubrobacter radiotolerans* RSPS-4. *Standards in Genomic Sciences*, 9(3), 1062–1075. <https://doi.org/10.4056/sigs.5661021>
- Empadinhas, N., Mendes, V., Simões, C., Santos, M. S., Mingote, A., Lamosa, P., ... Costa, M. S. d. (2007). Organic solutes in *Rubrobacter xylanophilus*: The first example of di-myo-inositol-phosphate in a thermophile. *Extremophiles*, 11(5), 667–673. <https://doi.org/10.1007/s00792-007-0084-z>
- Foesel, B. U., Geppert, A., Rohde, M., & Overmann, J. (2016). *Parviterribacter kavangonensis* gen. nov., sp. nov. and *Parviterribacter multiflagellatus* sp. nov., novel members of *Parviterribacteraceae* fam. nov. within the order *Solirubrobacterales*, and emended descriptions of the classes *Thermoleophilia* and *Rubrobacteria* and their orders and families. *International Journal of Systematic and Evolutionary Microbiology*, 66(2), 652–665. <https://doi.org/10.1099/ijsem.0.000770>
- França, L., Albuquerque, L., & da Costa, M. S. (2015). *Cavicella subterranea* gen. nov., sp. nov., isolated from a deep mineral-water aquifer, and emended description of the species *Perluclidibaca piscinae*. *International Journal of Systematic and Evolutionary Microbiology*, 65(11), 3812–3817. <https://doi.org/10.1099/ijsem.0.000493>

- Franca, L., Albuquerque, L., Sanchez, C., Farelaira, P., & da Costa, M. S. (2017). *Ampullimonas aquatilis* gen. nov., sp. nov. isolated from bottled mineral water. *International Journal of Systematic and Evolutionary Microbiology*, 66, 1459–1465. <https://doi.org/10.1099/ijsem.0.000903>
- França, L., Lopéz-Lopéz, A., Rosselló-Móra, R., & da Costa, M. S. (2015). Microbial diversity and dynamics of a groundwater and a still bottled natural mineral water. *Environmental Microbiology*, 17(3), 577–593. <https://doi.org/10.1111/1462-2920.12430>
- Guillot, E., & Leclerc, H. (1993). Biological specificity of bottled natural mineral waters: Characterization by ribosomal ribonucleic acid gene restriction patterns. *Journal of Applied Bacteriology*, 75(3), 292–298. <https://doi.org/10.1111/j.1365-2672.1993.tb02779.x>
- Hahnke, S. M., Moosmann, P., Erb, T. J., & Strous, M. (2014). An improved medium for the anaerobic growth of *Paracoccus denitrificans* Pd1222. *Frontiers in Microbiology*, 5(18), 1–9. <https://doi.org/10.3389/fmicb.2014.00018>
- Horsefield, R., Iwata, S., & Byrne, B. (2004). Complex II from a structural perspective. *Current Protein & Peptide Science*, 5(2), 107–118. <https://doi.org/10.2174/1389203043486847>
- Kawsar, H. I., Ohtani, K., Okumura, K., Hayashi, H., & Shimizu, T. (2004). Organization and transcriptional regulation of *myo*-inositol operon in *Clostridium perfringens*. *FEMS Microbiology Letters*, 235(2), 289–295. <https://doi.org/10.1016/j.femsle.2004.04.047>
- Kim, M. K., Na, J.-R., Lee, T.-H., Im, W.-T., Soung, N.-K., & Yang, D.-C. (2007). *Solirubrobacter soli* sp. nov., isolated from soil of a ginseng field. *International Journal of Systematic and Evolutionary Microbiology*, 57, 1453–1455. <https://doi.org/10.1099/ijms.0.64715-0>
- Lagesen, K., Hallin, P., Rødland, E. A., Stærfeldt, H. H., Rognes, T., & Ussery, D. W. (2007). RNAMmer: Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research*, 35(9), 3100–3108. <https://doi.org/10.1093/nar/gkm160>
- Laub, M. T., & Goulian, M. (2007). Specificity in two-component signal transduction pathways. *Annual Review of Genetics*, 41(1), 121–145. <https://doi.org/10.1146/annurev.genet.41.042007.170548>
- 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*. *International Journal of Systematic and Evolutionary Microbiology*, 63, 3793–3799. <https://doi.org/10.1099/ijms.0.050021-0>
- Leandro, T., França, L., Nobre, M. F., Schumann, P., Rosselló-Móra, R., & da Costa, M. S. (2012). *Nevskia aquatilis* sp. nov. and *Nevskia persephonica* sp. nov., isolated from a mineral water aquifer and the emended description of the genus *Nevskia*. *Systematic and Applied Microbiology*, 35(5), 297–301. <https://doi.org/10.1016/j.syapm.2012.05.001>
- Manske, C., Schell, U., & Hilbi, H. (2016). Metabolism of *myo*-inositol by *Legionella pneumophila* promotes infection of amoeba and macrophages. *Applied and Environmental Microbiology*, 82(16), 5000–5014. <https://doi.org/10.1128/AEM.01018-16>
- Morais, P. V., & da Costa, M. S. (1990). Alterations in the major heterotrophic bacterial populations isolated from a still bottled mineral water. *Journal of Applied Bacteriology*, 69(5), 750–757. <https://doi.org/10.1111/j.1365-2672.1990.tb01572.x>
- Müller, W. J., Tlalajoe, N., Cason, E. D., Litthauer, D., Reva, O., Brzuszkiewicz, E., & vanHeerden, E. (2016). Whole Genome Comparison of *Thermus* sp. NMX2.A1 Reveals Principal Carbon Metabolism Differences with Closest Relation *Thermus scotoductus* SA-01. G3: GENES, GENOMES, GENETICS, 6(9), 2791–2797. <https://doi.org/10.1534/g3.116.032953>
- Nielsen, P., Fritze, D., & Priest, F. G. (1995). Phenetic diversity of alkaliphilic *Bacillus* strains: Proposal for nine new species. *Microbiology*, 141(7), 1745–1761. <https://doi.org/10.1099/13500872-141-7-1745>
- Nobre, A., Alarico, S., Fernandes, C., Empadinhas, N., & da Costa, M. S. (2008). A unique combination of genetic systems for the synthesis of trehalose in *Rubrobacter xylanophilus*: Properties of a rare actinobacterial TreT. *Journal of Bacteriology*, 190(24), 7939–7946. <https://doi.org/10.1128/JB.01055-08>
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, 25(7), 1043–1055. <https://doi.org/10.1101/gr.186072.114>
- Rojo, F. (2009). Degradation of alkanes by bacteria: Minireview. *Environmental Microbiology*, 11(10), 2477–2490. <https://doi.org/10.1111/j.1462-2920.2009.01948.x>
- Shi, J., Jiang, Y., Jiang, Z., Wang, X., Wang, X., Zhang, S., ... Yang, C. (2015). Enzymatic conversion of carbon dioxide. *Chemical Society Reviews*, 44, 5981–6000. <https://doi.org/10.1039/c5cs00182j>
- Smits, T. H. M., Witholt, B., & van Beilen, J. B. (2003). Functional characterization of genes involved in alkane oxidation by *Pseudomonas aeruginosa*. *Antonie Van Leeuwenhoek, International Journal of General and Molecular Microbiology*, 84(3), 193–200. <https://doi.org/10.1023/A:1026000622765>
- Sohaskey, C. D., & Wayne, L. G. (2003). Role of narK2X and narGHJ in Hypoxic Upregulation of Nitrate Reduction by *Mycobacterium tuberculosis*. *Journal of Bacteriology*, 185(24), 7247–7256. <https://doi.org/10.1128/JB.185.24.7247-7256.2003>
- Vachée, A., Mossel, D., & Leclerc, H. (1997). Antimicrobial activity among *Pseudomonas* and related strains of mineral water origin. *Journal of Applied Microbiology*, 83(5), 652–658. <https://doi.org/10.1046/j.1365-2672.1997.00274.x>
- Van Beilen, J. B., Neuenschwander, M., Smits, T. H. M., Roth, C., Balada, S. B., & Witholt, B. (2002). Rubredoxins involved in alkane oxidation. *Journal of Bacteriology*, 184(6), 1722–1732. <https://doi.org/10.1128/JB.184.6.1722-1732.2002>
- Warburton, D. W. (1993). A review of the microbiological quality of bottled water sold in Canada. Part 2. The need for more stringent standards and regulations. *Canadian Journal of Microbiology*, 39(2), 158–168. <https://doi.org/10.1139/m93-022>
- Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., ... Wei, L. (2011). KOBAS 2.0: A web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Research*, 39, W316–W322. <https://doi.org/10.1093/nar/gkr483>
- Yoshida, K. I., Yamaguchi, M., Morinaga, T., Kinehara, M., Ikeuchi, M., Ashida, H., & Fujita, Y. (2008). *myo*-Inositol Catabolism in *Bacillus subtilis*. *Journal of Biological Chemistry*, 283(16), 10415–10424. <https://doi.org/10.1074/jbc.M708043200>
- Zarilla, K. A., & Perry, J. J. (1984). *Thermoleophilum album* gen. nov. and sp. nov., a bacterium obligate for thermophily and *n*-alkane substrates. *Archives of Microbiology*, 137(4), 286–290.

**How to cite this article:** Severino R, Froufe HJC, Barroso C, et al. High-quality draft genome sequence of *Gaiella occulta* isolated from a 150 meter deep mineral water borehole and comparison with the genome sequences of other deep-branching lineages of the phylum *Actinobacteria*. *MicrobiologyOpen*. 2019;8:e840. <https://doi.org/10.1002/mbo3.840>