



Article

Metagenomics Revealed a New Genus ‘*Candidatus Thiocaldithrix dubininis*’ gen. nov., sp. nov. and a New Species ‘*Candidatus Thiothrix putei*’ sp. nov. in the Family *Thiotrichaceae*, Some Members of Which Have Traits of Both Na⁺- and H⁺-Motive Energetics †

Nikolai V. Ravin ^{1,‡}, Maria S. Muntyan ^{2,*,‡}, Dmitry D. Smolyakov ³, Tatyana S. Rudenko ³, Alexey V. Beletsky ¹, Andrey V. Mardanov ¹ and Margarita Yu. Grabovich ^{3,*}

¹ Institute of Bioengineering, Research Center of Biotechnology, Russian Academy of Sciences, Leninsky Prospect, 33-2, 119071 Moscow, Russia; nravin@biengi.ac.ru (N.V.R.); mortu@yandex.ru (A.V.B.); mardanov@biengi.ac.ru (A.V.M.)

² Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Leninskie Gory, 119991 Moscow, Russia

³ Department of Biochemistry and Cell Physiology, Voronezh State University, Universitetskaya pl., 1, 394018 Voronezh, Russia; songolifreya@mail.ru (D.D.S.); ipigun6292@gmail.com (T.S.R.)

* Correspondence: muntyan@genebee.msu.ru (M.S.M.); margarita_grabov@mail.ru (M.Y.G.)

† The study is dedicated to the memory of the outstanding bioenergeticist Prof. V.P. Skulachev.

‡ These authors contributed equally to this work.



Citation: Ravin, N.V.; Muntyan, M.S.; Smolyakov, D.D.; Rudenko, T.S.; Beletsky, A.V.; Mardanov, A.V.; Grabovich, M.Y. Metagenomics Revealed a New Genus ‘*Candidatus Thiocaldithrix dubininis*’ gen. nov., sp. nov. and a New Species ‘*Candidatus Thiothrix putei*’ sp. nov. in the Family *Thiotrichaceae*, Some Members of Which Have Traits of Both Na⁺- and H⁺-Motive Energetics. *Int. J. Mol. Sci.* **2023**, *24*, 14199. <https://doi.org/10.3390/ijms241814199>

Academic Editor: Patrick C.Y. Woo

Received: 21 August 2023

Revised: 12 September 2023

Accepted: 14 September 2023

Published: 17 September 2023

Abstract: Two metagenome-assembled genomes (MAGs), GKL-01 and GKL-02, related to the family *Thiotrichaceae* have been assembled from the metagenome of bacterial mat obtained from a sulfide-rich thermal spring in the North Caucasus. Based on average amino acid identity (AAI) values and genome-based phylogeny, MAG GKL-01 represented a new genus within the *Thiotrichaceae* family. The GC content of the GKL-01 DNA (44%) differed significantly from that of other known members of the genus *Thiothrix* (50.1–55.6%). We proposed to assign GKL-01 to a new species and genus ‘*Candidatus Thiocaldithrix dubininis*’ gen. nov., sp. nov. GKL-01. The phylogenetic analysis and estimated distances between MAG GKL-02 and the genomes of the previously described species of the genus *Thiothrix* allowed assigning GKL-02 to a new species with the proposed name ‘*Candidatus Thiothrix putei*’ sp. nov. GKL-02 within the genus *Thiothrix*. Genome data first revealed the presence of both Na⁺-ATPases and H⁺-ATPases in several *Thiothrix* species. According to genomic analysis, bacteria GKL-01 and GKL-02 are metabolically versatile facultative aerobes capable of growing either chemolithoautotrophically or chemolithoheterotrophically in the presence of hydrogen sulfide and/or thiosulfate or chemoorganoheterotrophically.

Keywords: filamentous colorless sulfur bacteria; genome; metagenome-assembled genome; ‘*Candidatus Thiocaldithrix dubininis*’ gen. nov.; sp. nov.; ‘*Candidatus Thiothrix putei*’ sp. nov.; H⁺-motive ATPase; Na⁺-motive ATPase; Na⁺-motive PPase



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Before 2018, all the representatives of the filamentous colorless sulfur-oxidizing bacteria that form rosettes and accumulate elemental sulfur inclusions in their cells were united into the genus *Thiothrix* within the family *Thiotrichaceae*. In 2018, based on the sequence identity of 16S RNA and several other genes, Boden reclassified all the representatives of the genus *Thiothrix* into three families, *Thiotrichaceae*, *Thiolineaceae*, and *Thiofilaceae*, within the order *Thiotrichales*, and in addition to them included in this order another family *Leucotrichaceae* [1]. During that reclassification, two genera, *Leucothrix* and *Cocleimonas*, were assigned to the family *Leucotrichaceae*.

Somewhat later, after the symposium at FEMS-2019, it was decided that Bergey's Manual of Systematics of Archaea and Bacteria (BMSAB) would use the classification of the continuously updated Genome Taxonomy Database (GTDB) (<https://gtdb.ecogenomic.org/>, accessed on 28 June 2023) and its phylogenomic resource [2,3]. As a result, the order *Thiotrichales* (https://gtdb.ecogenomic.org/tree?r=o_Thiotrichales, accessed on 28 June 2023) includes only family *Thiotrichaceae* numbering seven genera (Figure 1), where five of them are validated genera, namely *Thiothrix*, *Thiolinea*, *Thiofilum*, *Leucothrix* and *Cocleimonas*, and two nonculturable candidate genera, HyVt-477 and S015-18 represented only by MAGs.

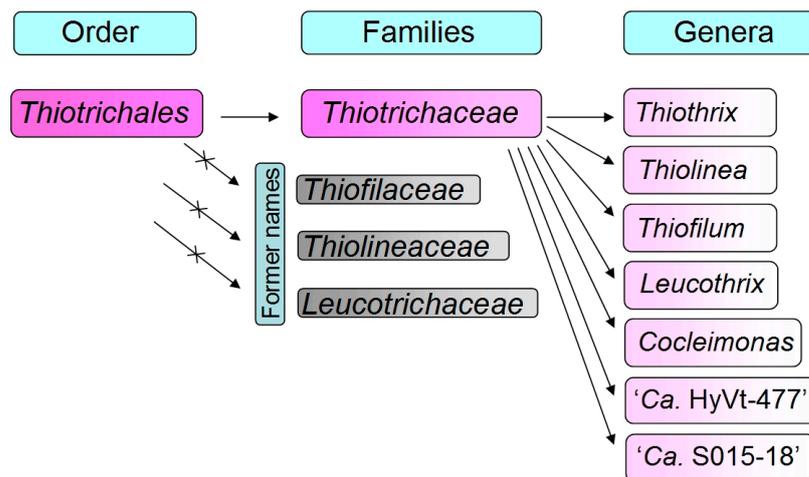


Figure 1. Reclassification of genera and families composing the order *Thiotrichales* according to genome taxonomic database (GTDB) (Release 08-RS214 (28 April 2023)).

Representatives of the first three genera are unified by the following common phenotype: they are filamentous colorless sulfur bacteria forming rosettes and subcellular inclusion of elemental sulfur under the growth in the presence of hydrogen sulfide and thiosulfate. Until now, their taxonomy is far from perfect. On the one hand, all the representatives of this group are microorganisms that are difficult to cultivate and, therefore, difficult to isolate as pure cultures. On the other hand, as a rule, all the representatives share a common pattern of major metabolic pathways. In general, all the above makes the reliable classification of new isolates noticeably more difficult.

At this stage of the studies, molecular phylogenetics brings significant contributions to the solution of the problem, while metagenomic analysis allowed the study of the communities of filamentous colorless sulfur bacteria by obtaining genomes of new representatives. These methods compensate for the limited number of species and genera as well as genome sequences in view of the lack of pure cultures.

Despite significant advances in the molecular phylogenetic classification of filamentous colorless sulfur bacteria (FCSB), the genera of the family *Thiotrichaceae* are still heterogeneous and include genetically unrelated species and genera. With numerous examples, genetic tools have been shown to differentiate FCSB species and genera. In our previous publications, we reported obtaining pure cultures and genomes assembled from metagenomes of sulfur biotopes, presenting them as new species of the genus *Thiothrix* [4]. These studies have noticeably extended the mentioned group of bacteria during the past several years, resulting in substantial changes in the taxonomy of this group [5–7]. Recently, we succeeded in assembling two genomes from a metagenome of microbial fouling from a hydrogen sulfide-rich thermal spring. In the presented study, phylogenetically, we assigned one of these genomes to a new genus within the family *Thiotrichaceae* and denoted it as '*Candidatus Thiocaldithrix dubininis*' gen. nov., sp. nov. GKL-01. The second genome we assigned to a new species within the genus *Thiothrix* and named '*Candidatus Thiothrix putei*' sp. nov. GKL-02.

As it was written earlier, members of the order *Thiotrichales* are often inhabitants of niches with microoxic conditions [6,8–12]. This environment is formed due to the high content of sulfur compounds with low redox potential. Among members of the order *Thiotrichales*, many species are also able to exist in both microoxic and anaerobic conditions. It is known that many bacteria capable of living under anaerobic conditions are often provided by energy-containing macroergic phosphates via the function of the Na^+ -motive energy-transforming pumps such as Na^+ -ATPases [13–16]. In the present study, we have elucidated which of the major energy transducers of the phosphate metabolism group could provide biological energy to members of this order. In the genomes of several species of the family *Thiotrichaceae*, we were able to identify Na^+ - and H^+ -pump genes, exemplified by energy-generating ATPases and PPases. Due to the availability of metagenomic data, which revealed a large number of new species within taxa of this order, it became possible to analyze a number of conditions and characteristics that could affect such a feature.

2. Results and Discussion

2.1. Metagenome and Genome Assembly

The specimen GKL for metagenome analysis was sampled from a sulfide mat in the zone of a sulfide hydrogen thermal spring in the North Caucasus ($44^{\circ}38' \text{ N}$, $39^{\circ}08' \text{ E}$), the town of Goryachiy Klyuch, Krasnodar region, Russia (Figure 2). Water characteristics in the sampling site were as follows: T 35°C , pH 7.5–8.0, total mineralization 1.56–1.65 g/L, oxygen concentration 0.5–1 mg/L, and sulfide concentration 0.13 mM.



Figure 2. Natural fouling with a predominance of representatives of the genus *Thiothrix* in the well of a natural hydrogen sulfide-rich thermal spring: general view of the well (**left panel**). The sampling site, highlighted in a magenta rectangle (**left panel**), is shown in $2.5 \times$ magnification (**right panel, top**). Fouling as white mats and rosettes on the surface of rocks and hydrogen sulfide-rich black slurry is selectively shown with magenta arrows (**right panel, top**). The mat from which the GKL metagenome was isolated (magenta arrow on the **left, right panel, top**) is shown in $9 \times$ magnification (**right panel, bottom**). Sulfide thermal spring in the town of Goryachiy Klyuch, Krasnodar region, Russia. Bar, 1 cm (**lower left corner, right panel, bottom**).

A total of 22,275 high-quality 16S rRNA gene sequences were determined to characterize the composition of the microbial mat at Goryachiy Klyuch spring. The microbial community was dominated by the representatives of several phyla: *Pseudomonadota* (former *Proteobacteria*, 60.8% of all 16S rRNA gene sequences including 60.1% of *Gammaproteobacteria*), *Campylobacterota* (16.2%), *Firmicutes* (5.6%), and *Bacteroidota* (4.5%). Representatives of the family *Thiotrichaceae* of the phylum *Pseudomonadota* accounted for 48.3% of the microbial mat community. Among them, two OTUs dominated with shares of 36.3% and 11.0%, which, according to the NCBI database, were identified as the representatives of the genus *Thiothrix*.

Two metagenome-assembled genomes (MAGs) were obtained from the GKL metagenome. *Thiotrichaceae* sp. GKL-01 and *Thiothrix* sp. GKL-02 MAGs were obtained as circular 3,251,546 b.p. and 4,277,058 b.p. long chromosomes, respectively. Both MAGs are 100% complete, according to CheckM2 estimates. The genome GKL-01 annotation revealed three copies of 16S-23S-5S rRNA operon, 49 tRNA genes, and 3021 genes encoding potential proteins. The GC content in the genome GKL-01 accounted for 44%. The genome GKL-02 annotation revealed three copies of 16S-23S-5S rRNA operon, 46 tRNA genes, and 4134 genes encoding potential proteins. GC content in the genome GKL-02 accounted for 50.08%. The 16S rRNA gene sequences of these two MAGs correspond to a pair of OTUs of the family *Thiotrichaceae* revealed in the microbial mat by 16S rRNA gene profiling. The main characteristics of the genomes are presented in Table 1, and further details according to the minimal standards for the description of MAGs [17] are provided in Supplementary Materials (Table S1).

Table 1. The general properties of genomes obtained from the metagenome of the microbial community of a sulfide thermal spring.

Genome (MAG)	Genome Assembly	Genome Size (MB)	GC-Content (%)	Genes		
				Protein-Coding	16S rRNA	tRNA
<i>Thiothrix</i> sp. GKL-02	GCA_029972225.1	4.28	50.8	4134	3	46
<i>Thiotrichaceae</i> sp. GKL-01	GCA_029972135.1	3.25	44.0	3021	3	49

2.2. Phylogenetic Analysis

A search against GTDB placed the bacterium GKL-01 in the family *Thiotrichaceae*, representing a novel genus. The average nucleotide identity (ANI) comparison values between MAG *Thiotrichaceae* sp. GKL-01 and other genomes of the genus *Thiothrix* accounted for 68.45–71.08%, which was lower than the threshold of 75% for the genera differentiation [18] (Figure 3). The average amino acid identity (AAI) comparison values between MAG GKL-01 and genomes of other members of the genus *Thiothrix* were in the range of 66.0–67.8%, which was slightly above the proposed threshold of 65% for the same genus [19] (Figure 4). MAG GKL-01 with AAI below 62% proved to be more distant from other *Thiotrichaceae* genera.

Phylogenetic position of MAG *Thiotrichaceae* sp. GKL-01, along with the previously described *Thiothrix* species, was analyzed by constructing a phylogenetic tree based on concatenated sequences of 120 conserved marker genes (Figure 5).

MAG *Thiotrichaceae* sp. GKL-01 formed a separate lineage at the root of the genus *Thiothrix*. Additionally, genome GC-content (44%), which meaningfully differs from the GC-content of other representatives of the genus *Thiothrix* (Figure 5), serves as a forcible argument for establishing the taxonomic position of MAG *Thiotrichaceae* sp. GKL-01 at the genus rank. Overall, the data obtained allowed us to identify MAG *Thiotrichaceae* sp. GKL-01 as a new species and genus within the family *Thiotrichaceae*, and we propose to name it '*Candidatus Thiocaldithrix dubininis*' gen. nov., sp. nov. GKL-01.

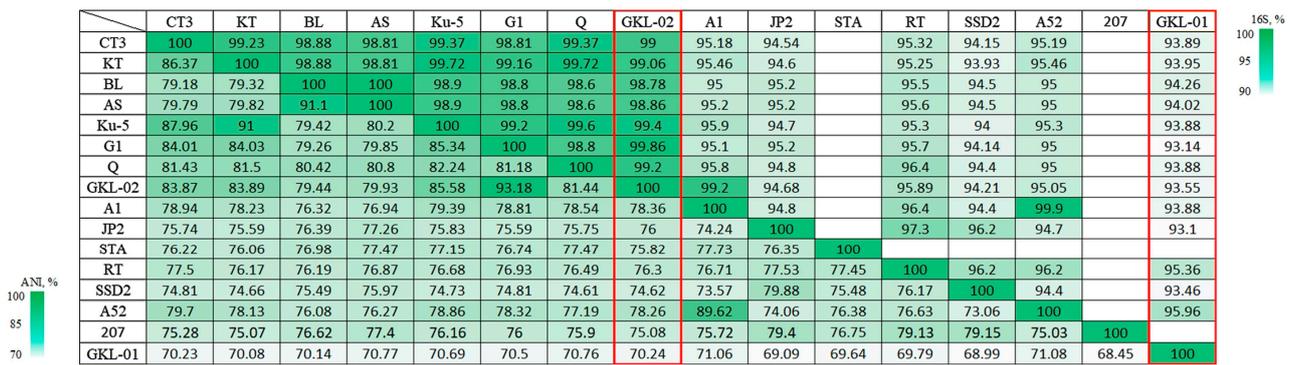


Figure 3. Heatmap of 16S rRNA gene sequence similarity and pairwise ANI values (%) for the assembled genomes of *Thiothrix*. *T. winogradskyi* CT3^T (GCA_021650935.1); ‘*Ca. Thiothrix sulfatifontis*’ KT (GCA_022828425.1); *T. lacustris* BL^T (GCF_000621325.1); *T. litoralis* AS^T (GCF_017901135.1); *T. subterranea* Ku-5^T (GCF_016772315.1); *T. caldifontis* G1^T (GCF_900107695.1); *Thiothrix* sp. STA_22 (GCA_028714775.1); *T. unzii* A1^T (GCA_017901175.1); *T. nivea* JP2^T (GCF_000260135.1); *T. fructosivorans* Q^T (GCA_017349355.1); *Ca. Thiothrix moscovensis* RT (GCA_016292235.1); *Ca. Thiothrix singaporensis* SSD2 (GCA_013693955.1); *Ca. Thiothrix anitrata* A52 (GCF_017901155.1); *Thiothrix* sp. 207 (GCA_018813855.1); MAG *Thiothrix* sp. GKL-02 (GCA_029972225.1); MAG *Thiotracheae* sp. GKL-01 (GCA_029972135.1). The red box indicates the new MAGs obtained in this work.

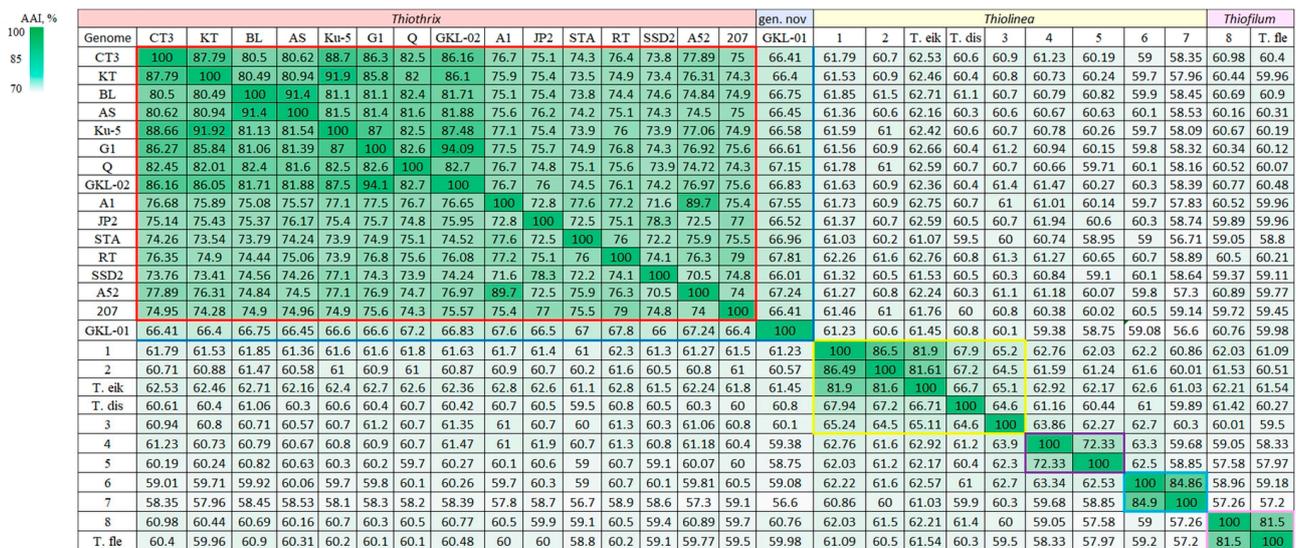


Figure 4. Heatmap of pairwise AAI values (%) for the assembled genomes of *Thiothrix*. *T. winogradskyi* CT3^T (GCA_021650935.1); ‘*Ca. Thiothrix sulfatifontis*’ KT (GCA_022828425.1); *T. lacustris* BL^T (GCF_0006

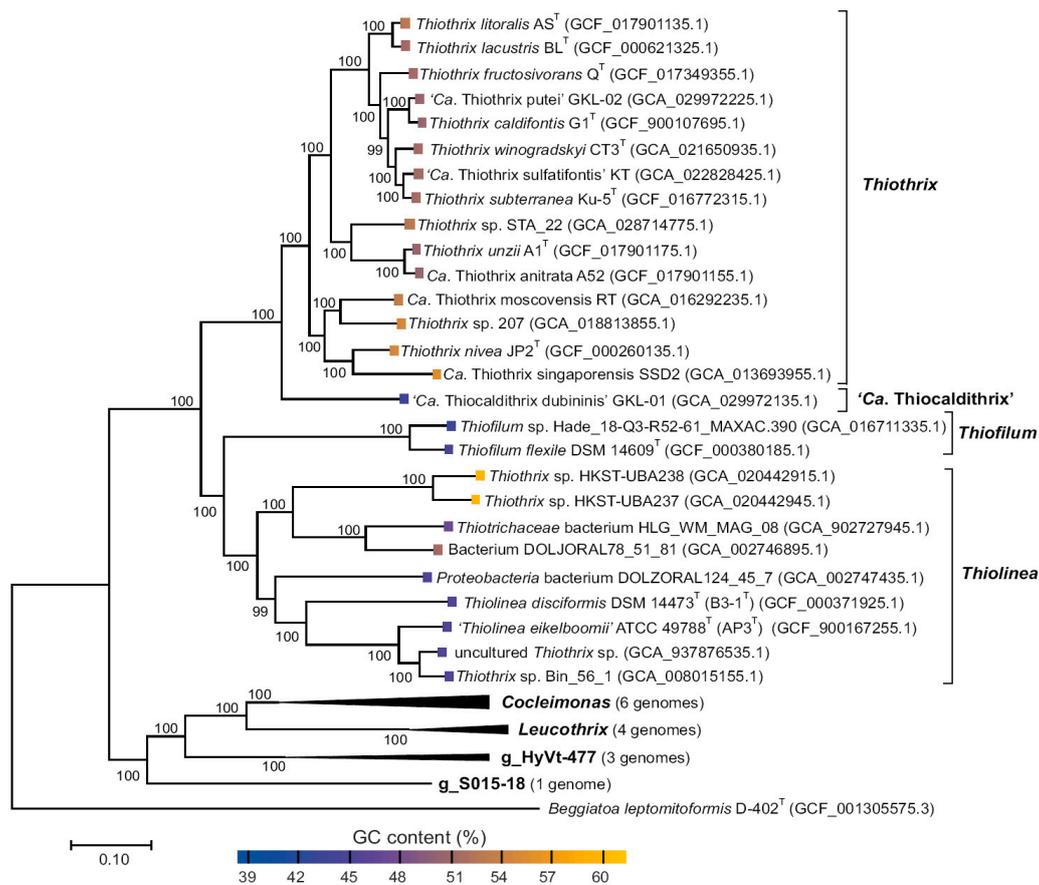


Figure 5. Genome-based phylogenetic tree of the family *Thiotrichaceae*. The genome positions were determined by the maximum likelihood method using concatenated sequences of 120 conserved marker genes. The GenBank assembly accession numbers are listed after the genome names. The genome of *Beggiatoa leptomitiformis* D-402^T (GCF_001305575.3) was used for tree rooting.

It was previously shown that the classical phylogenetic marker 16S rRNA is sometimes not suitable for reliable determination of the taxonomy position of representatives within the genus *Thiothrix* because 16S rRNA homology accounts for 93.7–100% within the genus *Thiothrix*. Thus, the 16S rRNA data obtained for MAG *Thiothrix* sp. GKL-02, when compared with *Thiothrix* representatives (94.21–99.86%), do not allow unambiguously attribute it to one of *Thiothrix* species (Figure 3). However, the pairwise ANI and dDDH comparison values between MAG *Thiothrix* sp. GKL-02 and genomes of the earlier described species of the genus *Thiothrix* ranged in the intervals that lay between 74.6–93.2% and 22.7–57.3%, respectively (Figure 3 and Figure S1). The obtained distances between the genomes allowed us to assign GKL-02 to a new species of the genus *Thiothrix* and propose the name ‘*Candidatus Thiothrix putei*’ sp. nov. GKL-02. The branching topology of the tree based on 120 conserved protein genes suggests that *T. caldifontis* G1^T is the closest relative of GKL-02 (Figure 5).

The representatives of the genera *Thiolinea* and *Thiofilum*, phylogenetically and phenotypically related to the genus *Thiothrix*, belong to FCSB. Analysis of genome sequences that were attributed to the genus *Thiolinea* according to GTDB showed that AAI indexes (less than 65%) suggested dividing this genus into three genera (Figure 4): 1—genus *Thiolinea* inclusive uncultured *Thiothrix* sp. (GCA_937876535.1), *Thiothrix* sp. Bin_56_1 (GCA_008015155.1), ‘*Thiolinea eikelboomii*’ AP3^T (GCA_900167255.1), *Thiolinea disciformis* B3-1^T (GCA_000371925.1), *Proteobacteria* bacterium DOLZORAL124_45_7 (GCA_002747435.1); 2—a new putative genus inclusive *Bacterium* DOLJORAL78_51_81 (GCA_002746895.1) and *Thiotrichaceae* bacterium HLG_WM_MAG_08 (GCA_902727945.1); 3—a new putative genus

inclusive *Thiothrix* sp. HKST-UBA238 (GCA_020442915.1) and *Thiothrix* sp. HKST-UBA237 (GCA_020442945.1).

2.3. Genome Analysis of ‘*Ca. Thiothrix putei*’ sp. nov. GKL-02 and ‘*Ca. Thiocaldithrix dubininis*’ gen. nov., sp. nov. GKL-01

Analysis of ‘*Ca. Thiothrix putei*’ sp. nov. GKL-02 genome showed the presence of a number of genes of dissimilatory sulfur metabolism, which is in accordance with the data for the members of filamentous colorless sulfur bacteria of the genus *Thiothrix* [20] (Table 2). Genomes of all members of the genus *Thiothrix*, including ‘*Ca. Thiothrix putei*’ GKL-02 contain genes of the branched SOX-system (*soxAXBYZ*) of oxidation thiosulfate to sulfur and sulfate; genes of the system of oxidation hydrogen sulfide to elemental sulfur *sqr*, sulfide:quinone oxidoreductase (*sqrF*, *sqrA*) and FCSD, flavocytochrome sulfide dehydrogenase (*fccAB*); genes of rDSR-complex (*dsrABEFHNEMKLJONR*) for oxidation sulfur to sulfite; genes of quinone-dependent sulfite dehydrogenase (*soeABC*) for direct oxidation of sulfite to sulfate and of ATP-sulfurylase, dissimilatory-type (*sat*), and APS-reductase (*aprAB*) for indirect oxidation.

Table 2. Characterization of the main metabolic pathways of the genera of the family *Thiotrichaceae*. *sqr*, sulfide: quinone oxidoreductase; *fccAB*, flavocytochrome sulfide dehydrogenase; *soxAXBYZ*, the branched SOX-system; rDsr, complex for sulfur oxidation *dsrABEFHNEMKLJONR*; *soeABC*, quinone-dependent sulfite dehydrogenase; *aprAB*, adenosine 5'-phosphosulfate reductase; *sat*, ATP-sulfurylase, dissimilatory-type; *phsABC*—thiosulfate reductase; *narG*, membrane-bound nitrate reductase; *napAB*, periplasmic nitrate reductase; *nasA*, assimilatory nitrate reductase; *nasB*—assimilatory nitrite reductase; *nirS*—dissimilatory nitrite reductase; *nirBD*—assimilatory nitrite reductase; *norBC*, nitric oxide reductase; *nosZ*, nitrous oxide reductase; *nif* gene cluster, nitrogenase genes; *mgo*, malate quinone oxidoreductase; *prk*, phosphoribulokinase; *hypABCDEF*, hydrogenase; SOD2, superoxide dismutase [Fe]; SOD1, superoxide dismutase, Cu-Zn family; *katG*, catalase/ peroxidase HPI; PRDX5, peroxiredoxin 5; *katE*, catalase; *ccp*, cytochrome c₅₅₁ peroxidase.

Metabolism	Characteristics	<i>Thiothrix</i>	‘ <i>Ca</i> <i>Thiocaldithrix</i> ’	<i>Thiolinea</i>		<i>Thiofilum Flexile</i>	
				<i>Thiolinea</i> <i>disciformis</i>	‘ <i>Thiolinea</i> <i>eikelboomii</i> ’		
Sulfur metabolism	Hydrogen sulfide oxidation systems	<i>sqrA, sqrF, fccAB</i>	<i>sqrA, sqrF, fccAB</i>	<i>sqrA, sqrF, fccAB</i>		<i>sqrA, sqrF, fccA</i>	
	SOX-system	<i>soxAXBYZ</i>	<i>soxAXBYZ</i>	<i>soxAXBYZ</i>		<i>soxAXBYZ</i>	
	Elemental sulfur oxidation systems-rDSR	<i>dsrABEFHNEM</i> <i>KLJONR</i>	<i>dsrABEFHNEM</i> <i>KLJONR</i>	<i>dsrABEFHNEM</i> <i>KLJONR</i>		<i>dsrABEFHNEM</i> <i>KLJONR</i>	
	Sulfite oxidation systems	Direct way <i>soeABC</i> Indirect way <i>sat, aprAB</i>	<i>soeABC</i> —	<i>soeABC</i> —	<i>phsABC</i> —	— <i>phsABC</i>	<i>soeABC</i> —
Nitrogen metabolism	Thiosulfate Reductase	<i>phsABC</i>	<i>phsABC</i>	<i>phsABC</i>	—	—	
	Dissimilatory nitrate reduction	$\text{NO}_3^- \rightarrow \text{NO}_2^-$	<i>narGHI, napAB</i> **	—	—	<i>narGHI</i>	<i>narGI</i>
		$\text{NO}_2^- \rightarrow \text{NH}_3$	<i>nirBD</i> **	—	—	—	<i>nirBD</i>
		$\text{NO}_2^- \rightarrow \text{NO}$	<i>nirS</i> **	—	—	—	—
		$\text{NO} \rightarrow \text{N}_2\text{O}$	<i>norBC</i> **	—	—	—	—
$\text{N}_2\text{O} \rightarrow \text{N}_2$	<i>nosZ</i> **	—	—	—	—	—	
Nitrogen fixation	<i>nifASUBNXX2YB</i> <i>2ENQVWMHDKZTO</i> **	—	—	—	—	—	
Carbon metabolism	Assimilatory nitrate reduction	<i>nasA, nasD</i> **	—	—	—	<i>nasA, nasD</i>	
	Type of malate dehydrogenase	<i>mgo</i>	<i>mgo</i>	<i>mgo</i>		<i>mgo, mdh</i>	
	Krebs cycle	+	+	+		+	
	Glyoxylate pathway	+	+	+		+	
	Pentose-phosphate pathway	+	+	+		+	
	Type of RuBisCo	IAq, IAc, II **	II	II	IAq	—	
AOS-system	The type of phosphoribulokinase	<i>prk1, prk2</i>	<i>prk1, prk2</i>	<i>prk1, prk2</i>		<i>prk1</i>	
	Hydrogenase	<i>hypABCDEF</i>	<i>hypABCDEF</i>	<i>hypABCDEF</i>		<i>hypABCDEF</i>	
	<i>sod2</i>	+	+	+		+	
	<i>sod1</i>	—	—	+		—	
	<i>prdx5</i>	+	+	+		—	
	<i>katG</i>	+ / - *	—	+		+	
	<i>katE</i>	—	+	+		—	
	<i>ccp</i>	+	+	+		+	
<i>gpx</i>	+	+	+		—		

*—*katG* is present in the genomes of all *Thiothrix* except for *T. unzii* A1^T, *Ca. Thiothrix moscovensis* RT and *Thiothrix* sp. 207. **—present in some species (see Table 3). The plus “+” and minus “-” symbols indicate presence or absence of a given gene(s).

Table 3. Distribution of genes for nitrogen metabolism and autotrophic carbon fixation in *Thiothrix* genomes. Genes: *narG*, membrane-bound nitrate reductase; *napAB*, periplasmic nitrate reductase; *nasA*, assimilatory nitrate reductase; *nasB*, assimilatory nitrite reductase; *nirS*, dissimilatory nitrite reductase; *nirBD*, assimilatory nitrite reductase; *norBC*, nitric oxide reductase; *nosZ*, nitrous oxide reductase; *nif* gene cluster, nitrogenase genes.

Strain	Type of RuBisCO	Dissimilatory Nitrate Reduction					Assimilatory Nitrate Reduction	N ₂ Fixation (<i>nif</i>)
		NO ₃ ⁻ → NO ₂ ⁻	NO ₂ ⁻ → NH ₃	NO ₂ ⁻ → NO	NO → N ₂ O	N ₂ O → N ₂		
<i>Thiothrix</i> sp. 207	IAq, IAc, II	<i>narGHI</i> , <i>napAB</i>	<i>nirBD</i>	<i>nirS</i>	<i>norBC</i>	<i>nosZ</i>	<i>nasA</i> , <i>nasD</i>	+
<i>Thiothrix</i> sp. STA_22	II	<i>narGHI</i>	<i>nirBD</i>	<i>nirS</i>	<i>norBC</i>	<i>nosZ</i>	<i>nasA</i> , <i>nasD</i>	+
<i>T. litoralis</i> AS ^T	IAq, IAc, II	<i>narGHI</i>	<i>nirBD</i>	<i>nirS</i>	<i>norBC</i>	–	<i>nasA</i> , <i>nasD</i>	+
<i>T. fructosivorans</i> Q ^T	IAq, II	<i>narGHI</i>	<i>nirBD</i>	<i>nirS</i>	<i>norBC</i>	–	<i>nasA</i> , <i>nasD</i>	–
<i>T. caldifontis</i> G1 ^T	IAq, IAc, II	<i>narGHI</i>	<i>nirBD</i>	<i>nirS</i>	<i>norBC</i>	–	<i>nasA</i> , <i>nasD</i>	+
<i>T. winogradskyi</i> CT3 ^T	IAc, II	<i>narGHI</i>	<i>nirBD</i>	<i>nirS</i>	<i>norBC</i>	–	<i>nasA</i> , <i>nasD</i>	–
<i>T. unzii</i> A1 ^T	IAq, IAc, II	<i>narGHI</i>	<i>nirBD</i>	<i>nirS</i>	<i>norBC</i>	–	<i>nasA</i> , <i>nasD</i>	+
<i>Ca. Thiothrix</i> singaporensis SSD2	IAq, II	<i>narGHI</i>	<i>nirBD</i>	<i>nirS</i>	<i>norBC</i>	–	<i>nasA</i> , <i>nasD</i>	+
<i>Ca. Thiothrix</i> moscovensis RT	IAq, IAc, II	<i>narGHI</i>	<i>nirBD</i>	–	<i>norBC</i>	–	<i>nasA</i> , <i>nasD</i>	+
' <i>Ca. Thiothrix putei</i> ' GKL-02	IAq, IAc	<i>narGHI</i>	–	–	<i>norBC</i>	–	–	–
<i>T. subterranea</i> Ku-5 ^T	IAq, IAc, II	<i>narGHI</i>	<i>nirBD</i>	–	–	–	<i>nasA</i> , <i>nasD</i>	+
<i>T. lacustris</i> BL ^T	IAq, II	<i>narGHI</i>	<i>nirBD</i>	–	–	–	<i>nasA</i> , <i>nasD</i>	–
' <i>Ca. Thiothrix</i> sulfatfontis' KT	IAq, IAc, II	–	<i>nirBD</i>	–	–	–	<i>nasA</i> , <i>nasD</i>	–
<i>T. nivea</i> JP2 ^T	IAq, IAc, II	<i>napAB</i>	<i>nirBD</i>	–	–	–	<i>nasA</i> , <i>nasD</i>	–
<i>Ca. Thiothrix anitrata</i> A52	IAq, IAc	–	–	–	–	–	–	–

High variability in the composition of nitrogen metabolism genes in members of the genus *Thiothrix* was previously determined in the gene sets of dissimilatory and assimilatory nitrate reduction as well as nitrogen fixation [6,7,20] (Table 3).

In the genome of '*Ca. Thiothrix putei*' GKL-02, we have observed only the genes of the membrane-bound nitrate reductase *narGHI*, participating in the dissimilatory reduction of nitrogen, and genes of nitric oxide reductase *norBC*, catalyzing the reduction of NO to N₂O that suggests the incomplete denitrification. Moreover, we have found in this genome the genes responsible for maturation of the nitrogenase complex (*nifASUVNWM*), while genes encoding catalytic subunits have not been identified (Table 3).

Genes of assimilatory nitrate reduction (*nasA*, *nasD*) are absent from the genome of '*Ca. Thiothrix putei*' GKL-02, similar to the genome of *Ca. Thiothrix anitrata* A52, but the classical amination genes (*glnB*, *gltBD*, *aspB*) were detected.

Similar to all members of the genus *Thiothrix*, the genome of '*Ca. Thiothrix putei*' GKL-02 harbors genes for carbon dioxide autotrophic assimilation in the Calvin–Benson–Bassham cycle. Of RuBisCO types IAq, IAc, and II encountered in the genus [20], types IAq and IAc were found in the GKL-02 genome (Table 3).

Similar to the earlier described thirteen members of the genus *Thiothrix*, in '*Ca. Thiothrix putei*' GKL-02, two phylogenetically distant copies of phosphoribulokinase were detected.

All genes of the Krebs cycle, oxidative pentose phosphate pathway, and glyoxylate shunt were found in the GKL-02 genome. The respiratory type of metabolism is determined by the presence of electron-transport chain genes (Table S1). Similar to that in all members of the genus *Thiothrix*, in the genome of GKL-02, the genes of FAD-dependent membrane-bound malate quinone oxidoreductase (*mqr*) are present instead of NAD-dependent malate dehydrogenase.

Phosphorus metabolism in '*Ca. Thiothrix putei*' GKL-02 is presented in a similar way as in the previously described members of the genus *Thiothrix* by the systems of gene expression regulation (*phoURB*) of inorganic phosphorus transmembrane transport and its concentration (*pstSACB*), formation, and hydrolysis of phosphodiester bonds by polyphosphate kinase and exopolyphosphatase (*ppk1*, *epp*).

The genome of '*Ca. Thiocaldithrix dubininis*' gen. nov., sp. nov. GKL-01 was analyzed in comparison with the previously described closely related genera, *Thiothrix*, *Thiolinea*, and *Thiofilum* (Table 2). The sulfur oxidation systems are the most conserved part of genomes in all four genera of the family *Thiotrichaceae*. The genes of oxidation of hydrogen sulfide to elemental sulfur (*sqrF*, *sqrA*, *fccAB*), elemental sulfur to sulfite (rDSR), direct oxidation of sulfite to sulfate (*soeABC*), as well as branched SOX-system (*soxAXBYZ*), were identified in all known members of the four genera, which argues for the close evolution in the systems with the high content of reduced sulfur compounds.

However, in contrast to the genus *Thiothrix*, '*Ca. Thiocaldithrix dubininis*' GKL-01, similar to the representatives of the genera *Thiolinea* and *Thiofilum*, has no genes for indirect sulfite oxidation (*sat*, *aprAB*). Comparative analysis of nitrogen metabolism genes in '*Ca. Thiocaldithrix dubininis*' GKL-01 and the closely related genus *Thiothrix* disclosed significant differences. Nitrogen metabolism in the GKL-01 bacterium is presented only by genes of classical amination (*glnB*, *gltBD*, *aspB*), whereas in representatives of the genus *Thiothrix*, there are genes of denitrification, assimilatory nitrate reduction, and nitrogen fixation. The profiles of nitrogen metabolism gene composition in '*Ca. Thiocaldithrix dubininis*' GKL-01 and genera *Thiolinea* and *Thiofilum* are close, all lacking genes for denitrification and nitrogen fixation—the composition of RuBisCo types in '*Ca. Thiocaldithrix dubininis*' GKL-01 (type II) is more close to genera *Thiolinea* and *Thiofilum*. It was found that the GKL-01 genome contains a catalase gene (*katE*) present in *Thiolinea* but absent in the representatives of genera *Thiothrix* and *Thiofilum*, but lacks *katG*, which is present in the representatives of *Thiolinea* and *Thiofilum*, and in most *Thiothrix* genomes.

2.4. Energy Converter Genes from the Phosphate Metabolism Group

The data presented in Tables 3 and 4 show that members of the family *Thiotrichaceae* are characterized by a wide range of metabolic capabilities [4]. This raises interest in their energy conversion systems. To date, energy-converting mechanisms operating on both hydrogen and sodium ions are known in the bacterial world [21–23]. While H⁺-motive energy converters have been known since the publication of P. Mitchell's theory [24], the first data on a wide variety of Na⁺-motive mechanisms, including Na⁺-decarboxylase [25], Na⁺-NADH-quinone oxidoreductase [26], Na⁺-ATPase [27], Na⁺-pyrophosphatase (PPase) [28], Na⁺-proteorhodopsin [29], and Na⁺-*cbb*₃ cytochrome oxidase [30,31], appeared much later.

As to the present study, we have confined it to the search for energy converter genes from the phosphate metabolism group. We have found that genomes of a number of *Thiotrichaceae* family members bear not only genes of operon *atpIBEFHAGDC* coding H⁺-motive F-ATPases but also genes of another operon *atpDCQRBEFG* coding potential Na⁺-motive F-ATPases which were previously proposed to be termed N-ATPases and referred to the N-subfamily of F-ATPases [32]. These ATPases form a distinct branch reliably separated from the cluster of H⁺-motive F-ATPases (Figure 6).

The new MAGs, '*Ca. Thiocaldithrix dubininis*' gen. nov., sp. nov. GKL-01, and '*Ca. Thiothrix putei*' sp. nov. GKL-02 appeared to bear only genes of the H⁺-motive F-ATPase. In addition, according to the phylogeny based on the amino acid sequences of the c-subunit of F-ATPase, the two new species are in separate clusters (Figure 6) and are distantly related, which favors the results of the genome-based phylogeny (Figure 5). In the c-subunit of N-ATPases of the *Thiotrichaceae* members, we found the following potential ligands of Na⁺ ions: Glu-32, Met-63, Glu-65, Ser-66, and Tyr-70 (numbering in c-subunit of *Ilyobacter tartaricus* Na⁺-ATPase), which is close to consensus sequence with a set of Na⁺-ligands in well-known Na⁺-motive F-ATPases of *Ilyobacter tartaricus* [33,34], *Propionigenium modestum* [27,35] and *Thermotoga maritima* [34,36].

Table 4. Interrelation of the physiological and ecological characteristics of the members of the family *Thiotrichaceae* and the presence of energy-coupling ATPase and membrane-bound PPase genes. Species bearing genes of Na⁺-pumps involved in phosphate metabolism are marked in pink color.

Genus	Strain	Biotope Conditions	Anaerobic Respiration		PPase, <i>hppA</i>		F ₀ F ₁ -ATPase	
			Thiosulfate Respiration (e ⁻ Acceptor)	Dissimilatory Nitrate Reduction	Coupling Ion			
			Na ⁺	H ⁺	Na ⁺	H ⁺	Na ⁺	H ⁺
	<i>Thiothrix</i> sp. STA_22	Unknown	–	+	–	+	+	+
	<i>T. litoralis</i> AS ^T	White Sea littoral, T 9–12 °C	+	+	–	+	+	+
	<i>T. fructosivorans</i> Q ^T	Activated sludge, T 4–28 °C	+	+	–	+	+	+
	<i>T. winogradskyi</i> CT3 ^T	Activated sludge, T 20–24 °C	–	+	–	+	+	+
	<i>T. unzii</i> A1 ^T	Active sludge of waste treatment facilities, T 20–28 °C	+	+	–	+	+	+
	<i>T. lacustris</i> BL ^T	Freshwater lake, T 9.3 °C	+	+	–	+	+	+
	<i>Ca. Thiothrix moscovensis</i> RT	Phosphorus removal bioreactor, T 20–24 °C	–	+	+	–	–	+
	<i>Thiothrix</i> sp. 207	Underground water, T 15 °C	+	+	–	+	–	+
<i>Thiothrix</i>	<i>T. caldifontis</i> G1 ^T	Hydrogen sulfide thermal spring, T 35–40 °C	+	+	–	+	–	+
	<i>Ca. Thiothrix singaporensis</i> SSD2	Phosphorus removal bioreactor, T 20–24 °C	+	+	–	+	–	+
	' <i>Ca. Thiothrix putei</i> ' GKL-02	Hydrogen sulfide thermal spring, T 35–40 °C	–	+	–	+	–	+
	<i>T. subterranea</i> Ku-5 ^T	Spout from the shaft, T 15 °C	+	+	–	+	–	+
	' <i>Ca. Thiothrix sulfatfontis</i> ' KT	Hydrogen sulfide spring, T 15–18 °C	+	–	–	+	–	+
	<i>T. nivea</i> JP2 ^T	Sulfide-containing well water, T 16 °C	+	+	–	+	–	+
	<i>Ca. Thiothrix anitrata</i> A52	Hydrogen sulfide spring, T 18 °C	+	–	–	+	–	+
' <i>Ca. Thiocaldithrix</i> '	' <i>Ca. Thiocaldithrix dubininis</i> ' GKL-01	Hydrogen sulfide thermal spring, T 35–40 °C	+	–	–	–	–	+
<i>Thiolinea</i>	' <i>Thiolinea eikelboomii</i> ' AP3 ^T	Activated sludge, T 13–15 °C	–	+	–	–	–	+
	<i>Thiolinea disciformis</i> B3-1 ^T	Activated sludge, T 14 °C	+	–	–	–	–	+
<i>Thiofilum</i>	<i>Thiofilum flexilie</i> EJ2M-B ^T	Activated sludge, T 10–14 °C	–	–	–	–	–	+
	<i>Leucothrix mucor</i> DSM 2157 ^T	Shoreline of the harbor, T 28 °C	–	–	–	+	+	+
<i>Leucothrix</i>	<i>Leucothrix arctica</i> IMCC9719 ^T	Arctic seawater, T 15 °C	–	–	–	–	+	+
	<i>Leucothrix pacifica</i> XH122 ^T	Surface water, T 28 °C	–	–	–	–	+	+
<i>Cocleimonas</i>	<i>Cocleimonas flava</i> KMM 3898 ^T	Internal tissue of a marine mollusk, T 25–30 °C	–	–	+	–	+	+
<i>Beggiatoa</i>	<i>Beggiatoa leptomitiformis</i> D-402 ^T	Freshwater spring, T 8–35 °C	–	–	–	+	–	+
	<i>Beggiatoa alba</i> B18LD ^T	Rice paddy, T 0–38 °C	–	–	–	+	–	+

*—Despite the presence of relevant genes (Table 2), anaerobic growth has not been experimentally documented;

‡—The genome cluster is incomplete and is represented only by *narGI*.

The N-ATPases of *Thiotrichaceae* representatives differ slightly from these F-ATPases in that the nonpolar Val-63 is replaced by nonpolar Met-63, and the polar Thr-67 by nonpolar Ile/Val-67. It should be noted here that Thr-67 apparently does not participate in the formation of the Na⁺-binding coordination sphere, and the above replacement is irrelevant. This follows from cryo-electron and atomic force microscopy as well as X-ray crystallography data, according to which the Na⁺-binding coordination sphere in the Na⁺-translocating membrane-bound c-subunit of the Na⁺-motive F-ATPase of *I. tartaricus* is formed by Gln-32, Val-63, Glu-65, Ser-66, and Tyr-70 [34,37]. Such a coordination sphere gives a total of six coordination bonds sufficient to bind Na⁺ ions. Moreover, the replacement of Val-63 by Met-63 evidently is not significant since, according to the X-ray crystallographic data, the ligand of the Na⁺ coordination sphere donated from Val-63 is the backbone carbonyl oxygen [34].

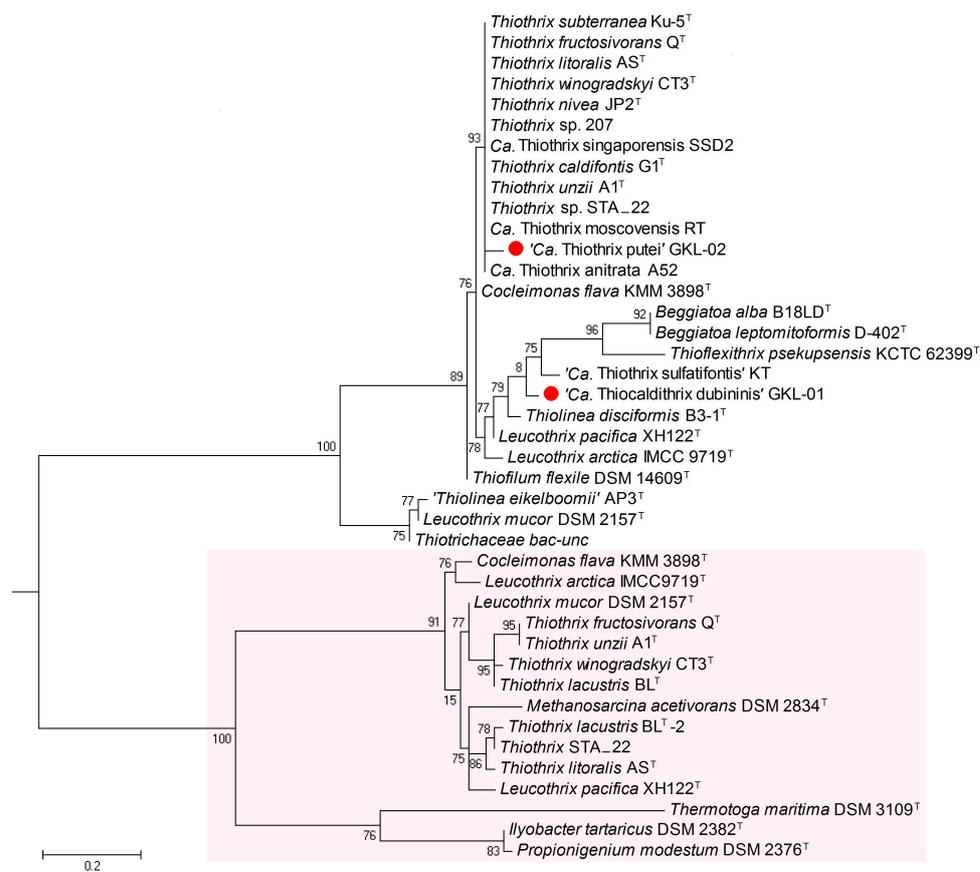


Figure 6. Maximum-likelihood tree based on the predicted amino acid sequences of F₀F₁-ATPase c-subunit mainly of the closest representatives of the family *Thiotrichaceae* and some other representatives of the order *Thiotrichales*. An unrooted tree with the highest log likelihood is shown. The bootstrap values (100 replicates) are shown next to the branches. The sequences were obtained from NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>), RAST server (<https://rast.nmpdr.org/rast.cgi>) and UniProtKB (<https://www.uniprot.org/uniprotkb/>). Two large clusters presenting two ATPase subfamilies are denoted: the prospective Na⁺-ATPases (pink-colored rectangle at the bottom) and H⁺-motive F₀F₁ATPases (uncolored clusters). Amino acid sequences of c-subunits of *Propionigenium modestum*, *Ilyobacter tartaricus*, *Thermotoga maritima*, and *Methanosarcina acetivorans* were used as proven Na⁺-ATPases. The newly proposed species '*Ca. Thiostrix putei*' GKL-02 and the new species of a new genus '*Ca. Thiocaldithrix dubininis*' GKL-01 in the family *Thiotrichaceae* are indicated by small red balls. The bar shows the scale of branch length in the number of substitutions per site. The gene accession numbers are listed in Supplementary Materials (Text S1).

Notably, modeling of the ion-coordination structure of the c-subunits of ATP synthase from *Methanosarcina acetivorans* demonstrates that even the presence of only 3 instead of 4–5 typical polar amino acid residues characteristic of tight Na⁺-binding in the Na⁺-coordination sphere of the c-subunit in *I. tartaricus* and its homolog in the *P. modestum* enzyme [27,35], is sufficient to display the Na⁺-motive behavior of *M. acetivorans* ATPase [38]. Moreover, according to the free energy computation, the selectivity of c-subunit for Na⁺ over H⁺ in the range of ATPases of several microorganisms smoothly decreases in compliance with the decrease in the number of specific ligands in the potential Na⁺-coordination sphere [38]. Thus, it can be assumed that a part of representatives of the family *Thiotrichaceae*, which forms a cluster, reliably separated on the phylogenetic tree from owners of H⁺-motive F-ATPases (marked in pink, Figure 6), possess Na⁺-motive F-ATPases.

The reasons for the appearance of duplicate ATPases with different ion specificity in the genomes of a number of bacteria are not entirely clear. For a number of bacteria, it has

been shown that GC content can reflect the adaptation of a species to certain environmental conditions; for example, an increase in GC correlates with an increase in the thermostability of bacteria [39]. Thus, the GC content could serve as a marker of the conditions and adaptive capabilities of a species. With the availability of a large number of genomes in the family *Thiotrichaceae*, an opportunity has arisen to use the GC index to test the correlation between the appearance of Na⁺-ATPases and the content of GC in the genome. As seen from Figure 7, the mean value of GC is the same in species that have only H⁺-motive F-ATPase genes and in species that have in their genomes both types of F-ATPases, H⁺-motive and Na⁺-motive.

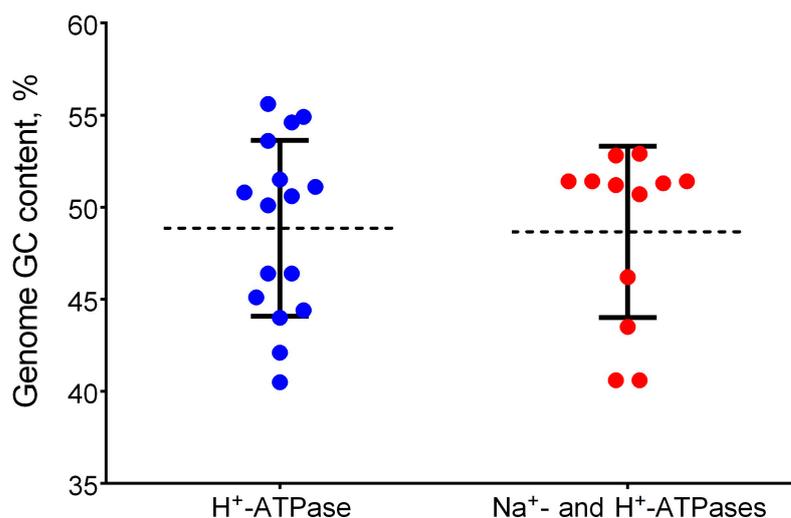


Figure 7. Genomic GC distribution depending on the presence or absence of Na⁺-ATPase genes in species of the order *Thiotrichales*. The mode of repeated measures one-way ANOVA data was used. Closed circles show genomic GC content of the only H⁺-ATPase-containing species (blue) and the species containing both H⁺- and Na⁺-ATPases (red); mean (dashed line, black) with SD (bars, black).

These data show the lack of correlation between the appearance of Na⁺-ATPases and GC content in the genome and mediate the probable absence of influence of factors correlating with changes in the GC content. Moreover, the presence or absence of genes of dissimilatory nitrate and thiosulfate reduction are not related to the appearance of sodium-type energy converters (Table 4). Interestingly, according to Dibrova et al. [32], there is a correlation between the GC content of N-ATPase operons and the genomes of their hosts in representatives of different orders. If we take into account the assumption of these authors about the spread of N-ATPase between organisms by horizontal transfer of the operon, then it is very likely that it must have taken quite a long time to match the GC profile of the host. If these assumptions are correct, then N-ATPases may be a very ancient acquisition of organisms.

We found that a number of *Thiotrichales* species have genes not only for Na⁺-dependent F-ATPase but also genes for other energy converters of the phosphate metabolism group, namely, genes for Na⁺-translocating membrane PPases. The membrane PPases we consider here, unlike membrane ATPases, are encoded by only one *hppA* gene and are capable of energizing membranes during pyrophosphate hydrolysis. Similar to membrane ATPases, these PPases are electrogenic reversible enzymes. In the genomes of most members of the family *Thiotrichaceae*, including the newly proposed species '*Ca. Thiothrix putei*' GKL-02, but not in '*Ca. Thiocaldithrix dubininis*' GKL-01, we found the *hppA* gene encoding a membrane K⁺-independent H⁺-motive PPase (Figure 8).

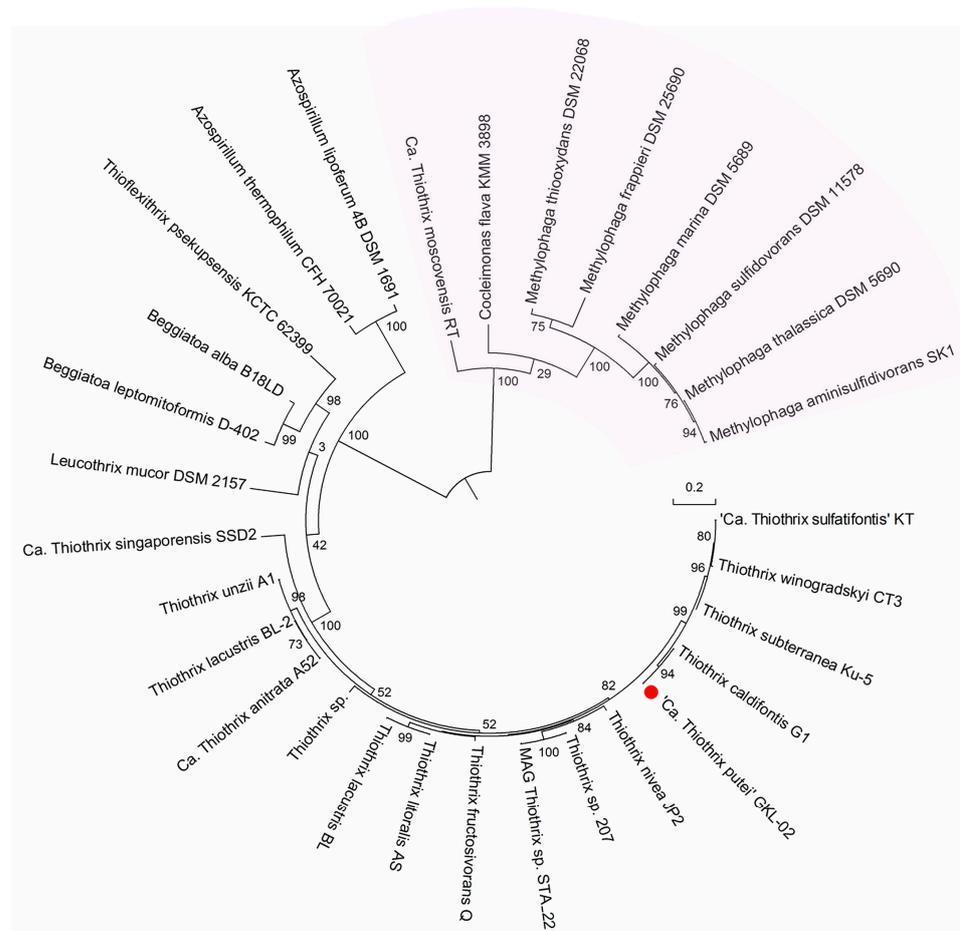


Figure 8. Maximum-likelihood tree based on the predicted amino acid sequences of membrane-bound PPases (*hpaA*) of close members of the order *Thiotrichales* and two species of *Azospirillum*. The bootstrap values (100 replicates) are shown next to the branches of the unrooted tree. The sequences were obtained from NCBI (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>), RAST server (<https://rast.nmpdr.org/rast.cgi>) and UniProtKB (<https://www.uniprot.org/uniprotkb/>). Two large clusters present two membrane-bound PPase subfamilies: Na⁺-PPases (pink-colored sector) and H⁺-PPases (uncolored sector). The newly proposed species ‘*Ca. Thiothrix putei*’ GKL-02 in the genus *Thiothrix* is indicated by a small red ball. The bar shows the scale of branch length in the number of substitutions per site. The gene accession numbers are listed in the Supplementary Materials (Text S2).

In contrast to K⁺-independent PPase, membrane-bound K⁺-dependent PPase, which is encoded by the same gene with a single specific substitution, is a Na⁺-pump. Such a membrane-bound K⁺-dependent Na⁺-motive PPase, similar to that demonstrated in anaerobic archaea [28,40], is found in the genomes of some representatives of *Thiotrichaceae*, including *Ca. Thiothrix muscovensis* RT. Interestingly, in the *Thiotrichaceae* species that we analyzed, we did not detect the co-presence of PPase genes with different ionic specificity similar to what we observed with F-ATPase genes. It is believed that ion-translocating PPases maintain the potential across the membrane when ATP deficiency occurs in the cell. Thus, it can be assumed that representatives of *Thiotrichaceae* are ready to reflect unfavorable circumstances.

It would be tempting to assume that potential Na⁺-motive ATPases detected in representatives of the order *Thiotrichales* may have a physiological function of ATP synthesis, i.e., perform the role of ATP synthetase. In this case, the reason for the duplication of the ATPase function and conservation of the Na⁺-specific ATPases in them by natural selection with the simultaneous presence of H⁺-motive ATPases could be explained. Since many representatives of the order *Thiotrichales* do not belong to marine species and do not live in

environments with high salt content, it seems unlikely that Na⁺-motive ATPases perform exclusively the function of sodium pumping from the cell. It seems to us that the presence of Na⁺-specific ATPases is rather related to the ability of many species of this group to inhabit niches with microoxic and maybe even anaerobic conditions.

During the transition to these conditions from normoxic conditions, the energy efficiency of the oxygen electron transport chain (ETC) decreases because the redox potential difference from the beginning of the ETC to its terminal site decreases. Under these conditions, the transition to sodium energetics would mean the transition to a more economical energy regime due to the fact that biomembranes have less leakage by sodium than by proton and consequently keep the electrochemical gradient of Na⁺ ions on the membrane longer. Then, under conditions of oxygen absence or deficiency, energy storage by oxidative phosphorylation involving Na⁺-motive ATP-synthetases would be more progressive than switching to substrate phosphorylation. At the same time, membrane PPases with both H⁺ and Na⁺ ion translocation functions could provide an additional safety net against a critical drop in the electrochemical gradient of H⁺ or Na⁺ ion concentration across the membrane.

2.5. Description of New Genus and Species

2.5.1. Description of 'Candidatus Thiocaldithrix' gen. nov.

Thiocaldithrix (Thi.o.cal.di'thrix. Gr. neut. n. *theion* sulfur, brimstone, L. transliteration *thium*, sulfur; L. fem. adj. *caldus* hot; Gr. fem. n. *thrix* thread; N.L. fem. n. *Thiocaldithrix* is a filament existing in thermal spring in the presence of hydrogen sulfide).

2.5.2. Description of 'Candidatus Thiocaldithrix dubininis' sp. nov.

- *dubininis* (du.bi'ni.nis. N.L. fem. gen. n. *dubininis* in honor of the Russian microbiologist Galina Dubinina, who made a significant contribution to the study of filamentous colorless sulfur bacteria).
- Not cultivated. Cells presumably perform a respiratory type of metabolism, facultative anaerobes, and lithoautotrophs. During lithoautotrophic growth, they obtain energy by oxidation of reduced sulfur compounds and carbon fixation in the Calvin–Benson–Bassham cycle. Potentially capable of anaerobic respiration in the presence of thiosulfate with the production of sulfide and sulfite due to thiosulfate reductase activity. Unable to participate in dissimilatory and assimilatory nitrate reduction and molecular nitrogen fixation. The 16S rRNA gene sequence is 94.7% identical to that of *T. nivea* JP2^T.
- Source: MAG GKL-01 was obtained from the bacterial fouling metagenome of a hydrogen sulfide thermal spring.
- GC fraction of genomic DNA (%): 44 (genome sequence).
- GenBank accession number (whole genome assembly): GCA_029972225.1.

2.5.3. Description of 'Candidatus Thiothrix putei' sp. nov.

- *putei* (pu.te'i. L. gen. n. *putei* of a well, from which the sample was taken, used to obtain this genome).
- Not cultivated. Cells presumably perform a respiratory type of metabolism, facultative anaerobes, and lithoautotrophs. During lithoautotrophic growth, they obtain energy by oxidation of reduced sulfur compounds and carbon fixation in the Calvin–Benson–Bassham cycle. Unable to participate in assimilatory nitrate reduction and molecular nitrogen fixation. Capable of dissimilatory anaerobic reduction of nitrate.
- Source: MAG GKL-02 was obtained from the bacterial fouling metagenome of a hydrogen sulfide thermal spring.
- GC fraction of genomic DNA (%): 50.8 (genome sequence).
- GenBank accession number (whole genome assembly): GCA_029972135.1.

3. Materials and Methods

3.1. Geography and Physicochemical Characteristics of Environmental Sampling Sites for Metagenomic Characterization of *Thiocaldithrix* sp. GKL-01 and *Thiothrix* sp. GKL-02

Biomass sampling of microbial fouling for metagenome analysis was performed in the sulfide hydrogen thermal spring in the town of Goryachiy Klyuch, Krasnodar region, Russia. The total community DNA was isolated from 500 mg of a microbial mat using a DNeasy PowerSoil DNA isolation kit (Qiagen, Hilden, Germany).

The physicochemical parameters of the water (pH, temperature, and redox potential) were measured with a HI18314F pH meter (Hanna Instruments, Vöhringen, Germany). The concentration of acid-labile sulfide in the samples was determined using the spectrophotometric method with *para*-phenylenediamine and by direct iodometric titration, preliminarily fixing the sulfide with 10% zinc acetate. The concentration of dissolved oxygen in the medium was determined using a HI 9142 oxygen meter (Romania). The total mineralization was determined by the method of electrical conductivity using a Multitest KSL-101 conductometer.

3.2. 16S rRNA Gene Profiling

Overall, 16S rRNA gene fragments were amplified from metagenomic DNA sample by PCR with the universal primers 341F (50-CCTAYGGGDBGCWSCAG) and 806R (50-GGACTACNVGGGTHCTAAT) [41]. The obtained PCR products were bar-coded using the Nextera XT Index Kit v. 2 (Illumina, San Diego, CA, USA) and sequenced on an Illumina MiSeq instrument in a paired reads mode (2 × 300 nt). Pairwise reads were merged using FLASH v.1.2.11 [42]. Obtained 16S rRNA gene sequences were filtered to exclude low-quality and chimeric sequences and clustered into operational taxonomic units (OTUs) at a 97% identity threshold using the USEARCH v. 11 program [43]. To calculate the OTU abundances, the obtained reads were mapped to OTU sequences at a 97% global identity threshold by USEARCH. The OTUs composed of only a single read were discarded. The taxonomic assignment of the OTUs was performed using the VSEARCH v. 2.14.1 program and SILVA v.138 rRNA reference database [44].

3.3. Metagenome-Sequencing and Assembly of GKL-01 and GKL-02 MAGs

Metagenomic DNA isolated from Goryachiy Klyuch was sequenced using Illumina and Oxford Nanopore techniques. Sequencing on a MinION device (Oxford Nanopore Technologies, Oxford, UK) was performed using the ligation sequencing kit 1D and FLOMIN110 cells. A total of 3,084,557 reads containing 10.2 Gb (N50 of 9695 nt) were generated. The library for Illumina sequencing was prepared using the NEBNext Ultra II DNA library preparation kit (New England Biolabs, Ipswich, MA, USA). Sequencing of this library on an Illumina HiSeq2500 in a paired-end format (2 × 150 nt) produced 172,847,806 read pairs (about 52 Gb). Adapter sequences were removed with Cutadapt v.3.4 [45], and low-quality ends (q = 30) were trimmed using Sickle v.1.33 (<https://github.com/najoshi/sickle>, accessed on 24 July 2023).

MinION reads were assembled into contigs using Flye v. 2.8.2 in metagenome mode [46]. Several contig assemblies were generated using different parts of the total number of obtained MinION reads (from 10% to 100%, in 10% increments). The sequences of the assembled contigs were corrected using Illumina reads with two iterations of NextPolish v.1.4.1 [47].

The obtained contigs were binned into MAGs using MetaBAT v.2.15 [48]. The obtained MAGs were taxonomically classified using the Genome Taxonomy Database Toolkit (GTDB-Tk) v.1.5.0 [49] and the Genome Taxonomy database (GTDB) [3]. Two MAGs, obtained as circular contigs upon assembly of 20% and 30% of all MinION reads (designated GKL-02 and GKL-01, respectively), were assigned to the family *Thiotrichaceae*.

CheckM2 v.1.0.1 [50] was used to evaluate the completeness and contamination values of obtained MAGs.

3.4. Genome Analysis and Annotation

Gene search and annotation were carried out using the RAST server 2 [51], followed by manual correction of the annotation by comparing the predicted protein sequences with the National Center for Biotechnology Information (NCBI) databases. ANI was calculated using an online resource (<https://www.ezbiocloud.net/tools/ani> (accessed on 28 April 2023)) based on the OrthoANI algorithm, using USEARCH [52]. AAI between the genomes was determined using the aai.rb script from the enveomics collection [53]. dDDH calculation was performed using the GGDC online platform (<https://ggdc.dsmz.de/ggdc.php#> (accessed on 6 May 2023)).

For genome-based phylogenetic analysis, GTDB-Tk v.1.5.0 [49] was used to identify 120 single-copy marker genes in the genomes and to create multiple sequence alignments of concatenated amino acid sequences. The maximum likelihood tree was estimated from the alignment by PhyML v. 3.3 [54] using default parameters (LG amino acid substitution model, 4 substitution rates categories modeled by discrete gamma distribution with estimated shape parameter, branch support values calculated by approximate Bayes method).

3.5. Phylogenetic Analysis of F_0F_1 -ATPase Subunits and Membrane-Bound PPases

Translated protein-coding genes were retrieved from the UniProtKB, NCBI, and RAST Server (version 2.0) databases. Protein-based phylogenetic analyses were fulfilled at the Phylogeny website (<http://www.phylogeny.fr/>, accessed on 8 June 2023) using Phylogeny Workspace Phylogenetic Analysis instruments as follows. The obtained amino acid sequences were aligned using MUSCLE software for multiple sequence alignment [55] and the Gblocks program to eliminate poorly aligned positions and divergent regions [56]. Phylogeny was estimated using the maximum likelihood method [57,58]. Final dendrograms were prepared using the MEGA 4.1 (beta) software [59].

4. Conclusions

Metagenomic analysis of a bacterial mat developing in a sulfidic thermal spring revealed two new candidate species of the family *Thiotrichaceae*, '*Candidatus* Thiothrix putei' sp. nov. GKL-02 and '*Candidatus* Thiocaldithrix dubininis' gen. nov., sp. nov. GKL-01, the latter representing a novel genus. Comparative genomic analysis showed that members of the family *Thiotrichaceae* have versatile metabolism, which indicates broad adaptive capabilities. An intriguing result was revealing features of two types of energetics in many members of *Thiotrichaceae*, which dispose of Na^+ -motive ATPases in addition to more common H^+ -motive ATPases. Another identified enzyme possessing Na^+ - and H^+ -motive forms was the less common membrane-bound PPase. The K^+ -independent and K^+ -dependent forms of this PPase, respectively, are capable of converting the energy of the proton and sodium ion gradient on the membrane into the high-energy bond of pyrophosphate and vice versa. Sodium-transporting forms of the above enzymes were previously described in anaerobic organisms. The reasons for the existence of such enzymes in facultatively aerobic members of *Thiotrichaceae* are not completely clear, but it can be assumed that their presence is associated with the functioning of the sodium cycle.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/ijms241814199/s1>.

Author Contributions: Conceptualization, N.V.R., M.Y.G. and M.S.M.; investigation, M.S.M., D.D.S., A.V.M. and T.S.R.; data curation, A.V.B. and N.V.R.; writing—original draft preparation, M.S.M., D.D.S. and T.S.R.; writing—review and editing, N.V.R., M.S.M., D.D.S., T.S.R. and M.Y.G.; supervision, N.V.R. and M.Y.G.; project administration, N.V.R. and M.Y.G.; funding acquisition, M.Y.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded partly by the Russian Science Foundation (grant number 20-14-00137, (<https://rscf.ru/en/project/20-14-00137/>, accessed on 15 May 2023) to M.Y.G., M.S.M., D.D.S., T.S.R. and A.V.B., genome-based analyses of metabolic pathways and energetics types, studies of physiological and ecological characteristics) and the Ministry of Science and Higher Education of the Russian Federation (N.V.R. and A.V.M., 16S rRNA gene sequencing and genome-based phylogenetic analysis).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Boden, R.; Scott, K.M. Evaluation of the genus *Thiothrix* Winogradsky 1888 (Approved Lists 1980) emend. Aruga et al. 2002: Reclassification of *Thiothrix disciformis* to *Thiolinea disciformis* gen. nov., comb. nov., and of *Thiothrix flexilis* to *Thiofilum flexile* gen. nov., comb. nov., with emended description of *Thiothrix*. *Int. J. Syst. Evol. Microbiol.* **2018**, *68*, 2226–2239. [[CrossRef](#)]
2. Parks, D.H.; Chuvochina, M.; Waite, D.W.; Rinke, C.; Skarshewski, A.; Chaumeil, P.A.; Hugenholtz, P. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* **2018**, *36*, 996–1004. [[CrossRef](#)]
3. Parks, D.H.; Chuvochina, M.; Rinke, C.; Mussig, A.J.; Chaumeil, P.A.; Hugenholtz, P. GTDB: An ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res.* **2022**, *50*, D785–D794. [[CrossRef](#)]
4. Grabovich, M.Y.; Ravin, N.V.; Boden, R. *Thiothrix*. In *Bergey's Manual of Systematics of Archaea and Bacteria*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2023; p. 21. [[CrossRef](#)]
5. Mardanov, A.V.; Gruzdev, E.V.; Smolyakov, D.D.; Rudenko, T.S.; Beletsky, A.V.; Gureeva, M.V.; Markov, N.D.; Berestovskaya, Y.Y.; Pimenov, N.V.; Ravin, N.V.; et al. Genomic and metabolic insights into two novel *Thiothrix* species from enhanced biological phosphorus removal systems. *Microorganisms* **2020**, *8*, 2030. [[CrossRef](#)]
6. Ravin, N.V.; Rossetti, S.; Beletsky, A.V.; Kadnikov, V.V.; Rudenko, T.S.; Smolyakov, D.D.; Moskvitina, M.I.; Gureeva, M.V.; Mardanov, A.V.; Grabovich, M.Y. Two new species of filamentous sulfur bacteria of the genus *Thiothrix*, *Thiothrix winogradskiji* sp. nov. and ‘*Candidatus* *Thiothrix sulfatfontis*’ sp. nov. *Microorganisms* **2022**, *10*, 1300. [[CrossRef](#)] [[PubMed](#)]
7. Ravin, N.V.; Rudenko, T.S.; Smolyakov, D.D.; Beletsky, A.V.; Rakitin, A.L.; Markov, N.D.; Fomenkov, A.; Sun, L.; Roberts, R.J.; Novikov, A.A.; et al. Comparative genome analysis of the genus *Thiothrix* involving three novel species, *Thiothrix subterranea* sp. nov. Ku-5, *Thiothrix litoralis* sp. nov. AS and ‘*Candidatus* *Thiothrix anitrata*’ sp. nov. A52, revealed the conservation of the pathways of dissimilatory sulfur metabolism and variations in the genetic inventory for nitrogen metabolism and autotrophic carbon fixation. *Front. Microbiol.* **2021**, *12*, 760289. [[CrossRef](#)] [[PubMed](#)]
8. Nelson, D.C. The Genus *Beggiatoa*. In *The Prokaryotes*; Balows, A., Trüper, H.G., Dworkin, M., Harder, W., Schleifer, K.H., Eds.; Springer: New York, NY, USA, 1992; pp. 3171–3180. [[CrossRef](#)]
9. Howarth, R.; Unz, R.F.; Seviour, E.M.; Seviour, R.J.; Blackall, L.L.; Pickup, R.W.; Jones, J.G.; Yaguchi, J.; Head, I.M. Phylogenetic relationships of filamentous sulfur bacteria (*Thiothrix* spp. and Eikelboom type 021N bacteria) isolated from wastewater-treatment plants and description of *Thiothrix eikelboomii* sp. nov., *Thiothrix unzii* sp. nov., *Thiothrix fructosivorans* sp. nov. and *Thiothrix defluvii* sp. nov. *Int. J. Syst. Bacteriol.* **1999**, *49*, 1817–1827. [[CrossRef](#)] [[PubMed](#)]
10. Grabovich, M.Y.; Patriitskaya, V.Y.; Muntyan, M.S.; Dubinina, G.A. Lithoautotrophic growth of the freshwater strain *Beggiatoa* D-402 and energy conservation in a homogeneous culture under microoxic conditions. *FEMS Microbiol. Lett.* **2001**, *204*, 341–345. [[CrossRef](#)] [[PubMed](#)]
11. Muntyan, M.S.; Grabovich, M.Y.; Patriitskaya, V.Y.; Dubinina, G.A. Regulation of metabolic and electron transport pathways in the freshwater bacterium *Beggiatoa leptomitiformis* D-402. *Microbiology* **2005**, *74*, 388–394. [[CrossRef](#)]
12. Teske, A.; Nelson, D.C. The Genera *Beggiatoa* and *Thioploca*. In *The Prokaryotes*; Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.H., Stackebrandt, E., Eds.; Springer: New York, NY, USA, 2006; Volume 6, pp. 784–810. [[CrossRef](#)]
13. Dimroth, P. Sodium ion transport decarboxylases and other aspects of sodium ion cycling in bacteria. *Microbiol. Rev.* **1987**, *51*, 320–340. [[CrossRef](#)] [[PubMed](#)]
14. Skulachev, V.P. The laws of cell energetics. *Eur. J. Biochem.* **1992**, *208*, 203–209. [[CrossRef](#)]
15. Häse, C.C.; Fedorova, N.D.; Galperin, M.Y.; Dibrov, P.A. Sodium ion cycle in bacterial pathogens: Evidence from cross-genome comparisons. *Microbiol. Mol. Biol. Rev.* **2001**, *65*, 353–370. [[CrossRef](#)]
16. Mulkidjanian, A.Y.; Galperin, M.Y.; Makarova, K.S.; Wolf, Y.I.; Koonin, E.V. Evolutionary primacy of sodium bioenergetics. *Biol. Direct.* **2008**, *3*, 13. [[CrossRef](#)] [[PubMed](#)]
17. Bowers, R.M.; Kyrpides, N.C.; Stepanauskas, R.; Harmon-Smith, M.; Doud, D.; Reddy, T.B.K.; Schulz, F.; Jarett, J.; Rivers, A.R.; Eloie-Fadrosch, E.A.; et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat. Biotechnol.* **2017**, *35*, 725–731. [[CrossRef](#)]

18. Konstantinidis, K.T.; Tiedje, J.M. Towards a genome-based taxonomy for prokaryotes. *J. Bacteriol.* **2005**, *87*, 6258–6264. [[CrossRef](#)] [[PubMed](#)]
19. Konstantinidis, K.; Rosselló-Móra, R.; Amann, R. Uncultivated microbes in need of their own taxonomy. *ISME J.* **2017**, *11*, 2399–2406. [[CrossRef](#)] [[PubMed](#)]
20. Ravin, N.V.; Rudenko, T.S.; Smolyakov, D.D.; Beletsky, A.V.; Gureeva, M.V.; Samylina, O.S.; Grabovich, M.Y. History of the study of the genus *Thiothrix*: From the first enrichment cultures to pangenomic analysis. *Int. J. Mol. Sci.* **2022**, *23*, 9531. [[CrossRef](#)]
21. Skulachev, V.P. Sodium bioenergetics. *Trends Biochem. Sci.* **1984**, *9*, 483–485. [[CrossRef](#)]
22. Skulachev, V.P. Membrane-linked energy transductions. Bioenergetic functions of sodium: H⁺ is not unique as a coupling ion. *Eur. J. Biochem.* **1985**, *151*, 199–208. [[CrossRef](#)]
23. Skulachev, V.P. The latest news from the sodium world. *Biochim. Biophys. Acta* **1994**, *1187*, 216–221. [[CrossRef](#)]
24. Mitchell, P. Chemiosmotic coupling in oxidative and photosynthetic phosphorylation. *Biol. Rev. Camb. Philos. Soc.* **1966**, *41*, 445–502. [[CrossRef](#)] [[PubMed](#)]
25. Dimroth, P. Reconstitution of sodium transport from purified oxaloacetate decarboxylase and phospholipid vesicles. *J. Biol. Chem.* **1981**, *256*, 11974–11976. [[CrossRef](#)] [[PubMed](#)]
26. Tokuda, H.; Unemoto, T. Na⁺ is translocated at NADH: Quinone oxidoreductase segment in the respiratory chain of *Vibrio alginolyticus*. *J. Biol. Chem.* **1984**, *259*, 7785–7790. [[CrossRef](#)]
27. Laubinger, W.; Dimroth, P. Characterization of the Na⁺-stimulated ATPase of *Propionigenium modestum* as an enzyme of the F1F0 type. *Eur. J. Biochem.* **1987**, *168*, 475–480. [[CrossRef](#)] [[PubMed](#)]
28. Malinen, A.M.; Belogurov, G.A.; Baykov, A.A.; Lahti, R. Na⁺-pyrophosphatase: A novel primary sodium pump. *Biochemistry* **2007**, *46*, 8872–8878. [[CrossRef](#)]
29. Inoue, K.; Ono, H.; Abe-Yoshizumi, R.; Yoshizawa, S.; Ito, H.; Kogure, K.; Kandori, H. A light-driven sodium ion pump in marine bacteria. *Nat. Commun.* **2013**, *4*, 1678. [[CrossRef](#)]
30. Muntyan, M.S.; Cherepanov, D.A.; Malinen, A.M.; Bloch, D.A.; Sorokin, D.Y.; Severina, I.I.; Ivashina, T.V.; Lahti, R.; Muyzer, G.; Skulachev, V.P. Cytochrome *cbb*₃ of *Thioalkalivibrio* is a Na⁺-pumping cytochrome oxidase. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 7695–7700. [[CrossRef](#)]
31. Muntyan, M.S.; Viryasov, M.B.; Sorokin, D.Y.; Skulachev, V.P. Sodium Energetic Cycle in the Natronophilic Bacterium *Thioalkalivibrio versutus*. *Int. J. Mol. Sci.* **2022**, *23*, 1965. [[CrossRef](#)] [[PubMed](#)]
32. Dibrova, D.V.; Galperin, M.Y.; Mulikjanian, A.Y. Characterization of the N-ATPase, a distinct, laterally transferred Na⁺-translocating form of the bacterial F-type membrane ATPase. *Bioinformatics* **2010**, *26*, 1473–1476. [[CrossRef](#)]
33. Neumann, S.; Matthey, U.; Kaim, G.; Dimroth, P. Purification and properties of the F₁F₀ ATPase of *Ilyobacter tartaricus*, a sodium ion pump. *J. Bacteriol.* **1998**, *180*, 3312–3316. [[CrossRef](#)]
34. Meier, T.; Polzer, P.; Diederichs, K.; Welte, W.; Dimroth, P. Structure of the rotor ring of F-Type Na⁺-ATPase from *Ilyobacter tartaricus*. *Science* **2005**, *308*, 659–662. [[CrossRef](#)]
35. Ludwig, W.; Kaim, G.; Laubinger, W.; Dimroth, P.; Hoppe, J.; Schleifer, K.H. Sequence of subunit c of the sodium ion translocating adenosine triphosphate synthase of *Propionigenium modestum*. *Eur. J. Biochem.* **1990**, *193*, 395–399. [[CrossRef](#)]
36. Kuhns, M.; Trifunović, D.; Huber, H.; Müller, V. The Rnf complex is a Na⁺ coupled respiratory enzyme in a fermenting bacterium, *Thermotoga maritima*. *Commun. Biol.* **2020**, *3*, 431. [[CrossRef](#)]
37. Von Ballmoos, C.; Wiedenmann, A.; Dimroth, P. Essentials for ATP synthesis by F₁F₀ ATP synthases. *Ann. Rev. Biochem.* **2009**, *78*, 649–672. [[CrossRef](#)]
38. Schlegel, K.; Leone, V.; Faraldo-Gómez, J.D.; Müller, V. Promiscuous archaeal ATP synthase concurrently coupled to Na⁺ and H⁺ translocation. *Proc. Nat. Acad. Sci. USA* **2012**, *109*, 947–952. [[CrossRef](#)] [[PubMed](#)]
39. Nishio, Y.; Nakamura, Y.; Kawarabayashi, Y.; Usuda, Y.; Kimura, E.; Sugimoto, S.; Matsui, K.; Yamagishi, A.; Kikuchi, H.; Ikeo, K.; et al. Comparative complete genome sequence analysis of the amino acid replacements responsible for the thermostability of *Corynebacterium efficiens*. *Genome Res.* **2003**, *13*, 1572–1579. [[CrossRef](#)]
40. Luoto, H.H.; Belogurov, G.A.; Baykov, A.A.; Lahti, R.; Malinen, A.M. Na⁺-translocating membrane pyrophosphatases are widespread in the microbial world and evolutionarily precede H⁺-translocating pyrophosphatases. *J. Biol. Chem.* **2011**, *286*, 21633–21642. [[CrossRef](#)]
41. Frey, B.; Rime, T.; Phillips, M.; Stierli, B.; Hajdas, I.; Widmer, F.; Hartmann, M. Microbial diversity in European alpine permafrost and active layers. *FEMS Microbiol. Ecol.* **2016**, *92*, fiw018. [[CrossRef](#)] [[PubMed](#)]
42. Magoc, T.; Salzberg, S.L. FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **2011**, *27*, 2957–2963. [[CrossRef](#)]
43. Edgar, R.C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **2010**, *26*, 2460–2461. [[CrossRef](#)] [[PubMed](#)]
44. Rognes, T.; Flouri, T.; Nichols, B.; Quince, C.; Mahé, F. VSEARCH: A versatile open source tool for metagenomics. *PeerJ* **2016**, *4*, e2584. [[CrossRef](#)] [[PubMed](#)]
45. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **2011**, *17*, 10–12. [[CrossRef](#)]
46. Rossmassler, K.; Hanson, T.E.; Campbell, B.J. Diverse sulfur metabolisms from two subterranean sulfidic spring systems. *FEMS Microbiol. Lett.* **2016**, *363*, fnw162. [[CrossRef](#)] [[PubMed](#)]
47. Hu, J.; Fan, J.; Sun, Z.; Liu, S. NextPolish: A fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* **2020**, *36*, 2253–2255. [[CrossRef](#)]

48. Kang, D.D.; Li, F.; Kirton, E.; Thomas, A.; Egan, R.; An, H.; Wang, Z. MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **2019**, *7*, e7359. [[CrossRef](#)]
49. Chaumeil, P.A.; Mussig, A.J.; Hugenholtz, P.; Parks, D.H. GTDB-Tk: A toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **2019**, *36*, 1925–1927. [[CrossRef](#)]
50. Chklovski, A.; Parks, D.H.; Woodcroft, B.J.; Tyson, G.W. CheckM2: A rapid, scalable and accurate tool for assessing microbial genome quality using machine learning. *Nat. Methods* **2023**, *20*, 1203–1212. [[CrossRef](#)] [[PubMed](#)]
51. Brettin, T.; Davis, J.J.; Disz, T.; Edwards, R.A.; Gerdes, S.; Olsen, G.J.; Olson, R.; Overbeek, R.; Parrello, B.; Pusch, G.D.; et al. RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci. Rep.* **2015**, *5*, 8365. [[CrossRef](#)] [[PubMed](#)]
52. Lee, I.; Ouk Kim, Y.; Park, S.C.; Chun, J. OrthoANI: An improved algorithm and software for calculating average nucleotide identity. *Int. J. Syst. Evol. Microbiol.* **2016**, *66*, 1100–1103. [[CrossRef](#)]
53. Rodriguez-R, L.M.; Konstantinidis, K.T. The enveomics collection: A toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ. Prepr.* **2016**, *4*, e1900v1. [[CrossRef](#)]
54. Guindon, S.; Dufayard, J.F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **2010**, *59*, 307–321. [[CrossRef](#)] [[PubMed](#)]
55. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [[CrossRef](#)] [[PubMed](#)]
56. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **2000**, *17*, 540–552. [[CrossRef](#)] [[PubMed](#)]
57. Guindon, S.; Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **2003**, *52*, 696–704. [[CrossRef](#)] [[PubMed](#)]
58. Anisimova, M.; Gascuel, O. Approximate likelihood ratio test for branches: A fast, accurate and powerful alternative. *Syst. Biol.* **2006**, *4*, 539–552. [[CrossRef](#)]
59. Tamura, K.; Dudley, J.; Nei, M.; Kumar, S. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **2007**, *24*, 1596–1599. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.