

Article

# Comparative Genomics of *Beggiatoa leptomitiformis* Strains D-401 and D-402<sup>T</sup> with Contrasting Physiology But Extremely High Level of Genomic Identity

Tatyana S. Rudenko <sup>1</sup>, Sergey V. Tarlachkov <sup>2,3</sup>, Nikolay D. Shatskiy <sup>1</sup> and Margarita Yu. Grabovich <sup>1,\*</sup>

- <sup>1</sup> Department of Biochemistry and Cell Physiology, Voronezh State University, Universitetskaya pl., 1, Voronezh 394018, Russia; ipigun6292@gmail.com (T.S.R.); kolyan9572@mail.ru (N.D.S.)
- <sup>2</sup> All-Russian Collection of Microorganisms (VKM), G.K. Skryabin Institute of Biochemistry and Physiology of Microorganisms, Pushchino Scientific Center for Biological Research of the Russian Academy of Sciences, Federal Research Center, pr. Nauki, 5, Pushchino 142290, Russia; sergey@tarlachkov.ru
- <sup>3</sup> Branch of Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry of the Russian Academy of Sciences, pr. Nauki, 6, Pushchino 142290, Russia
- \* Correspondence: margarita\_grabov@mail.ru; Tel.: +7-473-2208877; Fax: +7-473-2208755

Received: 23 May 2020; Accepted: 18 June 2020; Published: 19 June 2020



**Abstract:** Representatives of filamentous colorless sulfur-oxidizing bacteria often dominate in sulfide biotopes, preventing the diffusion of toxic sulfide into the water column. One of the most intriguing groups is a recently described *Beggiatoa leptomitiformis* including strains D-401 and D-402<sup>T</sup>. Both strains have identical genes encoding enzymes which are involved in the oxidation of hydrogen sulfide and thiosulfate. Surprisingly, the *B. leptomitiformis* strain D-401 is not capable to grow lithotrophically in the presence of reduced sulfur compounds and to accumulate elemental sulfur inside the cells, in contrast to the D-402<sup>T</sup> strain. In general, genomes of D-401 and D-402<sup>T</sup> have an extremely high level of identity and only differ in 1 single-letter substitution, 4 single-letter indels, and 16 long inserts. Among long inserts, 14 are transposons. It was shown that in the D-401 strain, a gene coding for a sulfur globule protein was disrupted by one of the mentioned transposons. Based on comparative genomics, RT-qPCR, and HPLC-MS/MS, we can conclude that this gene plays a crucial role in the formation of the sulfur globules inside the cells, and the disruption of its function prevents lithotrophic growth of *B. leptomitiformis* in the presence of reduced sulfur compounds.

**Keywords:** *Beggiatoa leptomitiformis*; comparative genomics; transposon; dissimilatory sulfur metabolism; lithotrophic growth; sulfur globule protein

## 1. Introduction

Colorless sulfur bacteria were discovered by S.N. Winogradsky at the end of the 19th century. Using *Beggiatoa* as an example, S.N. Winogradsky formulated the idea of chemosynthesis (chemolithoautotrophy) which would become a key to understanding the metabolism and biodiversity of bacteria and to developing a concept of the role of microorganisms in the biosphere.

Representatives of the genus *Beggiatoa* are typical inhabitants of sulfide environments. These motile microorganisms form mats on the surface of sulfide sediments and are located in the layer where inorganic oxidizing and reducing agents are encountered. They are typical gradient organisms and are able to use reduced sulfur compounds as electron donors for energy metabolism. Such oxidation processes are accompanied by an accumulation of elemental sulfur within cells [1,2]. Currently the

genus *Beggiatoa* includes two species with valid names for which pure cultures are available—*Beggiatoa alba* B18LD<sup>T</sup> (ATCC 33555<sup>T</sup>), *Beggiatoa alba* B15LD (DSM 1416), *Beggiatoa leptomitiformis* D-402<sup>T</sup> (DSM 14946<sup>T</sup>), and *Beggiatoa leptomitiformis* D-401 (DSM 14945<sup>T</sup>).

Surprisingly, *B. leptomitiformis* strains D-402<sup>T</sup> and D-401 possess metabolic and morphological differences in the presence of hydrogen sulfide and thiosulfate. It was found that strain D-401 was not capable of lithotrophic growth [3] and did not accumulate elemental sulfur inside cells, while strain D-402<sup>T</sup> was capable of lithoheterotrophic and lithoautotrophic growth, and in the presence of reduced sulfur, compounds deposited an abundant amount of elemental sulfur inside the cells which can reach 20 to 50% of dry weight [4].

The complete genomic sequences of *B. leptomitiformis* strains D-401 and D-402<sup>T</sup> were obtained in 2015 and 2018 [5,6]. Despite the fact that these strains were isolated from geographically remote habitats, the similarity of their genomes is more than 99.5% [5]. Therefore, elucidation of the molecular, biochemical, and evolutionary mechanisms responsible for such remarkable phenotypic differences with high genomic similarity is of great interest. This will contribute to the understanding of the evolution and ecology of the family Beggiatoaceae.

We studied the differences between genomes of strains D-401 and D-402<sup>T</sup> and determined the role of these differences in sulfur metabolism connected with the ability to lithotrophic growth and accumulation of elemental sulfur in the presence of reduced sulfur compounds.

## 2. Materials and Methods

### 2.1. Cultivation

Medium for lithoheterotrophic cultivation of the *Beggiatoa leptomitiformis* strains contained the following composition: ( $\text{g} \times \text{L}^{-1}$ ):  $\text{NaNO}_3$ , 0.62;  $\text{NaH}_2\text{PO}_4$ , 0.125;  $\text{CaCl}_2 \times 2\text{H}_2\text{O}$ , 0.3;  $\text{Na}_2\text{SO}_4$ , 0.5;  $\text{KCl}$ , 0.125;  $\text{MgCl}_2 \times 6\text{H}_2\text{O}$ , 0.5;  $\text{NaHCO}_3$ , 0.125; peptone, 0.2; lactate, 0.2;  $\text{Na}_2\text{S}_2\text{O}_3 \times 5\text{H}_2\text{O}$ , 1.0. The pH of the medium was adjusted to 7.0 before inoculation. Before inoculation, the medium was supplemented with vitamins and microelements [7]. Thiosulfate was not added to the medium for organoheterotrophic cultivation; for lithoautotrophic growth, organic compounds were not added and the concentration of  $\text{NaHCO}_3$  was increased to  $0.5 \text{ g} \times \text{L}^{-1}$ . Cultures were grown at 27 °C.

To create microaerobic conditions, 0.5 L bottles with polybutyl rubber gaskets and screwed metal caps were filled with 50 mL freshly boiled medium and purged with argon. After that, the required volume of air was introduced into the bottles: the final concentration of  $\text{O}_2$  in the gas phase was 2.0% [3].

### 2.2. Genome Polishing

DNA was isolated using the FastDNA<sup>TM</sup> SpinKit (MP Biomedicals, Irvine, CA, USA). The libraries were prepared using the KAPA HyperPlus kit with adapters from the KAPA Dual-Indexed Adapter Kit. The resulting libraries were sequenced on the Illumina MiSeq platform with  $2 \times 300$  bp reads. Raw sequencing data preprocessing was performed using Trimmomatic v. 0.39 [8], and then the resulted reads were mapped to the corresponding genomic assemblies using Bowtie2 v. 2.3.4.3 [9]. Genome corrections were performed using Pilon v. 1.23 [10].

### 2.3. Genome Analysis

The MEGA 6 program was used [11] for cluster analysis of long inserts. Sequence alignment was performed using ClustalW [12]. Trees were constructed by the neighbor-joining method [13] using the  $p$ -distance. Confidence values of branch nodes were evaluated using bootstrap analysis by constructing 1000 alternative trees. Homologous protein sequences were searched using the NCBI database (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). A search for conserved domains in the protein sequences was performed using the NCBI database (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). A search of signal peptides in amino acid sequences was performed using the SignalP service (<http://www.cbs.dtu.dk/services/SignalP>).

#### 2.4. Gene Expression Analysis

RNA was isolated using the ExtractRNA reagent (Evrogen, Moscow, Russia) in accordance with the manufacturer's protocol. RNA quality was evaluated by electrophoresis on a 2% agarose gel with 2.2 M formaldehyde added. RNA concentration was measured using an HS Qubit RNA assay kit (Thermo Fisher Scientific, Waltham, MA, USA) on a Qubit 2.0 fluorometer (Thermo Fisher Scientific, Waltham, MA, USA). Then, 1000 ng of RNA was reverse transcribed using M-MuLV (SybEnzyme, Moscow, Russia) according to the manufacturer's protocol. Quantitative RT-PCR was performed using SYBR Green I on a Bio-Rad CFX96™ real-time system (Bio-Rad, Hercules, CA, USA).

In order to find optimal amplification conditions, a temperature gradient was used. The final program included 95 °C, 3' + ((95°, 20" + 60 °C, 20" + 72 °C, 30") × 39). Fragments of genes encoding the proteins ALG67575.1, ALG68134.1, and ALG67202.2 were amplified using primers 5'-GGGGCAACAATGAATGGGGT-3' and 5'-GACGACCATAACCACGACCG-3' for ALG67575.1, 5'-GCTGAAAAATGGGACGCAGG-3' and 5'-TCACCCACGTTGCCATAACC-3' for ALG68134.1, 5'-CTTTGTGATTTGGGCGGCAA-3' and 5'-AGATGCGACAAGTGATGCGG-3' for ALG67202.2; all primers were designed with PrimerBLAST (<http://www.ncbi.nlm.nih.gov/tools/primer-blast>).

#### 2.5. Enzymatic Assay

Sulfite oxidase activity (SOE) (EC 1.8.5.6) was measured in a buffer containing 50 mM Tris-HCl pH 8.0, 2 mM ferricyanide, 2 mM sodium sulfite, EDTA 2.9 mM. The activity was determined spectrophotometrically at 420 nm by a decrease in optical density as a result of ferricyanide reduction [14].

The protein concentration in the sample for the measurement of the enzyme activity was in the range of 100–500 µg mL<sup>-1</sup> (crude cell extract).

#### 2.6. Sample Preparation for Proteomic Analysis

Cells were resuspended in 200 µL of lysis buffer containing 1% dithiothreitol, 4% CHAPS, 7M urea, 2M thiourea, and 5% ampholytes 3/10. Lysate was prepared using a Potter homogenizer. The homogenate was centrifuged at 5000× g for 10 min at 20 °C and the supernatant was collected. Protein concentration was determined by the Bradford method using a commercial set Coomassie Protein Assay Reagent, Thermo Fisher according to the manufacturer's instructions.

#### 2.7. Two-Dimensional Gel Electrophoresis (2-DE)

2-DE was performed as described in [15] on a PROTEAN II xi 2-D Cell system (Bio-Rad, Hercules, CA, USA).

When conducting isoelectric focusing, the pH gradient was from 3 to 10 (servalites, Serva Electrophoresis GmbH, Heidelberg, Germany). The amount of the sample was 120 µg of protein per tube. Isofocusing was performed for 16 h under the following modes: 100 V-45 min, 200 V-45 min, 300 V-45 min, 400 V-45 min, 500 V-45 min, 600 V-45 min, 700 V-10 h, and 900 V-1.5 h.

Electrophoresis of the samples obtained after isoelectric focusing was carried out in a gradient acrylamide gel with sodium dodecyl sulfate (7.5–25%) at a voltage of 300 V. Before applying to the second dimension, the samples were incubated for 20 min in a solution containing dithiothreitol (6M urea, 2% sodium dodecyl sulfate, 10 mM dithiothreitol, 0.5M Tris-HCl, pH 6.8) to prevent oxidation of sulfhydryl groups in proteins. For visual analysis of the distribution of protein components and for mass spectrometric analysis, the gels were stained with Brilliant Blue R Staining Solution (Sigma, Georgetown, SC, USA).

The Infinity1000/26MX gel documentation system (VilberLourmat, Collégien, France) was used to obtain protein maps.

### 2.8. High Performance Liquid Chromatography in Combination with Tandem Mass Spectrometry (HPLC-MS/MS)

Proteins from the strains D-401 and D-402<sup>T</sup> were subjected to trypsin (Promega, Madison, WI, USA) and chymotrypsin (Promega, Madison, WI, USA) digestion according to the FASP protocol on 10 kDa-MWCO filters (Microcon Centrifugal Filter 10 kDa, Millipore, Burlington, MA, USA), as described previously [16].

The peptides were separated with high-performance liquid chromatography (HPLC, Ultimate 3000 Nano LC System, Thermo Scientific, Rockwell, IL, USA) in a 15-cm long C18 column with an inner diameter of 75 µm (Acclaim PepMap RSLC, Thermo Fisher Scientific, Rockwell, IL, USA). The peptides were eluted with a gradient of buffer B (80% acetonitrile, 0.1% formic acid) at a flow rate of 0.3 µL min<sup>-1</sup>. Total run time was 90 min which included the initial 4 min of column equilibration to buffer A (0.1% formic acid), then the gradient from 5 to 35% of buffer B over 65 min, then 6 min to reach 99% of buffer B, flushing 10 min with 99% of buffer B, and 5 min re-equilibration to buffer A. MS analysis was performed in triplicate with a Q Exactive HF mass spectrometer (Q Exactive HF Hybrid Quadrupole-Orbitrap Mass spectrometer, Thermo Fisher Scientific, Rockwell, IL, USA). Mass spectra were acquired at a resolution of 120,000 (MS) and 15,000 (MS/MS) in a *m/z* range of 350–1500 (MS) and 100–2000 (MS/MS). An isolation threshold of 100,000 counts was determined for precursor's selection and the top 10 precursors were chosen for fragmentation with high-energy collisional dissociation at 30 NCE and 100 ms accumulation time. Precursors with a charged state of +1 were rejected and all measured precursors were excluded from measurement for 20 s.

### 2.9. Proteomic Data Processing

The proteins of the strains were identified with the SearchGUI v. 3.3.17 program using simultaneous X!Tandem, OMSSA, and MS-GF+ algorithms [17] with protein sequences databases TrEMBL *Beggiatoa leptomitiformis* D-401 (3445 entries) and *Beggiatoa leptomitiformis* D-402<sup>T</sup> (3447 entries). Parameters were set as follows: no enzyme specificity; MS tolerance of 5 ppm and 0.01 Da tolerance for MS/MS ions; with carbamidomethylation of C as a fixed modification and oxidation of M as a variable modification. The SearchGUI output was analyzed and visualized in PeptideShaker [18] v. 1.16.44. Peptide-Spectrum Matches, peptides, and proteins were validated at a 1.0% False Discovery Rate estimated using the decoy hit distribution. Only proteins having at least two unique peptides were considered as positively identified.

## 3. Results

### 3.1. Improving Assembly Quality

In the present research, we use complete genomic sequences of *B. leptomitiformis* D-402<sup>T</sup> and D-401, previously obtained using PacBio RSII platform [5,6]. Recent studies reported that PacBio RSII sequencing technology has a high error rate in homopolymer regions [19]. At the same time, our preliminary analysis of genomic sequences of strains D-402<sup>T</sup> and D-401 showed a rather large number of homopolymer sites. Thus, to improve the quality of assemblies, we polished previously published sequences using Illumina reads obtained in the present research. Polished assemblies had been published in GenBank under accession numbers CP012373.2 and CP018889.2 for *B. leptomitiformis* D-402<sup>T</sup> and D-401, respectively. In total, we corrected 65 sites for the genome of strain D-401 and 22 sites for strain D-402<sup>T</sup>. After correction, the number of pseudogenes decreased approximately by factor 2 due to the correction of the shifts of the reading frames.

The refined sequences of the complete genomes of *B. leptomitiformis* strains allowed us to conduct a comparative analysis of homologous genes encoding all the necessary enzymes involved in the oxidation of hydrogen sulfide, thiosulfate, elemental sulfur, and sulfite (Table 1). In particular, genes of periplasmic sulfide-oxidizing enzymes were found in the genomes of the strains D-401 and D-402<sup>T</sup>: sulfide:quinoneoxygenase (*sqrQ*) type I (*sqrA*) and type VI (*sqrF*); flavocytochrome c-sulfide

dehydrogenase (FCSD) type *fccB*. Additionally, genes encoding enzymes that can be involved in the oxidation of thiosulfate due to the branched pathway were identified in two strains and SoxB activity was detected only in the strain D-402<sup>T</sup> [20].

**Table 1.** Genes encoding the enzymes of dissimilation sulfur metabolism of *Beggiatoa leptomitiformis* D-401 and D-402<sup>T</sup>.

<i>B. leptomitiformis</i> D-401		<i>B. leptomitiformis</i> D-402 <sup>T</sup>		Gene
Locus Tag	Protein Accession	Locus Tag	Protein Accession	
BLE401_14705	AUI69816.1	AL038_09565	ALG67913.1	<i>soxY</i>
BLE401_14710	AUI69817.1	AL038_09560	ALG67912.1	<i>soxZ</i>
BLE401_13150	AUI69545.1	AL038_11110	ALG68158.1	<i>soxAX</i>
BLE401_00760	AUI67367.1	AL038_05225	ALG67220.1	<i>soxB</i>
BLE401_03540	AUI67863.1	AL038_02505	ALG66791.1	<i>soxF</i>
BLE401_07405	AUI68548.1	AL038_16835	ALG69045.1	<i>sqr</i>
BLE401_16835	AUI70200.1	AL038_07480	ALG67570.1	<i>dsrE</i>
BLE401_16840	AUI70201.1	AL038_07475	ALG67569.1	<i>dsrF</i>
BLE401_16845	AUI70202.1	AL038_07470	ALG67568.1	<i>dsrH</i>
BLE401_16850	AUI70203.1	AL038_07465	ALG67567.1	<i>dsrC</i>
BLE401_14535	AUI70655.1	AL038_09735	ALG69423.1	<i>soeA</i>
BLE401_14540	AUI69788.1	AL038_09730	ALG67939.1	<i>soeB</i>
BLE401_14545	AUI69789.1	AL038_09725	ALG67938.1	<i>soeC</i>

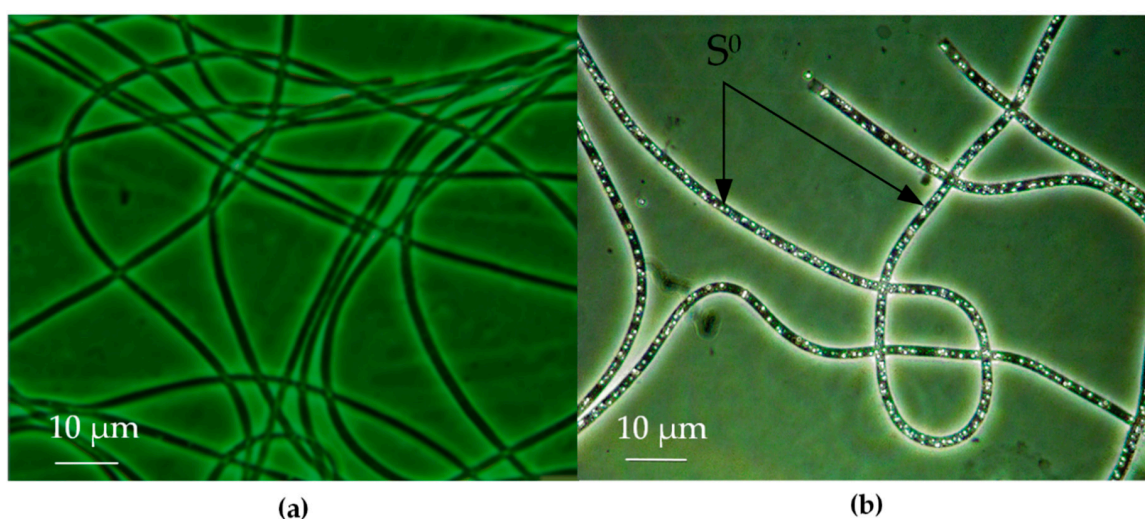
The mechanism of oxidation of sulfur to sulfite in representatives of the genus *Beggiatoa* is still not fully understood.

Oxidation of sulfite to sulfate by prokaryotes can occur in two ways: direct and indirect. Genes (*aprAB*, *sat/sopT*, *apt*) encoding the enzymes of the indirect pathway for the oxidation of sulfite to sulfate were not identified in any of the analyzed *B. leptomitiformis* genomes, and we did not find corresponding enzymatic activity [20]. The direct pathway for sulfite oxidation can be catalyzed by several enzymes: soluble periplasmic sulfite:ferric cytochrome c oxidoreductase, cytoplasmic sulfite:quinone oxidoreductase SoeABC membrane-bound complex. The genes of both SorAB subunits were not identified in the genomes of *B. leptomitiformis* strains, while the genes encoding SoeABC were found and the enzymatic activity of the complex was shown (50–70 nmol × min<sup>-1</sup> × mg protein<sup>-1</sup>) with lithoautotrophic growth of bacteria under microaerobic conditions.

It can be concluded that both strains carry the genes necessary for the dissimilation of sulfur metabolism. However, only strain D-402<sup>T</sup> is capable of lithoheterotrophic and lithoautotrophic growth and an accumulation of elemental sulfur inside cells in the presence of thiosulfate or hydrogen sulfide (Figure 1) [20]. Strain D-401 is not able to use reduced sulfur compounds as an electron donor for energy metabolism, it shows only organoheterotrophic growth.

Moreover, a comparative analysis of the coding and promoter regions of these genes did not reveal any differences between strain D-402<sup>T</sup> and D-401. It can be suggested that other genes associated with the dissimilation of sulfur metabolism can also be changed.

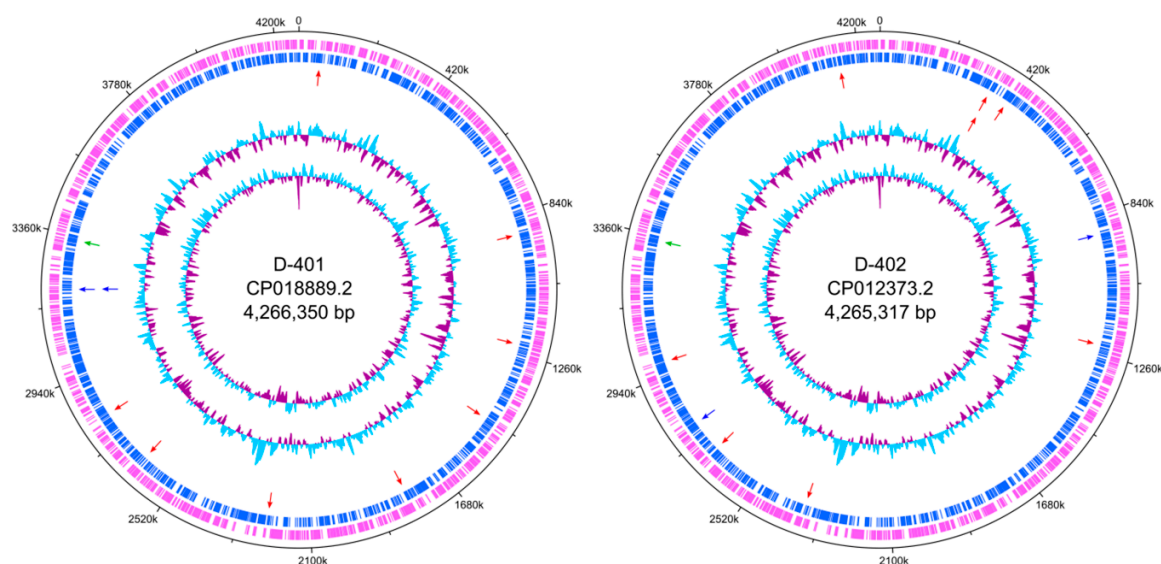




**Figure 1.** Phase contrast micrographs showing morphology of *Beggiatoa leptomitiformis* D-401 (a) and D-402<sup>T</sup> (b) when growing on a medium with thiosulfate and lactate. Arrows indicate elemental sulfur (S<sup>0</sup>) inclusions in D-402<sup>T</sup> cells.

### 3.2. Comparison of Complete Genomes

In order to identify any genetic differences that cause a physiology contrast between the strains, we performed a whole-genome alignment. There are no large rearrangements in the genomes. However, we identified 16 long inserts: eight inserts were found in the genome of strain D-402<sup>T</sup> and eight inserts were found in the genome of strain D-401. These inserts can be divided into several types by length: 10 inserts with length equal to 853 bp, four inserts with length equal to 782 bp, one insert with length equal to 378 bp, and one insert with length equal to 11 bp. We also found four one-letter indels and one one-letter substitution. Identified loci that distinguish genomes do not form clusters and are evenly distributed over the genomes (Figure 2, Supplementary Tables S1 and S2).



**Figure 2.** Map of differences between *B. leptomitiformis* D-401 and D-402<sup>T</sup> genomes. Designations from the outer circle to the inner one: circle 1 and 2—predicted open reading frames in forward and reverse orientation, respectively, 3—long inserts (red), single-letter inserts (blue), single-letter substitution (green), 4—GC-plot, showing deviations from the mean value, 5—GC-skew. Coordinates are specified in bp.

### 3.3. Long Inserts

Cluster analysis showed that long inserts can be divided into three separate types (Supplementary Figure S1). Within each type of sequence, a high percent identity (>97%) is shown. Interestingly, long inserts of the second type are found only in strain D-401. We found direct and inverted repeats at the ends of sequences of types 1 and 2 (Supplementary Table S3), identical within each type which indicates that these sequences are transposons. For sequences of type 3, repeats were not found.

A search for transposons by the exact match of the inverted repeats in the genome of strain D-402<sup>T</sup> additionally revealed 34 sequences of the first type and 12 sequences of the second type, and for strain D-401 we identified 32 and 16 sequences, respectively. Among them, two transposons of greater length (1731 and 1660 bp) were identified, belonging to type 1 and 2, respectively. These sequences found in the genome of strain D-402<sup>T</sup> are completely identical to the sequences found in the genome of strain D-401. The full-sized sequences of each type contain four protein-coding genes (Supplementary Tables S4 and S5) which are various transposases and nucleases. This further confirms that these sequences are transposons. Thus, strains D-402<sup>T</sup> and D-401 have at least two types of transposons (types 1 and 2), and the sequence of type 3 is not a transposon.

### 3.4. Candidate Genes

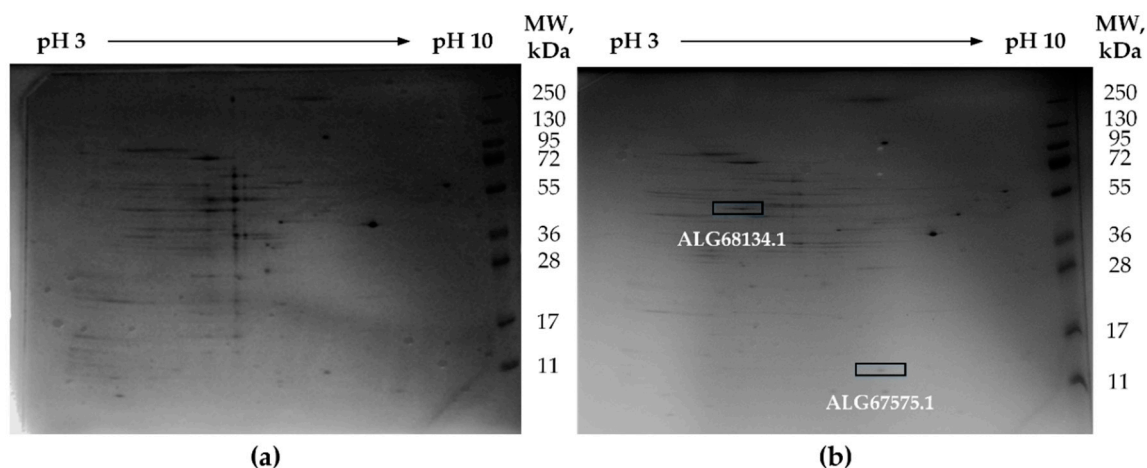
The one-letter substitution falls into the coding region of the gene which encodes the response regulator, but it does not change the amino acid sequence of the protein. Therefore, we did not consider this difference as significant. Single-letter indels lead to frameshifts in the genes encoding the following putative proteins: Bama/TamA family outer membrane protein, long-chain acyl-CoA synthetase, and transposase. We suppose that these genes do not play a significant role in the lithotrophy processes in *B. leptomitiformis*. Three transposons found in the genome of strain D-401 fall into coding sequences and can disrupt the functioning of corresponding genes.

The first and most likely gene can encode the protein of the sulfur globules envelope (ALG67575.1). In representatives of the family Beggiatoaceae, elemental sulfur is deposited in cytoplasmic invaginations and is surrounded by a specific protein envelope. A 15-kDa protein was found around sulfur inclusions in *Beggiatoa alba* B18LD, when strain cultivated in medium with sulfide [21,22].

This class of proteins generally does not contain markable conserved domains. However, the sequence of the found protein enriched with regularly spaced prolines contains a cleavable N-terminal peptide necessary for transport to the periplasmic space. These features correspond to those for the previously described envelope proteins [23]. Similar genes were also found in phylogenetically close neighbors of *B. leptomitiformis* D-402<sup>T</sup>—*B. alba* B18LD and *Thioflexithrix pseukupensis* D3 (Supplementary Figure S2).

The two other genes are hypothetical proteins. One of them can encode porin (ALG68134.1) which can form channels in the outer membrane through which small molecules, such as thiosulfate are able to penetrate. Thus, this gene can play a role in lithotrophic growth processes, acting as a transmembrane transporter. Another gene encodes a protein (ALG67202.2) containing a sensory domain with an unknown function.

To show the lack of full-sized proteins of the found genes in strain D-401, a proteomic analysis of the proteins of both strains which were cultivated in the presence of lactate and thiosulfate was performed. During the study of proteomic maps by the method of two-dimensional gel electrophoresis, differences in protein expression profiles were revealed (Figure 3). In strain D-401, none of the claimed proteins were found. We identified and analyzed proteins that corresponded to the calculated values of the molecular weights of the studied proteins without and with inserts. Mass-spectrometric analysis allowed us to identify peptides in the D-402<sup>T</sup> sample corresponding to the full-size proteins of the sulfur globules envelope (ALG67575.1) and the hypothetical porin protein (ALG68134.1) (Supplementary Table S6). Protein ALG67202.2 was not identified in the sample of strain D-402<sup>T</sup>. The reason may be the low representation of the corresponding protein in the sample.



**Figure 3.** Protein expression profiles in *B. leptomitiformis* strains D-401 (a) and D-402<sup>T</sup> (b) after two-dimensional gel electrophoresis on proteomic maps. Cultivation in the presence of thiosulfate and lactate.

Additionally, for the analysis of proteins of interest, a more sensitive approach for proteomic analysis was used—high performance liquid chromatography in combination with tandem mass spectrometry. During the analysis of strain D-401 by HPLC-MS/MS, 847 proteins were reliably identified. When analyzing proteins of strain D-402<sup>T</sup>, 606 protein sequences were reliably identified, among which proteins ALG68134.1 (identifier in TrEMBL UPI0007069FE2) and ALG67575.1 (identifier in TrEMBL UPI00070625ED) were found.

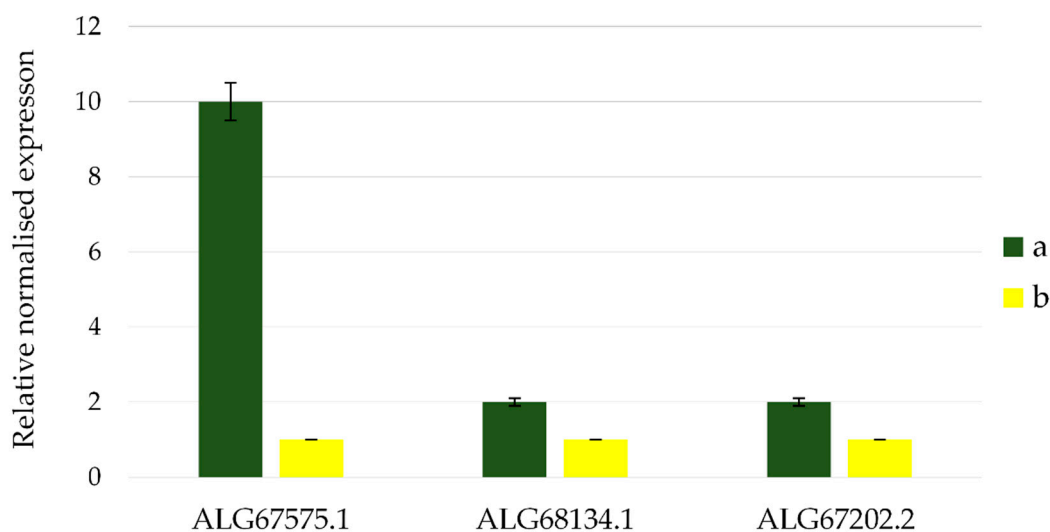
Supplementary Figure S3 shows the protein sequences of candidate genes, where the peptides which were identified by MALDI mass spectrometry and HPLC-MS/MS (LC-MS/MS) are highlighted in bold and underlined, respectively. Supplementary Figure S3a shows that the protein ALG68134.1 is fairly well covered with peptides identified by HPLC-MS/MS which fully include the sequences of peptides detected by MALDI mass spectrometry. The sequence coverage of this protein is 71.6%. For protein ALG67575.1 (Supplementary Figure S3b), the peptide in the N-terminal region was detected only by MALDI mass spectrometry. The C-terminal region of the protein is covered with peptides which were detected using both mass-spectrometric approaches, and the total protein coverage of ALG67575.1 is 40.4%.

### 3.5. Candidate Genes Expression

To determine the involvement of the proteins ALG67575.1, ALG68134.1, and ALG67202.2 in the oxidation processes of reduced sulfur compounds in the strain D-402<sup>T</sup>, we studied expression of genes coding for these proteins in the presence or absence of thiosulfate in the culture medium.

As a result of RT-PCR, it was shown that the presence of thiosulfate in the medium does not affect the expression of genes of proteins ALG67202.2 and ALG68134.1. In the case of the gene which encodes the sulfur globules envelope protein (ALG67575.1), it was found that the addition of thiosulfate to the medium increases gene expression by 10 times compared with the growth without thiosulfate (Figure 4).





**Figure 4.** The expression of genes coding for ALG67575.1, ALG68134.1, and ALG67202.2 proteins of *B. leptomitiformis* strain D-402<sup>T</sup> during lithoheterotrophic (a) and organotrophic (b) cultivation.

#### 4. Discussion

The inability of the *B. leptomitiformis* strain D-401, in contrast to the D-402<sup>T</sup> strain, to grow lithotrophically in the presence of reduced sulfur compounds may suggest the existence of mutations in the genes related to dissimilatory sulfur metabolism. However, we did not find any differences in the genes of sulfur metabolism between strain D-401 and D-402<sup>T</sup> in both coding and promoter regions. However, in strain D-401, the presence of a transposon in the gene that encodes sulfur globules envelope protein was found which leads to morphological and physiological differences between strains D-401 and D-402<sup>T</sup>.

For some sulfur bacteria capable of accumulating elemental sulfur intracellularly, electron micrographs clearly show that sulfur globules form in the periplasmic space [22,24]. In these organisms, sulfur globules are surrounded by a protein envelope, for example, in representatives of the genus *Thiothrix* [25], purple sulfur bacteria [21,26–28], *Beggiatoa* [29], and *Thiovulum* [30].

Today, little is known about the protein of the sulfur globules envelope in representatives of the genus *Beggiatoa*. The vast majority of what is known regarding sulfur globules envelope comes from studies of the purple sulfur bacteria of the Chromatiaceae family [21,26–28]. In *Allochrochromatium vinosum*, the sulfur envelope is represented by three proteins (SgpA, SgpB, SgpC) [31]. The sulfur envelope of *Thiocapsa roseopersicina* contains two proteins of 10.7 and 8.7 kDa, both of which are homologous to large and small proteins of *A. vinosum* [32]. The presence of a single protein of 18.5 kDa of the sulfur globule in *Isochromatium buderi* [33] and the 15-kDa protein in *Beggiatoa* was found when bacteria were grown in a medium supplemented with sulfide [22]. Today, there is no clarity regarding the subunit structure of the protein of the sulfur globules envelope from *I. buderi* and *Beggiatoa*.

Our proteomic data confirm the presence of a protein of 15 kDa which performs the function of a sulfur globules envelope in *B. leptomitiformis* D-402<sup>T</sup> and confirm its absence in strain D-401. Moreover, for strain D-402<sup>T</sup>, we have shown that the addition of thiosulfate to the culture medium increases the gene expression of this protein by 10 times compared with the cultivation without thiosulfate. Thus, obtained data allow us to assume that the protein ALG67575.1 is strictly necessary for lithotrophic growth in the presence of reduced sulfur compounds. Our data are corresponding to experiments with the purple sulfur bacterium *Allochrochromatium vinosum* [31]. The sulfur envelope proteins in *A. vinosum* are encoded by three genes *sgpABC*. It was shown that *sgpBC* double mutant completely lost the ability to use H<sub>2</sub>S as an electron donor and accumulate elemental sulfur inside the cells.

Comparison of biotopes where *Beggiatoa leptomitiformis* D-401 and D-402<sup>T</sup> strains develop shows that strain D-401 lives in silt sediments of treatment facilities (Yaroslavl Oblast, Russia), where the

concentration of H<sub>2</sub>S is not high (0.5–1.0 mg × L<sup>-1</sup>), while strain D-402<sup>T</sup> was isolated from sulfur mat formed on irrigation fields contaminated with residential and agricultural wastewater (Moscow Region, Russia), where the concentration of H<sub>2</sub>S is 2–5 mg × L<sup>-1</sup>.

The incorporation of transposon into the corresponding gene in D-401 led to the “shutdown” of the lithotrophy process in the presence of reduced sulfur compounds which, in turn, allowed D-401 to develop in a biotope with a lower H<sub>2</sub>S and other bioenergetics nutrient content (organic compounds) than the biotope of D-402<sup>T</sup>.

It is considered that transposons can easily leave genes and move within the genome. However, traces of ‘cutting out’ transposons were not revealed in the studied strains, which indicate genome stability of the strains D-401 and D-402<sup>T</sup>, and we have been observing their properties for more than 35 years.

Along with our research, further study of representatives of the genus *Beggiatoa* could have far-reaching implications in environmental microbiology and biogeochemical cycling.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2076-2607/8/6/928/s1>, Figure S1: Dendrogram of 15 long inserts found in the genomes of *B. leptomitiformis* strains D-401 and D-402<sup>T</sup>, Figure S2: The protein sequences which are supposed to be a sulfur globule protein. N-terminal transport peptides are marked blue; prolines are marked yellow, Figure S3: Covering the protein sequence of ALG68134.1 (a) and ALG67575.1 (b) with identified peptides. Peptides which were identified by HPLC-MS/MS are shown in bold; peptide sequences which were identified by MALDI mass spectrometry are underlined, Table S1: Inserts of *B. leptomitiformis* strain D-401, Table S2: Inserts of *B. leptomitiformis* strain D-402<sup>T</sup>, Table S3: Inverted and direct repeats at the ends of long inserts, Table S4: Genes in a full-length transposon type 1 sequence, Table S5: Genes in a full-length transposon type 2 sequence, Table S6: Identification results of *B. Leptomitiformis* proteins by MALDI mass spectrometry.

**Author Contributions:** Conceptualization, T.S.R., S.V.T., and M.Y.G.; methodology, S.V.T.; software, S.V.T.; visualization, T.S.R. and S.V.T.; formal analysis, T.S.R., S.V.T., N.D.S., and M.Y.G.; investigation, T.S.R., S.V.T., and N.D.S.; project administration, M.Y.G; writing—original draft preparation, T.S.R., S.V.T., and M.Y.G.; writing—review and editing, T.S.R., S.V.T., and M.Y.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the Russian Foundation for Basic Research [grant 18-04-00556].

**Acknowledgments:** HPLC-MS/MS analysis was performed using the equipment of the Centre of Collective Usage “Human proteome” (IBMC, Moscow).

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

## References

- Nelson, D.C.; Jannasch, H.W. Chemoautotrophic growth of a marine *Beggiatoa* in sulfide-gradient cultures. *Arch. Microbiol.* **1983**, *136*, 262–269. [[CrossRef](#)]
- Strohl, W.R.; Larkin, J.M. Enumeration, isolation, and characterization of *Beggiatoa* from freshwater sediments. *Appl. Environ. Microbiol.* **1978**, *36*, 755–770. [[CrossRef](#)]
- Grabovich, M.Y.; Patrikskaya, V.Y.; Muntyan, M.S.; Dubinina, G.A. Lithoautotrophic growth of the fresh-water strain *Beggiatoa* D-402 and energy conservation in a homogeneous culture under microoxic conditions. *FEMS Microbiol. Lett.* **2001**, *204*, 341–345. [[CrossRef](#)] [[PubMed](#)]
- Nelson, D.C.; Castenholz, R.W. Use of reduced sulfur compounds by *Beggiatoa* sp. *J. Bacteriol.* **1981**, *147*, 140–154. [[CrossRef](#)] [[PubMed](#)]
- Fomenkov, A.; Sun, Z.; Vincze, T.; Dubinina, G.; Orlova, M.; Tarlachkov, S.V.; Anton, B.P.; Grabovich, M.Y.; Roberts, R.J. Complete genome sequence of the freshwater bacterium *Beggiatoa leptomitiformis* strain D-401. *Genome Announc.* **2018**, *6*, e00311-18. [[CrossRef](#)]
- Fomenkov, A.; Vincze, T.; Grabovich, M.Y.; Dubinina, G.; Orlova, M.; Belousova, E.; Roberts, R.J. Complete genome sequence of the freshwater colorless sulfur bacterium *Beggiatoa leptomitiformis* neotype strain D-402<sup>T</sup>. *Genome Announc.* **2018**, *3*, e01436-15. [[CrossRef](#)]
- Pfennig, N.; Lippert, K. Über das vitamin B12-bedurfuis phototropher Schwefelbakterien. *Arch. Mikrobiol.* **1966**, *55*, 245–256. [[CrossRef](#)]
- Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)]

9. Langmead, B.; Salzberg, S. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [[CrossRef](#)]
10. Walker, B.J.; Abeel, T.; Shea, T.; Priest, M.; Abouelliel, A.; Sakthikumar, S.; Cuomo, C.A.; Zeng, Q.; Wortman, J.; Young, S.K.; et al. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **2014**, *9*, e112963. [[CrossRef](#)]
11. Tamura, K.; Stecher, G.; Peterson, D.; Filipski, A.; Kumar, S. MEGA 6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **2013**, *30*, 2725–2729. [[CrossRef](#)] [[PubMed](#)]
12. Thompson, J.D.; Higgins, D.G.; Gibson, T.J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673–4680. [[CrossRef](#)] [[PubMed](#)]
13. Saitou, N.; Nei, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **1987**, *4*, 406–425. [[CrossRef](#)] [[PubMed](#)]
14. Denger, K.; Weinitschke, S.; Smits, T.H.; Schleheck, D.; Cook, A.M. Bacterial sulfite dehydrogenases in organotrophic metabolism: Separation and identification in *Cupriavidus necator* H16 and in *Delftia acidovorans* SPH-1. *Microbiology* **2008**, *154*, 256–263. [[CrossRef](#)] [[PubMed](#)]
15. O'Farrell, P.H. High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **1975**, *250*, 4007–4021. [[PubMed](#)]
16. Anselm, V.; Novikova, S.; Zgoda, V. Re-adaption on earth after spaceflights affects the mouse liver proteome. *Int. J. Mol. Sci.* **2017**, *18*, 1763. [[CrossRef](#)] [[PubMed](#)]
17. Barsnes, H.; Vaudel, M. SearchGUI: A highly adaptable common interface for proteomics search and de novo engines. *J. Proteome Res.* **2018**, *17*, 2552–2555. [[CrossRef](#)] [[PubMed](#)]
18. Vaudel, M.; Burkhart, J.M.; Zahedi, R.P.; Oveland, E.; Berven, F.S.; Sickmann, A.; Martens, L.; Barsnes, H. PeptideShaker enables reanalysis of MS-derived proteomics data sets. *Nat. Biotechnol.* **2015**, *33*, 22–24. [[CrossRef](#)] [[PubMed](#)]
19. Weirather, J.L.; de Cesare, M.; Wang, Y.; Piazza, P.; Sebastiano, V.; Wang, X.J.; Buck, D.; Au, K.F. Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Research* **2017**, *6*, 100. [[CrossRef](#)]
20. Dubinina, G.; Savvichev, A.; Orlova, M.; Gavrish, E.; Verbarq, S.; Grabovich, M. *Beggiatoa leptomitiformis* sp. nov., the first freshwater member of the genus capable of chemolithoautotrophic growth. *Int. J. Syst. Evol. Micr.* **2017**, *67*, 197–204. [[CrossRef](#)]
21. Nicolson, G.L.; Schmidt, G.L. Structure of the *Chromatium* sulfur particle and its protein membrane. *J. Bacteriol.* **1971**, *105*, 1142–1148. [[CrossRef](#)] [[PubMed](#)]
22. Schmidt, M.T.; Vinci, A.V.; Strohl, R.W. Protein synthesis by *Beggiatoa alba* B18LD in the presence and absence of sulfide. *Arch. Microbiol.* **1986**, *144*, 158–162. [[CrossRef](#)]
23. Dahl, C. Sulfur metabolism in phototrophic bacteria. In *Modern Topics in the Phototrophic Prokaryotes*; Hallenbeck, P.C., Ed.; Springer: Cham, Switzerland, 2017; pp. 27–66.
24. Schmidt, M.T.; DiSpirito, A.A. Spectral characterization of c-type cytochromes purified from *Beggiatoa Alba*. *Arch. Microbiol.* **1990**, *154*, 453–458. [[CrossRef](#)]
25. Bland, J.A.; Staley, J.T. Observations on the biology of *Thiothrix*. *Arch. Microbiol.* **1978**, *117*, 79–87. [[CrossRef](#)]
26. Remsen, C.C. Comparative subcellular architecture of photosynthetic bacteria. In *The Photosynthetic Bacteria*; Clayton, R.K., Sistrom, W.R., Eds.; Plenum Press: New York, NY, USA, 1978; pp. 31–60.
27. Remsen, C.C. Structural attributes of membranous organelles in bacteria. *Int. Rev. Cytol.* **1982**, *76*, 195–223. [[CrossRef](#)] [[PubMed](#)]
28. Remsen, C.C.; Trüper, H.G. The fine structure of *Chromatium buderi*. *Arch. Mikrobiol.* **1973**, *90*, 269–280. [[CrossRef](#)] [[PubMed](#)]
29. Strohl, W.R.; Geffers, I.; Larkin, J.M. Structure of the sulfur inclusion envelopes from four *Beggiatoas*. *Curr. Microbiol.* **1981**, *6*, 75–79. [[CrossRef](#)]
30. Wirsén, C.O.; Jannasch, H.W. Physiological and morphological observations on *Thiovulum* sp. *J. Bacteriol.* **1978**, *136*, 765–774. [[CrossRef](#)]
31. Prange, A.; Engelhardt, H.; Trüper, H.G.; Dahl, C. The role of the sulfur globule proteins of *Allochromatium vinosum*: Mutagenesis of the sulfur globule protein genes and expression studies by real-time RT-PCR. *Arch. Microbiol.* **2004**, *182*, 165–174. [[CrossRef](#)]

32. Brune, D.C. Isolation and characterization of sulfur globule proteins from *Chromatium vinosum* and *Thiocapsa roseopersicina*. *Arch. Microbiol.* **1995**, *163*, 391–399. [[CrossRef](#)]
33. Imhoff, J.F.; Süling, J.; Petri, R. Phylogenetic relationships among the *Chromatiaceae*, their taxonomic reclassification and description of the new genera *Allochromatium*, *Halochromatium*, *Isochromatium*, *Marichromatium*, *Thiococcus*, *Thiohalocapsa* and *Thermochromatium*. *Int. J. Syst. Evol. Micr.* **1998**, *48*, 1129–1143. [[CrossRef](#)] [[PubMed](#)]



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).