microorganisms

MDPI

Article

Bioinformatics Investigations of Universal Stress Proteins from Mercury-Methylating *Desulfovibrionaceae*

Raphael D. Isokpehi [1,*], Dominique S. McInnis [1], Antoinette M. Destefano [1], Gabrielle S. Johnson [1], Akimio D. Walker [1], Yessenia A. Hall [1], Baraka W. Mapp [1], Matilda O. Johnson [2] and Shaneka S. Simmons [3]

[1] College of Science, Engineering and Mathematics, Bethune-Cookman University, Daytona Beach, FL 32114, USA; smithmcinnisd@cookman.edu (D.S.M.); antoinette.m.destefano@students.cookman.edu (A.M.D.); gabrielle.s.johnson@students.cookman.edu (G.S.J.); akimio.d.walker@students.cookman.edu (A.D.W.); yessenia.a.hall@students.cookman.edu (Y.A.H.); williamsbaraka@gmail.com (B.W.M.)

[2] College of Nursing and Health Sciences, Bethune-Cookman University, Daytona Beach, FL 32114, USA; johnsonma@cookman.edu

[3] Department of Science and Mathematics, Jarvis Christian College, Hawkins, TX 75765, USA; ssimons@jarvis.edu

* Correspondence: isokpehir@cookman.edu

Abstract: The presence of methylmercury in aquatic environments and marine food sources is of global concern. The chemical reaction for the addition of a methyl group to inorganic mercury occurs in diverse bacterial taxonomic groups including the Gram-negative, sulfate-reducing *Desulfovibrionaceae* family that inhabit extreme aquatic environments. The availability of whole-genome sequence datasets for members of the *Desulfovibrionaceae* presents opportunities to understand the microbial mechanisms that contribute to methylmercury production in extreme aquatic environments. We have applied bioinformatics resources and developed visual analytics resources to categorize a collection of 719 putative universal stress protein (USP) sequences predicted from 93 genomes of *Desulfovibrionaceae*. We have focused our bioinformatics investigations on protein sequence analytics by developing interactive visualizations to categorize *Desulfovibrionaceae* universal stress proteins by protein domain composition and functionally important amino acids. We identified 651 *Desulfovibrionaceae* universal stress protein sequences, of which 488 sequences had only one USP domain and 163 had two USP domains. The 488 single USP domain sequences were further categorized into 340 sequences with ATP-binding motif and 148 sequences without ATP-binding motif. The 163 double USP domain sequences were categorized into (1) both USP domains with ATP-binding motif (3 sequences); (2) both USP domains without ATP-binding motif (138 sequences); and (3) one USP domain with ATP-binding motif (21 sequences). We developed visual analytics resources to facilitate the investigation of these categories of datasets in the presence or absence of the mercury-methylating gene pair (*hgcAB*). Future research could utilize these functional categories to investigate the participation of universal stress proteins in the bacterial cellular uptake of inorganic mercury and methylmercury production, especially in anaerobic aquatic environments.

Keywords: aquatic environments; anaerobic bacteria; bioinformatics; biofilms; *Desulfovibrionaceae*; genomes; mercury; methylation; *Pseudodesulfovibrio mercurii*; stress response; universal stress protein

Citation: Isokpehi, R.D.; McInnis, D.S.; Destefano, A.M.; Johnson, G.S.; Walker, A.D.; Hall, Y.A.; Mapp, B.W.; Johnson, M.O.; Simmons, S.S. Bioinformatics Investigations of Universal Stress Proteins from Mercury-Methylating *Desulfovibrionaceae*. *Microorganisms* **2021**, *9*, 1780. https://doi.org/10.3390/microorganisms9081780

Academic Editors: Sylvie Chevalier and Pierre Cornelis

Received: 16 June 2021
Accepted: 6 August 2021
Published: 21 August 2021

1. Introduction

Mercury is a trace metal, which in both its organic (methyl mercury) and elemental form (Hg) is known to be highly toxic to all life forms [1,2]. Exposure to mercury can occur through inhalation of toxic elemental mercury vapors [3], through dietary sources and non-dietary sources [4,5]. The presence of methylmercury in aquatic environments and marine food sources is of global concern [2,6]. In the United States, mercury-impaired waterbodies have concentrations of mercury in fish tissue that have exceeded 1.0 mg/kg

total mercury [7–10]. The chemical reaction for the addition of a methyl group to inorganic mercury occurs in diverse bacterial taxonomic groups including the Gram-negative, sulfate-reducing *Desulfovibrionaceae* family of the delta subdivision of proteobacteria [11–13]. The genera in the *Desulfovibrionaceae* family include *Bilophila*, *Desulfobaculum*, *Desulfocurvibacter*, *Desulfocurvus*, *Desulfohalovibrio*, *Desulfovibrio*, *Halodesulfovibrio*, *Humidesulfovibrio*, *Lawsonia*, and *Pseudodesulfovibrio* [14]. The availability of whole-genome sequence datasets for some members of the *Desulfovibrionaceae* [15–17] presents opportunities to understand the microbial mechanisms that contribute to methylmercury production in water bodies.

The genomes of bacteria that are able to methylate mercury have a two-gene cluster, *hgcA* and *hgcB*, respectively encoding a corrinoid protein and a ferredoxin [12,18]. Bacteria containing the *hgcAB* gene pair occur in a wide range of habitats including extreme natural environments such as coastal dead zones, deep-sea anaerobic sediments, thawing permafrost soils, and hypersaline ecosystems [19]. A list of *Desulfovibrionaceae* genomes predicted to be mercury methylators according to the presence of the *hgcAB* gene pair are available on the data page of the Biogeochemical Transformations at Critical Interfaces project of the Oak Ridge National Laboratory's Mercury Science Focus Area [12,20]. We are interested in genes that encode the universal stress protein (USP) domain (Protein Family (Pfam) Identifiers: Usp, pfam00582 or PF00582), since they aid bacteria in (1) responding to extreme conditions; and (2) the formation as well as maintenance of adherent bacteria communities termed biofilms [21–24]. Biofilms can methylate mercury (Hg) at higher rates than unattached bacteria and are a location for mercury methylation in the environment [11]. The USP gene count per genome has not been compiled for the *Desulfovibrionaceae* genomes to enable comparisons between genomes of mercury methylators and those that are not mercury methylators. This research article bridges the knowledge gap on USP gene content of the *Desulfovibrionaceae* genomes.

The universal stress proteins can be composed of one USP domain; two USP domains in tandem; or one or two USP domains together with other functional domains including transporters, kinases, permeases, transferases, and bacterial sensor proteins [23,25]. The three-dimensional structure of universal stress proteins provides evidence for associated molecular functions, biological processes and cellular components. Adenosine-5′-triphosphate (ATP) functions as coenzyme as well as energy molecule [26] and its binding to USPs provides a basis for the functional categorization of USPs [27]. The ATP-binding amino acid motif of G2XG9XG(S/T) categorizes the USP domain into two groups: ATP-binding and non-ATP-binding [27,28]. The categorization of USP domains of the mercury-methylating *Desulfovibrionaceae* will allow for the new bioinformatics investigations and the design of experiments to determine the participation of USPs in bacterial mercury methylation.

Genes for universal stress proteins were predicted from the genome sequencing of *Desulfovibrionaceae* members, including those that methylate mercury and inhabit extreme environments [29–32]. Thus, the aim of the research reported here was to investigate the protein sequence features encoded by the predicted universal stress protein sequences of mercury-methylating *Desulfovibrionaceae*. We have focused our bioinformatics investigations on protein sequence analytics by developing interactive visualizations to categorize *Desulfovibrionaceae* universal stress proteins by protein domain composition and functionally important amino acids (functional sites).

We applied bioinformatics resources and developed visual analytics resources to categorize a collection of 719 putative universal stress proteins predicted from 93 genomes of *Desulfovibrionaceae*. We identified a subset of 651 *Desulfovibrionaceae* universal stress protein sequences into 488 sequences with one USP domain and 163 with two USP domains. Additionally, the sequences were categorized by (1) the presence of ATP-binding functional sites and (2) the presence of mercury methylation gene pair in the bacterial genome. The findings provide foundations to investigate the participation of universal stress proteins in the bacterial cellular uptake of inorganic mercury and methylmercury production, especially in anaerobic aquatic environments.

2. Materials and Methods

2.1. Overview—Applying Bioinformatics Resources and Developing Visual Analytics Resources

The flowchart describing the stages of the bioinformatics investigations is presented in Figure 1. The U.S. Department Joint Genome Institute's (JGI) Integrated Microbial Genomes and Microbiomes (IMG/M) system [33] was the key bioinformatics resource for collecting and interacting with protein sequence data. We also applied the Batch Web Conserved Domain Search (CD-Search) Tool of the National Center for Biotechnology Information (NCBI) [34] to obtain the number and the protein domain composition as well as the amino acid functional sites.

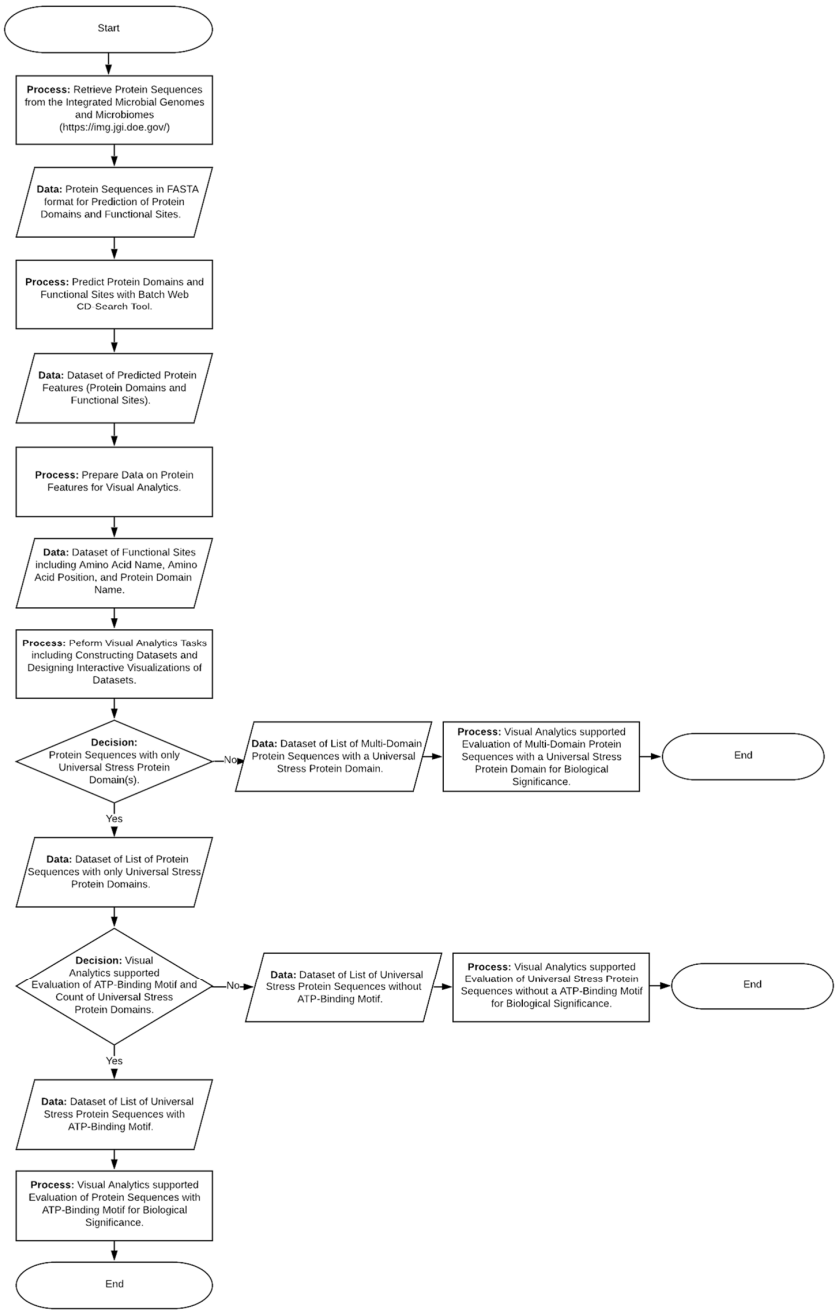


Figure 1. Overview of bioinformatics data investigations of universal stress proteins relevant to bacterial mercury methylation. The process integrates bioinformatics resources and visual analytics resources to categorize universal stress proteins by protein features such as protein domain composition (count and type) as well as the presence of the ATP-binding motif.

We typically constructed the results from bioinformatics tasks into datasets that serve as data sources for visual analytics tasks [35]. Bioinformatics tasks that we performed include searching for genes with specific annotation as well as predicting the conserved domains and functional amino acids. The visual analytics tasks include designing interfaces to support interaction, analysis and representation of datasets from bioinformatics tasks [35]. We implemented interactive visualizations (visual representations) in version 2020.4 of Tableau (Tableau, Seattle, WA, USA), a visual analytics software The framework for interaction design for complex cognitive activities with visual representations guided our designs of the interactive visual representations [36,37]. This interaction design framework defines the type of visualizations (e.g., enclosure tables, box plots, and bar plots) and action patterns (e.g., filtering, selecting and transforming) that promotes complex cognitive activities such as decision making, planning, knowledge discovery and understanding [37].

2.2. Retrieval of Genome List, Gene List and Protein Sequences annotated with Universal Stress Protein Domain

We applied the Find Genomes and the Find Function tools of the IMG/M system to retrieve, respectively, lists of *Desulfovibrionaceae* genomes and genes annotated with pfam00582. We exported the genome lists and gene lists with annotations from IMG/M into text files for visual analytics tasks. The retrieval of the genome list and gene list in IMG/M generates an Analysis Cart that includes functionalities for exporting protein sequences (in FASTA format). A text file with protein sequences predicted from genes was the input to the Batch Web Conserved Domain Search (CD-Search) Tool of the National Center for Biotechnology Information (NCBI) [34].

2.3. Prediction of Protein Domain Composition and Functional Amino Acid Sites

According to the amino acid sequence of the ATP-binding universal stress protein (MJ0577) of *Methanocaldococcus jannaschii*, there are 12 functional sites where amino acids contact the ATP molecule [38]. Thus, we submitted to a bioinformatics resource (NCBI Web Batch CD-Search Tool) a text file containing FASTA formatted amino acid sequences of the *Desulfovibrionaceae* proteins predicted by the IMG/M system to contain the pfam00582 (Usp) domain. We also submitted to the NCBI Web Batch CD-Search Tool a set of 3470 protein sequences predicted from the genome of *Desulfovibrio desulfuricans* ND132. This additional prediction approach could identify potential universal stress proteins that we did not retrieve with the IMG/M pfam00582 function keyword search. It also demonstrates that our categorization process by functional features can be applied beyond the universal stress protein family. The results generated for the protein sequences were Domain hits, Align details, and Features. We downloaded the Features into a file and removed the comment section such that the dataset on functional sites is in a tab-delimited file ready as input for visual analytics.

The data fields in the Features file are (1) Query (obtained from FASTA header); (2) Type of protein domain (e.g., specific or superfamily); (3) Title (e.g., Ligand-Binding Site); (4) coordinates (amino acid and position, e.g., P9, V10, D11, C39, M108, G109, R111, G112, G122, S123, V124, T125); (5) complete size (the expected number of functional sites, e.g., 12); (6) mapped size (observed functional sites, e.g., 12); and (7) source domain (protein domain source of functional sites, e.g., 23,812 for the Usp domain). The data file has a record for each protein domain present in the sequence. Thus, it was possible to identify sequences with more than one protein domain including the tandem-type Usp domains. We constructed patterns from the coordinates to facilitate tasks on visual representations, interactions and analyses (such as categorizing and comparing sequences) in a visual analytics software. We developed Perl code to extract patterns from the amino acid coordinates. For example, from coordinates "P9, V10, D11, C39, M108, G109, R111, G112, G122, S123, V124, T125", the amino acid pattern "PVDCMGRGGSVT" and the amino acid position pattern "9_10_11_39_108_109_111_112_123_124_125" were extracted. Additional information on the Perl code and application beyond the amino acid sequences of the universal stress proteins is presented in the Appendix A (Figure A1). For comparison and accuracy verification

of the ATP-binding motif detection procedure, we extracted patterns from the sequences of 10 universal stress proteins from *Mycobacterium tuberculosis* (lab strain H37Rv), whose USPs were extensively investigated for ATP-binding capacity. We performed scripting tasks on computing hardware including a large memory computer cluster (carbonate.uits.iu.edu) configured to support high-performance, data-intensive computing at the National Center Genome Analysis Support (NCGAS), Indiana University [39].

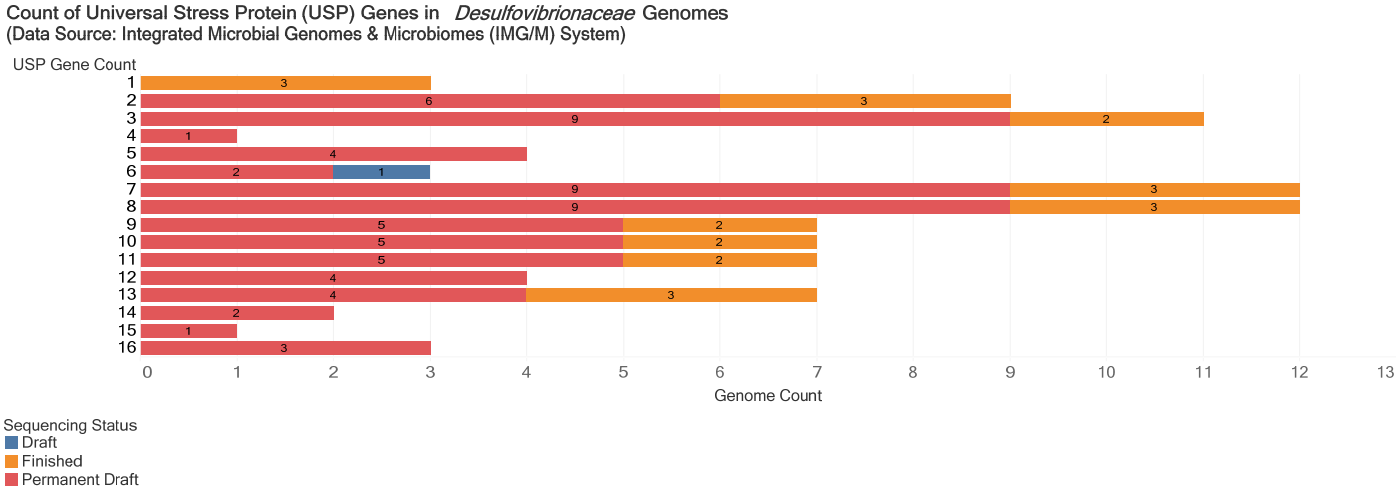
3. Results

3.1. Count of Universal Stress Protein Genes in Desulfovibrionaceae Genomes

Our search on the Integrated Microbial Genomes and Microbiomes (IMG/M) system for genes encoding the universal stress protein domain (pfam00582) retrieved 716 genes from 93 *Desulfovibrionaceae* genomes. The genera represented in the dataset are *Bilophila*, *Desulfobaculum*, *Desulfocurvibacter*, *Desulfocurvus*, *Desulfohalovibrio*, *Desulfovibrio*, *Halodesulfovibrio*, *Lawsonia*, *Mailhella*, and *Pseudodesulfovibrio*. According to the sequencing status, there were 23 finished genomes, 69 permanent draft genomes and 1 draft genome.

Based on automated Pfam entry annotation, the observed counts of USP gene per genome ranged from 1 to 16. An example genome in each count type is: 1 (*Lawsonia intracellularis* N343); 2 (*Desulfovibrio desulfuricans desulfuricans* ATCC 27774); 3 (*Desulfovibrio fairfieldensis* CCUG 45958); 4 (*Desulfovibrio cuneatus* DSM 11391); 5 (*Desulfovibrio frigidus* DSM 17176); 6 (*Desulfovibrio vexinensis* DSM 17965); 7 (*Desulfovibrio salexigens* DSM 2638); 8 (*Pseudodesulfovibrio piezophilus* C1TLv30); 9 (*Pseudodesulfovibrio aespoeensis* Aspo-2, DSM 10631); 10 (*Desulfovibrio magneticus* RS-1); 11 (*Desulfovibrio gigas* DSM 1382, ATCC 19364); 12 (*Desulfovibrio paquesii* DSM 16681); 13 (*Desulfovibrio vulgaris vulgaris* Hildenborough); 14 (*Halodesulfovibrio marinisediminis* DSM 17456), 15 (*Desulfovibrio alcoholivorans* DSM 5433); and 16 (*Halodesulfovibrio aestuarii aestuarii* ATCC 29578). We observed strains with two genome sequencing projects: *Desulfovibrio alkalitolerans* DSM 16529 (8 USP genes); *Desulfovibrio gigas* DSM 1382, ATCC 19364 (11 USP genes); *Desulfovibrio hydrothermalis* AM13, DSM 14728 (5 USP genes); and *Pseudodesulfovibrio indicus* J2 (7 USP genes).

Figure 2 is an overview of the distribution of the USP genes in *Desulfovibrionaceae* genomes according to sequencing status and USP gene count. Our bioinformatics investigation for protein domain composition of 3407 protein sequences from *Desulfovibrio desulfuricans* ND132, a bacterial mercury methylation, identified three additional USP genes to make 13 USP genes. Therefore, we collected into a text file 719 FASTA formatted protein sequences annotated to contain the universal stress protein domain.



Grouping of *Desulfovibrionaceae* Genomes by Count of Universal Stress Protein (USP) Genes and Sequencing Status.
The Total Gene Count for each Genome is also shown.

USP Gene Count	Sequencing Status	Genome ID	Genome Name	
1	Finished	637000145	Lawsonia intracellularis PHE/MN1-00	1398
		2521172708	Lawsonia intracellularis N343	1418
		2775506998	Candidatus Desulfovibrio trichonymphae Rs-N31	1399
10	Finished	644736352	Desulfovibrio magneticus RS-1	4760
		2503754015	Desulfovibrio desulfuricans ND132	3534
	Permanent Draft	2524023062	Desulfovibrio aminophilus DSM 12254	3311
		2526164711	Desulfovibrio africanus DSM 2603	4062
		2571042914	Desulfohalovibrio reitneri L21-Syr-AB	3251
		2574179706	Desulfovibrio vietnamensis DSM 10520	3546
		2574180452	Desulfovibrio idahonensis DSM 15450	3290
16	Permanent Draft	2518645586	Halodesulfovibrio aestuarii aestuarii ATCC 29578	3251
		2524614816	Halodesulfovibrio aestuarii DSM 10141	3071
		2585428154	Desulfovibrio desulfuricans aestuarii DSM 17919	3208

Figure 2. The count of universal stress protein (USP) genes in *Desulfovibrionaceae* Genomes. The bar plot shows the distribution of the gene count by genome count categorized by sequencing status (draft, finished and permanent draft) of the genome. The enclosure table shows examples of genomes with particular gene counts. We have shown genomes with 1, 10 and 16 USP gene count types and the associated total gene count. Among the genomes with USP gene count of 10 is *Desulfovibrio desulfuricans* ND132, a model for bacterial mercury methylation.

In Figure 3, we present a comparison of the protein sequence features for six *Desulfovibrionaceae* genomes including from five that encode the gene pair for mercury methylation. *Desulfovibrio africanus*, *Desulfovibrio desulfuricans* ND132 and *Desulfovibrio halophilus* DSM 5663 are mercury-methylating *Desulfovibrionaceae* species. Among the genomes of the mercury-methylating *Desulfovibrio africanus* (reclassified as *Desulfocurvibacter africanus*), there is an additional gene for strain Walvis Bay encoding a 150 aa universal stress protein (Figure 3). Furthermore, strain DSM 2603 has an additional 282 aa universal stress protein. In the case of *Desulfovibrio desulfuricans* ND132, the groups of amino acid lengths (aa) observed are 139, 146, 148, 156, 162, 265, 288, 294, 295, 297, 310 and 629. The *Desulfovibrio gilichinskyi* K3S genome encodes five universal stress proteins including a USP with 630 aa. *Desulfovibrio halopilus* DSM 5663 encodes seven universal stress proteins including two protein sequences with lengths 146 aa and 297 aa) that were also predicted from the genomes of *Desulfovibrio desulfuricans* ND132 and *Desulfovibrio gilichinskyi* K3S.

Figure 3. An overview of the protein sequence features for six *Desulfovibrionaceae* genomes including from five that encode the gene pair for mercury methylation. *Desulfovibrio africanus*, *Desulfovibrio desulfuricans* ND132 and *Desulfovibrio halophilus* DSM 5663 are mercury-methylating *Desulfovibrionaceae* species. The comparison visual reveals differences and commonalities in the amino acid sequence length, gene count and protein family (pfam) annotation that were obtained from the annotation in the Integrated Microbial Genomes and Microbiomes (IMG/M) system. For example, among the mercury-methylating *Desulfovibrio africanus* (reclassified as *Desulfocurvibacter africanus*), there is an additional one gene encoding a 150 aa universal stress protein. The color of the bar plots represents the arrangement of the protein domains. Null means no protein domain predicted in the IMG/M resource.

The website to the interactive version of figures generated from visual analytics software is available in the Supplementary Materials section.

3.2. Protein Domain Composition and Functional Sites of Desulfovibrionaceae Universal Stress Proteins

The results of the NCBI Batch Web Conserved Domain Search (CD-Search) bioinformatics tool for the 719 protein sequences included predictions on the type and position of the functionally relevant amino acid residues as well as the protein domain(s). The four types of protein domains with functional sites were (1) Universal Stress Protein family; (2) USP domain located between the N-terminal sensor domain and C-terminal catalytic domain of Osmosensitive K+ channel histidine kinase family; (3) old yellow enzyme (OYE)-like Flavin Mono-Nucleotide (FMN)-binding domain; and (4) histidine kinase-like ATPase domain. The NCBI CD-Search Tool identified conserved domains in 718 of the 719 sequences submitted.

We identified 651 universal stress protein sequences that have at least one conserved USP domain model (Position Specific Scoring Matrix Identifier (PSSM-ID) for the USP domain is 23,812). Additionally, we observed 353 patterns (signatures) of amino acid residues (functional sites) associated with 247 amino acid position patterns. For example, an amino acid pattern "AVDVMGHGGSVA" had the highest occurrence in 54 and is associated with amino acid position patterns:

(1) 11_12_13_41_113_114_116_117_127_128_129_130 (7 sequences)
(2) 10_11_12_40_112_113_115_116_126_127_128_129 (44 sequences)
(3) 9_10_11_39_111_112_114_115_125_126_127_128 (3 sequences).

The amino acid position pattern of 9_10_11_39_111_112_114_115_125_126_127_128 was restricted to sequences from three *Bilophila* species with Locus Tags (HMPREF0178_03304, T370DRAFT_02139, and HMPREF0179_03080). Based on the ATP-binding motif of G2XG9XG(S/T), our algorithm (a calculated field in the visual analytics software) classified the 353 functional site amino acid patterns into two motif types: 236 (non-ATP-binding motif) and 117 (ATP-binding motif). We designed a visual representation to grouped the 651 protein sequences by amino acid sequence length, amino acid pattern and amino acid position pattern (Figure 4 shows a subset for three genomes: *D. desulfuricans* ND132, *D. halophilus* DSM 5663 and *D. gilichiniskyi* K3S). The design allowed us to identify proteins with identical amino acid sequence length and pattern of functional site (for example, the 146 aa and 297 aa sets encoded by the three genomes). We observed 13 types of functional site patterns from 10 of the 13 universal stress protein sequences predicted from *Desulfovibrio desulfuricans* ND132 (Figure 5). Protein sequences for DND132_1399, DND132_2319, and DND132_2657 have evidence for ATP-binding. The NCBI CD-Search did not report functional sites for DND132_1176 (IMG/M Gene ID 2503785994), DND132_1371 (IMG/M Gene ID 2503786190), and DND132_1376 (IMG/M Gene ID 2503786195).

Amino Acid Sequence Length (aa)	Amino Acid Pattern	Position Pattern	Gene ID	Locus Tag	Genome Name	ATP-Binding?
139	PVDVMGKGGSVT	9_10_11_38_107_108_110_111_121_122_123_124	2503787147	DND132_2319	Desulfovibrio desulfuricans ND132	■
140	PVDCMGKGGSVT	9_10_11_39_108_109_111_112_122_123_124_125	2574159592	BR24DRAFT_1171	Desulfovibrio halophilus DSM 5663	■
			2709101797	Ga0139011_0808	Desulfovibrio gilichinskyi K3S	■
146	AVDVMGHGGSVA	10_11_12_40_112_113_115_116_126_127_128_129	2503787489	DND132_2657	Desulfovibrio desulfuricans ND132	■
			2574159560	BR24DRAFT_1139	Desulfovibrio halophilus DSM 5663	■
			2709103576	Ga0139011_2587	Desulfovibrio gilichinskyi K3S	■
148	ATHVIGRQTA	8_9_10_38_116_117_119_120_132_133	2574161143	BR24DRAFT_2727	Desulfovibrio halophilus DSM 5663	●
152	PTDVMSRGGGVT	9_10_11_39_114_115_119_120_130_131_132_133	2574160033	BR24DRAFT_1613	Desulfovibrio halophilus DSM 5663	●
162	AVDIIGRGGSVS	9_10_11_40_129_130_132_133_143_144_145_146	2503786218	DND132_1399	Desulfovibrio desulfuricans ND132	■
163	AVDIVGRGGSVS	9_10_11_40_130_131_133_134_144_145_146_147	2574159000	BR24DRAFT_0578	Desulfovibrio halophilus DSM 5663	■
288	AVDILGHHGSVP	6_7_8_36_131_132_135_136_143_144_146_147	2503786189	DND132_1370	Desulfovibrio desulfuricans ND132	●
294	AIGALGKGNSVC	8_9_10_39_122_123_125_126_136_137_138_139	2503787237	DND132_2406	Desulfovibrio desulfuricans ND132	●
295	ATDVMAYSGSTT	157_158_159_187_255_256_258_259_270_271_272_273	2503786307	DND132_1487	Desulfovibrio desulfuricans ND132	●
	AVTAMGHTDTTI	8_9_10_38_106_107_109_110_124_125_126_127	2503786307	DND132_1487	Desulfovibrio desulfuricans ND132	●
297	ATDAVAHSGSTM	164_165_166_194_259_260_262_263_274_275_276_277	2503786371	DND132_1547	Desulfovibrio desulfuricans ND132	●
	ATSVMGSSGSTL	8_9_10_38_114_115_117_118_130_131_132_133	2503786371	DND132_1547	Desulfovibrio desulfuricans ND132	●
	ATTVMGSGGSTL	8_9_10_38_114_115_118_119_130_131_132_133	2709102955	Ga0139011_1966	Desulfovibrio gilichinskyi K3S	●
	GTDAMAHTGSTV	164_165_166_194_259_260_262_263_274_275_276_277	2709102955	Ga0139011_1966	Desulfovibrio gilichinskyi K3S	●
	PTDVMAHGSTV	159_160_161_189_257_258_260_272_273_274_275	2574160690	BR24DRAFT_2272	Desulfovibrio halophilus DSM 5663	●
301	ATDVMAHSGSTV	158_159_160_188_244_245_247_248_259_260_261_262	2503786205	DND132_1386	Desulfovibrio desulfuricans ND132	●
	GVTVMGHEGSTL	8_9_10_38_105_106_108_114_124_125_126_127	2503786205	DND132_1386	Desulfovibrio desulfuricans ND132	●
310	CIGVVGSPRPLA	12_13_14_42_150_151_153_156_166_168_169_170	2503786952	DND132_2126	Desulfovibrio desulfuricans ND132	●
629	ALGVMGHGGETV	496_497_498_526_598_599_601_602_612_613_614_615	2503787549	DND132_2717	Desulfovibrio desulfuricans ND132	●
630	ALGVMGHGGETV	502_503_504_532_598_599_601_602_612_613_614_615	2709103738	Ga0139011_2749	Desulfovibrio gilichinskyi K3S	●

ATP-Binding?
● N
■ Y

Figure 4. The patterns of amino acid type and amino acid positions for selected *Desulfovibrionaceae* universal stress proteins. We annotated the proteins for the presence of ATP-binding motif (filled shape, square for presence of ATP-binding motif and circle for absence of ATP-binding motif). *Desulfovibrio desulfuricans* ND132 and *Desulfovibrio halophilus* DSM 5663 are mercury methylating. *Desulfovibrio gilichinskyi* K3S is included for comparison of the 629 aa universal stress protein from *Desulfovibrio desulfuricans* ND132.

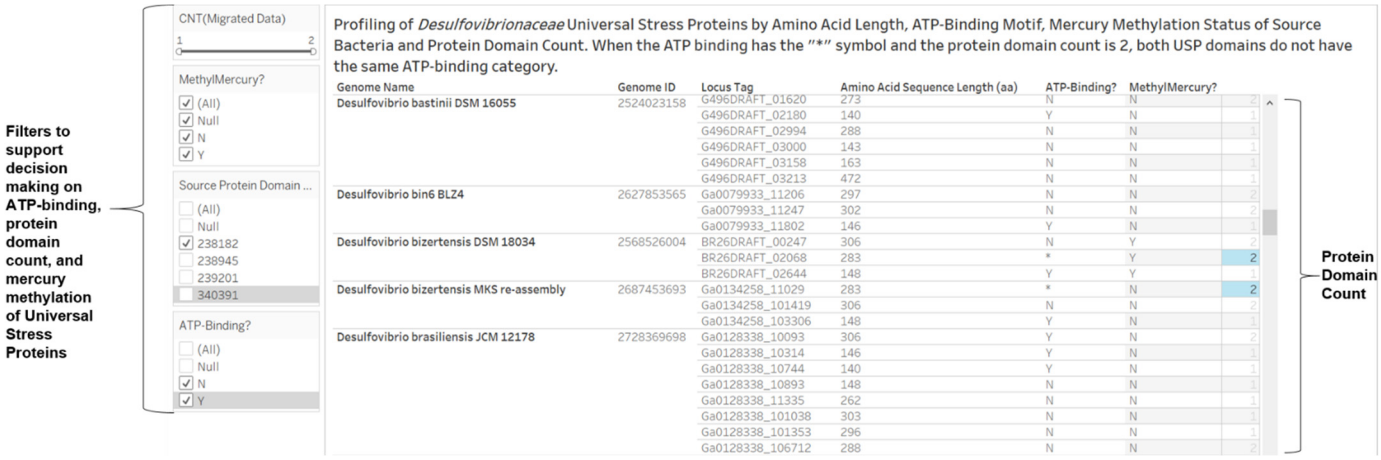


Figure 5. Profiling of *Desulfovibrionaceae* universal stress proteins by amino acid length, ATP-binding motif, mercury methylation status of source bacteria and protein domain count. When the ATP-binding has the "*" symbol and the protein domain count is 2, both USP domains do not have the same ATP-binding category (BR26DRAFT_02068 and Ga0134258, which are 283 aa USP of strains of *Desulfovibrio bizertensis*). Protein domain count of "1" indicates protein sequence has only one USP domain.

The DND132_2657 gene for a 146 aa protein was among the 54 *Desulfovibrionaceae* USP genes encoding the ATP-binding functional site pattern "AVDVMGHGGSVA". The protein domain arrangement of DND132_2717, a 629 aa protein sequence, comprised of a metal ion transporter domain and a USP domain with functional sites that do not conform with the ATP-binding motif. Comparison of amino acid sequence length and protein domain composition provided evidence that among the *Desulfovibrionaceae* genomes investigated the combination of metal ion transport domain and universal stress protein unique to *D. desulfuricans* ND132 and *D. gilichinskyi* K3S (previously named *Desulfovibrio algoritolerance* K3S). The IMG/M Gene ID, Locus Tag and amino acid sequence length for the equivalent gene of DND132_2717 in *Desulfovibrio gilichinskyi* K3S is 2709103738 and Ga0139011_2749 and 630 aa, respectively. Both protein sequences have an identical amino acid pattern of "ALGVMGHGGETV". Figure A2 in the Appendix A presents profiling of 92 *Desulfovibrionaceae* USPs by amino acid pattern, amino acid length, ATP-Binding motif, and mercury methylation status of source bacteria.

In Figure 5, the visual analytics design integrates the amino acid length, ATP-binding prediction of the USP domain, and the mercury methylation status of the source bacteria. The view among other functions allow for the categorization of a tandem-type USP by the types of ATP-binding prediction (Y = ATP-binding for both domains; N = Both domains are not ATP-binding; and * = One domain is ATP-binding and other does not bind ATP). The findings were confirmed with the NCBI Conserved Domains resource for three tandem-type USPs encoded in *Desulfovibrionaceae* genomes that encode the hgcA and hgcB proteins (Figure 6).

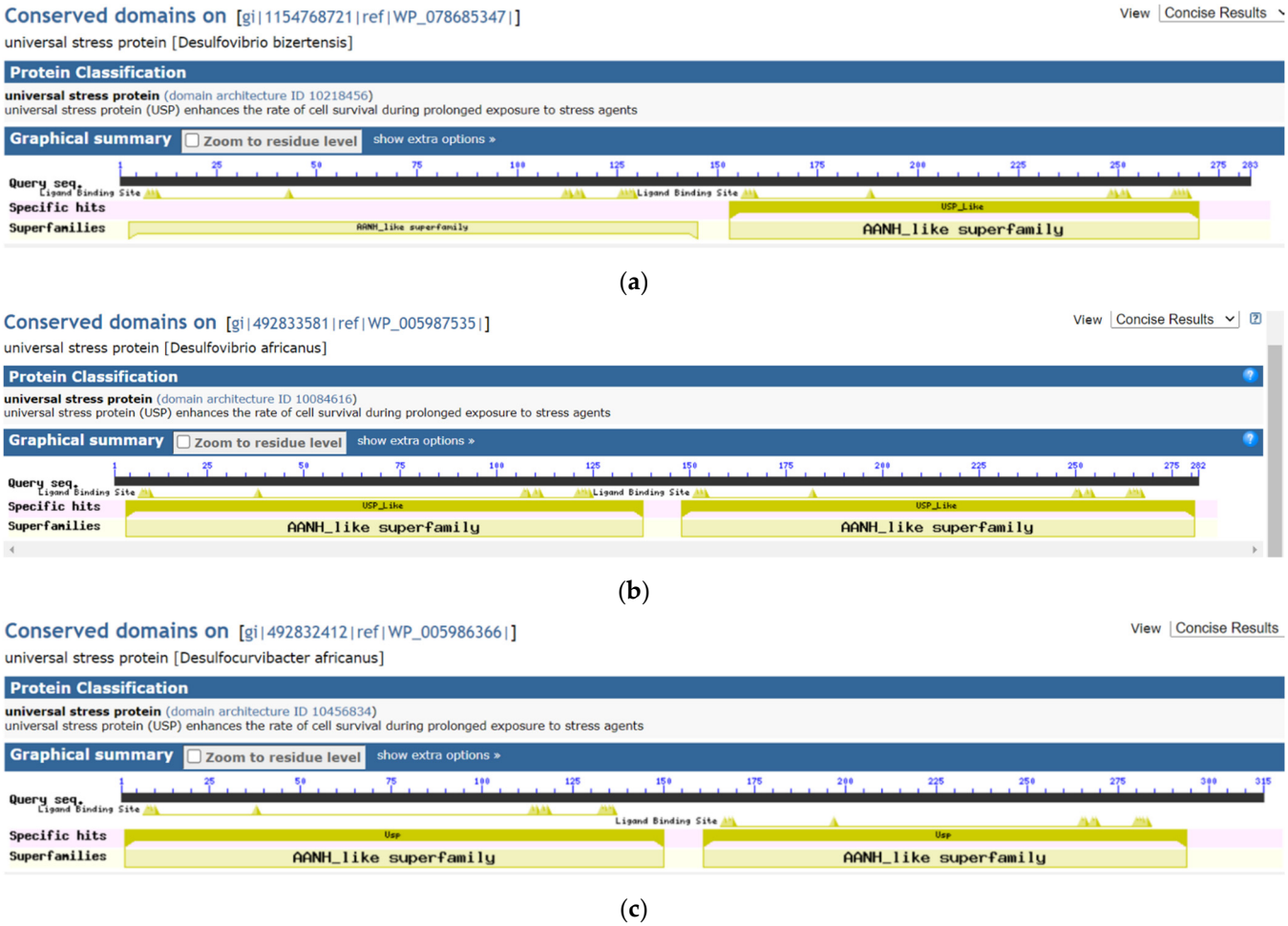


Figure 6. Protein domain compositions of three tandem-type universal stress proteins from mercury-methylating *Desulfovibrionaceae* bacteria. (**a**) In WP_0005987535 of *Desulfovibrio bizertensis* DSM 18034, the USP domains do not have the same ATP-binding category. (**b**) In WP_005986366 of *Desulfocurvicter africanus* PCS, the USP domains are both ATP-binding. (**c**) In WP_005986366 of *Desulfocurvicter africanus* PCS, the two USP domains are not ATP-binding. Interactive version of each protein domain composition is available at the National Center for Biotechnology Information (NCBI) Conserved Domains resource by searching for the protein sequence identifier.

4. Discussion

We have conducted bioinformatics investigations on the universal stress proteins encoded in *Desulfovibrionaceae* genomes including genomes of strains that methylate mercury. Prior to our study, the characterization of *Desulfovibrionaceae* USPs was limited to genome-wide transcriptome or proteome analyses [17,32,40]. Our report provides findings from protein sequence analytics to guide further research on the molecular functions, biological processes and cellular components associated with *Desulfovibrionaceae* universal stress proteins (USPs). In our prior publications, we have applied bioinformatics and developed visual analytics resources to understand the universal stress proteins of taxonomic groups namely viridiplantae, *Bacillus*, *Schistosoma*, *Alcanivorax*, *Brucella* and *Lactobacillus* [35,41–46]. In this report, we have made noteworthy findings on a collection of 719 *Desulfovibrionaceae* USPs regarding (1) protein domain arrangement; and (2) functional amino acid residues (Figure 1).

The observed counts of USP gene per genome among 93 *Desulfovibrionaceae* genomes ranged from 1 to 16 (Figure 2). The count of USP genes per genome could reflect the diverse phenotypic properties and habitats of the *Desulfovibrionaceae* members. The number of USP gene per genomes of *Escherichia coli* and *Mycobacterium tuberculosis* are six and ten, respectively [23,47]. The genomes of three *Halodesulfovibrio aestuarii* strains had 16 USP

genes, the highest observed among the 93 genomes investigated. The *Halodesulfovibrio* species tolerates up to 6% (*w*/*v*) sodium chloride (NaCl) with optimum growth at 1.5–3.5% (*w*/*v*) [48]. Future research could investigate the relationship between universal stress protein function and mercury methylation in the NaCl tolerant *Desulfovibrio halophilus* DSM 5663.

The genomes of three *Desulfovibrio desulfuricans desulfuricans* strains namely ATCC 27774, DSM 642 and DSM 7057 had only two USPs compared to 10 USP genes (retrieved from the IMG/M resource) for strain *D. desulfuricans* ND132. The finding of an excess number of USP genes further supports the reclassification of strain ND132. Recent phylogenetic analyses have clustered strain ND132 with validly published and reclassified members of *Pseudodesulfovibrio* genus including mercury-methylating *Pseudodesulfovibrio hydrargyri* BerOc1 [49,50]. A February 2021 publication formally described strain ND132 as *Pseudodesulfovibrio mercurii* ND132 [51]. We recommend comparative analysis of the universal stress proteins of *Pseudodesulfovibrio* strains to determine the effects of protein domain composition and genomic context of USP genes on stress response and methylmercury production.

Based on the ATP-binding motif of G2XG9XG(S/T), our algorithm (a calculated field in the visual analytics software, Tableau) categorized the 353 functional site amino acid patterns into two motif types 236 (non-ATP-binding motif) and 117 (ATP-binding motif) (Figure 5). For tandem-type USPs, we developed visual analytics views that provides three categories according to ATP-binding (Figure 6). Future research can investigate the biological significance of these categories. The *Desulfovibrio desulfuricans* ND132 protein sequences for DND132_1399, DND132_2319, and DND132_2657 have evidence for ATP-binding. Research investigations are required to understand the molecular function, biological processes and cellular components of the predicted ATP-binding USPs of strain ND132. The ATP-binding universal stress proteins are predicted to function in energy-dependent biological processes [52]. Examples of ATP (energy)-regulated processes are: (1) the regulation of entry into chronic persistent growth phase in *Mycobacterium tuberculosis* [28]; (2) the response to acid stress condition during the exponential growth phase in *Listeria innocua* [53]; (3) susceptibility of *Mycobacterium tuberculosis;* and (4) survival of *Mycobacterium smegmatis* in human monocyte cells [52]. The visual analytics resource accompanying this report provides a resource for interacting with the datasets on predicted ATP-binding status. Further, the genomic context or neighborhood of the USP genes can provide insights on the molecular function, biological processes and cellular components of the universal stress proteins of strain ND132.

Among the four universal stress protein sequences of strain ND132 that contain two protein domains (DND132_1487, DND132_1547, DND132_1386, and DND132_2717), only DND132_2717 (a 629 aa protein) has a metal ion transporter domain (pfam01566 or PF01566: natural resistance-associated macrophage protein (NRAMP) domain) (Figure 3). The transmembrane NRAMP family of transporters function as divalent metal ion transporters from bacteria to humans [54]. Thus, we recommend research to determine (1) if DND132_2717 transports inorganic divalent mercury ions (Hg^{2+}); (2) if DND132_2717 localizes to the membrane; and (3) if DND132_2717 function is regulated by the universal stress protein domain. The divalent metal cation transporter is listed among metal transporters impacted by the deletion of *hgcAB* genes of strain DND132 [55]. A yeast divalent cation transporter DMT1 of participates in the uptake of inorganic mercury [56]. Research publications on the uptake of inorganic mercury in mercury-methylating *Desulfovibrionaceae* species and related organisms could guide these future studies [57–60].

The results of bioinformatics investigations are influenced by several factors including the version of software and updates to datasets. The taxonomy of the *Desulfovibrionaceae* has recently been updated including reclassification and formal description of strain ND132 [51,61,62]. Our investigation has considered these limitations and have included information on when the investigations were conducted. We also use multiple approaches, databases and genomic data to achieve consensus results. We have provide results as part of visual analytics resources to support the formulation of new problems for investigations

beyond those reported here. The visual analytics resources can also serve as resources for educational interventions for learning biological data investigation [63]. We are also using the methods and findings to investigate denitrification potential of bacterial communities of Eastern Oyster (*Crassostrea virginica*) found in benthic environments [64].

5. Conclusions

We have determined protein domain composition and ATP-binding functional sites to categorize a collection of 719 genes predicted to encode the universal stress protein (USP) domains in 93 *Desulfovibrionaceae* genomes. The key findings are the categories of universal stress protein sequences according to (1) USP domain count; and (2) presence of ATP-binding motif (functional sites). We have identified 651 *Desulfovibrionaceae* universal stress protein sequences, of which 488 sequences had only one USP domain and 163 had two protein USP domains. The 488 single USP domain sequences were further categorized into 340 sequences with ATP-binding motif and 148 sequences without ATP-binding motif. The 163 double USP domain sequences were categorized into (1) both USP domains with ATP-binding motif (3 sequences); (2) both USP domains without ATP-binding motif (138 sequences); and (3) one USP domain with ATP-binding motif (21 sequences). We developed visual analytics resources to facilitate the investigation of these categories of datasets in the presence or absence of the mercury-methylating gene pair (hgcAB). Future research could utilize these functional categories to investigate the participation of universal stress proteins in the bacterial cellular uptake of inorganic mercury and methylmercury production, especially in anaerobic aquatic environments.

Supplementary Materials: The online versions of the interactive analytics resources produced are available at https://public.tableau.com/app/profile/qeubic/viz/uspdesulfofamily/overview. Figure A1: Evidence that the process for analytics of functional sites of protein sequences can be applied beyond the universal stress proteins; Figure A2: Profiling of 92 *Desulfovibrionaceae* universal stress proteins by amino acid pattern (functional sites for ATP binding), protein domain count, amino acid length, ATP-binding motif, and mercury methylation status of source bacteria.

Author Contributions: Conceptualization, R.D.I., D.S.M., A.M.D., B.W.M., M.O.J. and S.S.S.; methodology, R.D.I., D.S.M., A.M.D., B.W.M., S.S.S.; software, R.D.I., D.S.M., A.M.D., G.S.J., A.D.W. and Y.A.H.; validation, R.D.I., D.S.M., A.M.D., G.S.J., A.D.W., Y.A.H., B.W.M. and S.S.S.; formal analysis, R.D.I., D.S.M., A.M.D., G.S.J., A.D.W., Y.A.H., B.W.M. and S.S.S.; investigation, R.D.I., D.S.M., A.M.D., G.S.J., A.D.W. and Y.A.H.; resources, R.D.I., B.W.M. and M.O.J.; data curation, R.D.I., D.S.M., A.M.D., G.S.J., A.D.W. and Y.A.H.; writing—original draft preparation, R.D.I., D.S.M., A.M.D., G.S.J., A.D.W., Y.A.H., B.W.M., M.O.J. and S.S.S.; writing—review and editing, R.D.I., D.S.M., A.M.D., B.W.M., M.O.J., S.S.S.; visualization, R.D.I., D.S.M. and A.M.D.; supervision, R.D.I. and B.W.M.; project administration, R.D.I., B.W.M.; funding acquisition, R.D.I., B.W.M., M.O.J. and S.S.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the U.S. Department of Energy Minority Serving Institution Partnership Program (MSIPP) managed by the Savannah River National Laboratory under SRNS contract DE-AC09-08SR22470; Department of Education Title III Program (P031B170091); National Science Foundation (CSE-1829717, BIO-1901377 and EHR-2029363). The Article Processing Charge (APC) was funded by Bethune-Cookman University.

Data Availability Statement: The Perl code, input sequences, and output datasets used in this report for the analytics of the conserved protein domains are available on the GitHub software development platform at https://github.com/qeubic/protein_features (accessed on 20 August 2021).

Acknowledgments: Bethune-Cookman University for administrative support of the project.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A

The computer programming language code for preparing the output from NCBI Conserved Domain search can be applied to any collection of fasta-formatted protein sequences or list of NCBI protein identifiers. In this report, we applied the predicted protein sequences from the genome sequence of *Desulfovibrio desulfuricans* ND132. The input sequences are from the Integrated Microbial Genomes/Microbiomes (IMG/M) resource. The Perl code, input sequences, and output datasets used in this report for the analytics of the conserved protein domains are available on the GitHub software development platform at https://github.com/qeubic/protein_features (accessed on 20 August 2021).

A visual analytics resource for protein sequence analytics is available at https://public.tableau.com/app/profile/qeubic/viz/uspdesulfofamily/figureA1 (accessed on 20 August 2021). The designs are for interacting with the data on the protein families including protein domain composition and functional sites. Figure A1 shows that the protein sequence analytics procedure can be applied to other protein groups (e.g., nitrogen metabolism protein groups that have "nitr" in the gene/protein name).

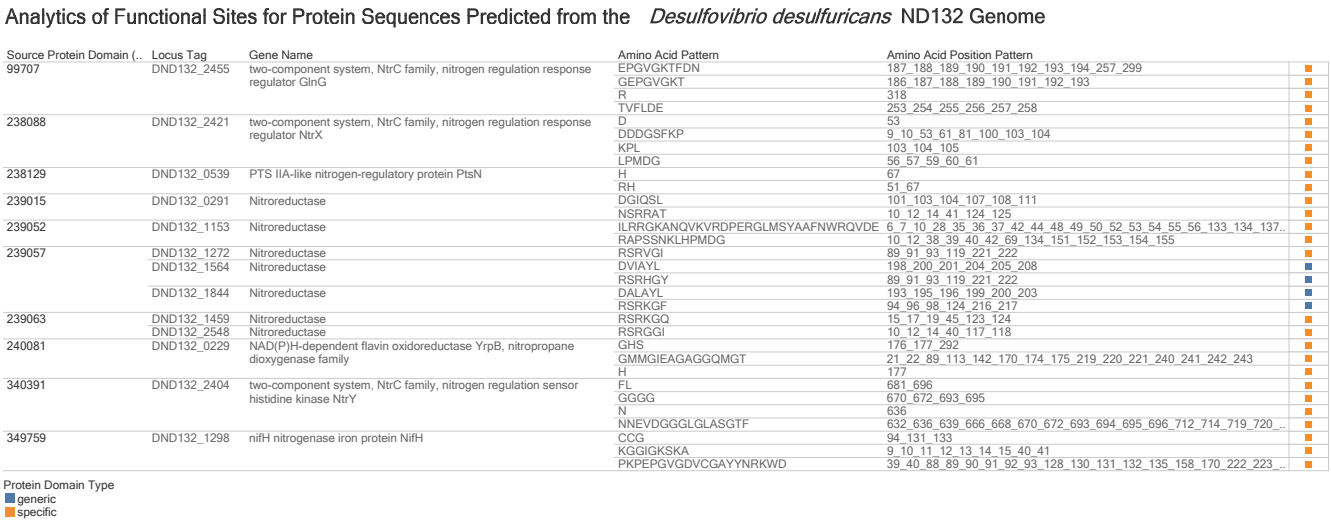


Figure A1. Evidence that the process for analytics of functional sites of protein sequences can be applied beyond the universal stress proteins. The visual representation integrates the protein domain identifier (source protein domain (PSSM-ID)), locus tag, gene name, amino acid pattern and amino acid position pattern. The protein sequences for *Desulfovibrio desulfuricans* ND132 were obtained from the Integrated Microbial Genomes and Microbiomes (IMG/M) system. The protein sequences with annotation for nitrogen metabolism are relevant to our research on the denitrification by microbial communities in oysters [64].

The integration of disparate features of the universal stress proteins from 50 *Desulfovibrionaceae* genomes can facilitate comparative analysis and planning of future research. Therefore, we have constructed a visualization that integrates the amino acid pattern, amino acid length, ATP-binding motif, and mercury methylation status of source bacteria (Figure A2. The mercury methylation status was obtained from the data page of the Biogeochemical Transformations at Critical Interfaces project of the Oak Ridge National Laboratory's Mercury Science Focus Area [12,20]. A total of 92 *Desulfovibrionaceae* USPs were profiled according to 12 amino acid patterns, 28 amino acid lengths, two ATP-binding motifs, and mercury methylation status of the source bacteria.

The 92 USPs included 60 USPs with ATP-binding motifs, 27 USPs from 12 genomes of mercury methylators and 15 single-domain USPs that are ATP-binding and from the genomes of mercury methylators. The genomes encoding mercury methylation are *Desulfohalovibrio reitneri* L21-Syr-AB, *Desulfovibrio africanus* DSM 2603, *Desulfovibrio africanus* PCS, *Desulfovibrio africanus* Walvis Bay, *Desulfovibrio alkalitolerans* DSM 16529, *Desulfovibrio desulfuricans* ND132, *Desulfovibrio halophilus* DSM 5663, *Desulfovibrio inopinatus* DSM 10711,

Desulfovibrio longus DSM 6739, *Desulfovibrio oxyclinae* DSM 11498, *Desulfovibrio* sp. X2, and *Pseudodesulfovibrio aespoeensis* Aspo-2, DSM 10631 [18].

Profiling of *Desulfovibrionaceae* Universal Stress Proteins by Amino Acid Pattern, Amino Acid Length, ATP-Binding Motif, and Mercury Methylation Status of Source Bacteria

Amino Acid Pattern	USP Domain Count	Amino Acid Sequence Length (aa)	Genome ID	Genome Name	Gene ID	Locus Tag	ATP-Binding? Y	ATP-Binding? N
ALGVMGHGGETV	1	629	2503754015	Desulfovibrio desulfuricans ND132	2503787549	DND132_2717		□
		630	2708742538	Desulfovibrio gilichinskyi K3S	2709103738	Ga0139011_2749		○
ATDAVAHSGSTM	2	297	2503754015	Desulfovibrio desulfuricans ND132	2503786371	DND132_1547		□
ATDVMAHSGSTV	2	296	2518645586	Halodesulfovibrio aestuarii aestuarii ATCC 29578	2518923941	F461DRAFT_00002		○
			2585428154	Desulfovibrio desulfuricans aestuarii DSM 17919	2588094275	EK04DRAFT_00002		○
		300	649633037	Pseudodesulfovibrio aespoeensis Aspo-2, DSM 10631	649851577	Daes_0785		□
		301	2503754015	Desulfovibrio desulfuricans ND132	2503786205	DND132_1386		□
			2523533628	Desulfovibrio zosterae DSM 11974	2524002313	H589DRAFT_1748		○
			2599185335	Desulfovibrio ferrireducens DSM 16995	2600104667	Ga0056056_2093		○
			2651869878	Pseudodesulfovibrio indicus J2	2653206634	AWY79_07325		○
			2687453294	Pseudodesulfovibrio indicus J2	2688215740	Ga0133372_111431		○
			2784746795	Pseudodesulfovibrio indicus DSM 101483	2785470980	Ga0244724_101331		○
		302	2829737207	Desulfobaculum xiamenense DSM 24233	2829738152	Ga0373198_946		○
		307	648276636	Desulfovibrio fructosovorans JJ	648709038	DesfrDRAFT_1886		○
			2561511137	Desulfovibrio alcoholivorans DSM 5433	2562217101	Q368DRAFT_03791		○
		309	2524614631	Desulfovibrio inopinatus DSM 10711	2525215207	G451DRAFT_04512		□
		314	2568526008	Desulfovibrio gracilis DSM 16080	2568551905	BR08DRAFT_00479		○
		315	2523533522	Desulfovibrio longus DSM 6739	2523632238	G452DRAFT_1371		□
		317	2823931327	Desulfovibrio ferrophilus IS5	2823933873	Ga0374997_2547		○
		322	2574179706	Desulfovibrio vietnamensis DSM 10520	2574204304	BP82DRAFT_02389		○
ATDVMAYSGSTT	2	295	2503754015	Desulfovibrio desulfuricans ND132	2503786307	DND132_1487		□
		296	2651869878	Pseudodesulfovibrio indicus J2	2653206495	AWY79_06630		○
			2687453294	Pseudodesulfovibrio indicus J2	2688215878	Ga0133372_111569		○
			2784746795	Pseudodesulfovibrio indicus DSM 101483	2785470843	Ga0244724_101194		○
ATSVMGSSGSTL	2	297	2503754015	Desulfovibrio desulfuricans ND132	2503786371	DND132_1547		□
			2651869878	Pseudodesulfovibrio indicus J2	2653206428	AWY79_06295		○
			2687453294	Pseudodesulfovibrio indicus J2	2688215945	Ga0133372_111636		○
			2784746795	Pseudodesulfovibrio indicus DSM 101483	2785470780	Ga0244724_101131		○
AVDIIGRGGSVS	1	162	2503754015	Desulfovibrio desulfuricans ND132	2503786218	DND132_1399	□	
AVDILGHHGSVP	1	288	2503754015	Desulfovibrio desulfuricans ND132	2503786189	DND132_1370		□
AVDVMGHGGSVA	1	146	644736353	Desulfovibrio salexigens DSM 2638	644838317	Desal_0180	○	
			649633037	Pseudodesulfovibrio aespoeensis Aspo-2, DSM 10631	649853836	Daes_3019	□	
			2503754015	Desulfovibrio desulfuricans ND132	2503787489	DND132_2657	□	
			2515154157	Desulfovibrio oxyclinae DSM 11498	2515861573	B149DRAFT_01255	□	
			2523533539	Desulfovibrio hydrothermalis AM13, DSM 14728	2523709406	H588DRAFT_01000	○	
			2523533628	Desulfovibrio zosterae DSM 11974	2524001971	H589DRAFT_1406	○	
			2524023158	Desulfovibrio bastinii DSM 16055	2524275037	G496DRAFT_00821	○	
			2540341170	Pseudodesulfovibrio piezophilus C1TLV30	2540825530	BN4_11796	○	
			2541046937	Desulfocurvus vexinensis DSM 17965	2541190038	G495DRAFT_02150	○	
			2565956520	Desulfovibrio frigidus DSM 17176	2566032372	BR06DRAFT_01277	○	
			2568526008	Desulfovibrio gracilis DSM 16080	2568551853	BR08DRAFT_00427	○	
			2571042916	Desulfovibrio halophilus DSM 5663	2574159560	BR24DRAFT_1139	□	
			2599185335	Desulfovibrio ferrireducens DSM 16995	2600102932	Ga0056056_0356	○	
			2627853564	LKpool_bin5_Desulfovibrio	2628079096	Ga0079932_10553	○	
			2627853565	Desulfovibrio bin6 BLZ4	2628082753	Ga0079933_11802	○	
			2648501899	Desulfovibrio hydrothermalis AM13, DSM 14728	2651740980	Ga0045936_122639	○	
			2651869878	Pseudodesulfovibrio indicus J2	2653206023	AWY79_04270	○	
			2687453294	Pseudodesulfovibrio indicus J2	2688216361	Ga0133372_112053	○	
			2708742538	Desulfovibrio gilichinskyi K3S	2709103576	Ga0139011_2587	○	
			2775506867	Pseudodesulfovibrio profundus 500-1	2776113880	Ga0259004_121682	○	
			2784746795	Pseudodesulfovibrio indicus DSM 101483	2785472404	Ga0244724_104221	○	
			2811995097	Desulfovibrio sp. SP109	2813129853	Ga0266297_101766	○	
			2823931327	Desulfovibrio ferrophilus IS5	2823934391	Ga0374997_3065	○	
		147	647000236	Desulfovibrio carbinoliphilus oakridgensis FW1012B	647338606	DFW101DRAFT_1316	○	
			648276636	Desulfovibrio fructosovorans JJ	648707840	DesfrDRAFT_0699	○	
			2508501038	Desulfovibrio sp. U5L	2508666996	DesU5LDRAFT_0819	○	
			2540341244	Desulfovibrio cf. magneticus IFRC170	2541021702	K366DRAFT_1497	○	
			2561511137	Desulfovibrio alcoholivorans DSM 5433	2562213519	Q368DRAFT_00206	○	
			2574180452	Desulfovibrio idahonensis DSM 15450	2576855954	BR09DRAFT_00821	○	
			2622736540	Desulfovibrio mexicanus DSM 13116	2623321993	Ga0070557_0782	○	
			2636415935	Desulfovibrio sp. TomC	2638627572	Ga0078453_10979	○	
			2788500052	Desulfovibrionaceae bacterium CG1_02_65_16	2789373593	Ga0302021_11553	○	
		148	2576861818	Desulfovibrio sp. DMSS-1	2579733615	H034DRAFT_4069	○	
			2791354756	Desulfovibrio sp. An276	2791375294	Ga0303299_10156	○	
		149	648276636	Desulfovibrio fructosovorans JJ	648708860	DesfrDRAFT_1708	○	
			2513237322	Bilophila sp. 4_1_30	2514481677	HMPREF0178_03304	○	
			2561511128	Bilophila wadsworthia ATCC 49260	2562181888	T370DRAFT_02139	○	
			2561511137	Desulfovibrio alcoholivorans DSM 5433	2562215893	Q368DRAFT_02582	○	
			2562617176	Bilophila wadsworthia 3_1_6	2563290109	HMPREF0179_03080	○	
			2724679721	Desulfovibrio sp. Bin 1	2727894400	Ga0182992_13124	○	
					2727895257	Ga0182992_14618	○	
		150	647000236	Desulfovibrio carbinoliphilus oakridgensis FW1012B	647338537	DFW101DRAFT_1249	○	
			2503754016	Desulfovibrio africanus Walvis Bay	2503789945	DesafDRAFT_1965	□	
			2508501038	Desulfovibrio sp. U5L	2508669883	DesU5LDRAFT_3707	○	
			2519899530	Desulfovibrio africanus PCS	2520046563	PCS_02374	□	
			2523533522	Desulfovibrio longus DSM 6739	2523631860	G452DRAFT_0993	□	
			2524614631	Desulfovibrio inopinatus DSM 10711	2525211881	G451DRAFT_01182	□	
			2526164711	Desulfovibrio africanus DSM 2603	2527068051	H585DRAFT_02055	□	
		151	2524023062	Desulfovibrio aminophilus DSM 12254	2524114690	H587DRAFT_02833	○	
			2571042912	Desulfovibrio alkalitolerans DSM 16529	2574146249	BS30DRAFT_1110	□	
			2597489948	Desulfovibrio sp. X2	2598072065	Ga0038523_00011	□	
			2619619068	Desulfovibrio alkalitolerans DSM 16529	2620572104	Ga0032409_12650	□	
		152	2571042914	Desulfohalovibrio reitneri L21-Syr-AB	2574154204	N911DRAFT_1582	□	
		153	2523231030	Desulfovibrio putealis DSM 16056	2523323852	G453DRAFT_00597	○	
AVTAMGHTDTTI	2	295	2503754015	Desulfovibrio desulfuricans ND132	2503786307	DND132_1487		□
CIGVVGSPRPLA	1	310	2503754015	Desulfovibrio desulfuricans ND132	2503786952	DND132_2126		□
GVTVMGHEGSTL	2	301	2503754015	Desulfovibrio desulfuricans ND132	2503786205	DND132_1386		□
PVDVMGKGGSVT	1	139	2503754015	Desulfovibrio desulfuricans ND132	2503787147	DND132_2319	□	
			2651869878	Pseudodesulfovibrio indicus J2	2653205644	AWY79_02375	○	
			2687453294	Pseudodesulfovibrio indicus J2	2688216746	Ga0133372_112438	○	
			2784746795	Pseudodesulfovibrio indicus DSM 101483	2785473413	Ga0244724_11073	○	
		140	2568526008	Desulfovibrio gracilis DSM 16080	2568553986	BR08DRAFT_02567	○	

MethylMercury?
○ N
□ Y

Figure A2. Profiling of 92 *Desulfovibrionaceae* universal stress proteins by amino acid pattern (functional sites for ATP binding), protein domain count, amino acid length, ATP-binding motif, and mercury methylation status of source bacteria. The interactive version of the visual is available at https://public.tableau.com/app/profile/qeubic/viz/uspdesulfofamily/figureB1 (accessed on 20 August 2021).

References

1. Bjørklund, G.; Tinkov, A.A.; Dadar, M.; Rahman, M.M.; Chirumbolo, S.; Skalny, A.V.; Skalnaya, M.G.; Haley, B.E.; Ajsuvakova, O.P.; Aaseth, J. Insights into the potential role of mercury in Alzheimer's disease. *J. Mol. Neurosci.* **2019**, *67*, 511–533. [CrossRef]
2. Driscoll, C.T.; Mason, R.P.; Chan, H.M.; Jacob, D.J.; Pirrone, N. Mercury as a global pollutant: Sources, pathways, and effects. *Environ. Sci. Technol.* **2013**, *47*, 4967–4983. [CrossRef]
3. Higueras, P.; Oyarzun, R.; Kotnik, J.; Esbrí, J.M.; Martínez-Coronado, A.; Horvat, M.; López-Berdonces, M.A.; Llanos, W.; Vaselli, O.; Nisi, B.; et al. A compilation of field surveys on gaseous elemental mercury (gem) from contrasting environmental settings in europe, south america, south africa, and china: Separating fads from facts. *Environ. Geochem. Health* **2014**, *36*, 713–734. [CrossRef] [PubMed]
4. Davis, M.A.; Gilbert-Diamond, D.; Karagas, M.R.; Li, Z.; Moore, J.H.; Williams, S.M.; Frost, H.R. A dietary-wide association study (DWAS) of environmental metal exposure in US children and adults. *PLoS ONE* **2014**, *9*, e104768. [CrossRef]
5. Vahabzadeh, M.; Balali-Mood, M. Occupational metallic mercury poisoning in gilders. *Int. J. Occup. Environ. Med.* **2016**, *7*, 116. [CrossRef] [PubMed]
6. Sakamoto, M.; Nakamura, M.; Murata, K. Mercury as a global pollutant and mercury exposure assessment and health effects. *Nihon Eiseigaku Zasshi* **2018**, *73*, 258–264. [CrossRef] [PubMed]
7. Stone, J.J.; McCutcheon, C.M.; Stetler, L.D.; Chipps, S.R. Interrelationships between fish tissue mercury concentrations and water quality for South Dakota natural lakes and impoundments. *Water Air Soil Pollut.* **2011**, *222*, 337–349. [CrossRef]
8. Betemariam, H.H.; McCutcheon, C.M.; Davis, A.D.; Stetler, L.D.; DeSutter, T.M.; Penn, M.R.; Stone, J.J. Geochemical behavior and watershed influences associated with sediment-bound mercury for South Dakota lakes and impoundments. *Water Air Soil Pollut.* **2013**, *224*, 1–14. [CrossRef]
9. Brent, R.N.; Kain, D.G. Development of an empirical nonlinear model for mercury bioaccumulation in the South and South Fork Shenandoah Rivers of Virginia. *Arch. Environ. Contam. Toxicol.* **2011**, *61*, 614–623. [CrossRef]
10. Brooks, S.C.; Southworth, G.R. History of mercury use and environmental contamination at the Oak Ridge Y-12 Plant. *Environ. Pollut.* **2011**, *159*, 219–228. [CrossRef] [PubMed]
11. Lin, T.Y.; Kampalath, R.A.; Lin, C.-C.; Zhang, M.; Chavarria, K.; Lacson, J.; Jay, J.A. Investigation of mercury methylation pathways in biofilm versus planktonic cultures of *Desulfovibrio desulfuricans*. *Environ. Sci. Technol.* **2013**, *47*, 5695–5702. [CrossRef]
12. Parks, J.M.; Johs, A.; Podar, M.; Bridou, R.; Hurt, R.A.; Smith, S.D.; Tomanicek, S.J.; Qian, Y.; Brown, S.D.; Brandt, C.C. The genetic basis for bacterial mercury methylation. *Science* **2013**, *339*, 1332–1335. [CrossRef] [PubMed]
13. Voordouw, G. The genus *Desulfovibrio*: The centennial. *Appl. Environ. Microbiol.* **1995**, *61*, 2813. [CrossRef]
14. Galushko, A.; Kuever, J. Desulfovibrionaceae. In *Bergey's Manual of Systematics of Archaea and Bacteria*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2015; pp. 1–13. [CrossRef]
15. Heidelberg, J.F.; Seshadri, R.; Haveman, S.A.; Hemme, C.L.; Paulsen, I.T.; Kolonay, J.F.; Eisen, J.A.; Ward, N.; Methe, B.; Brinkac, L.M. The genome sequence of the anaerobic, sulfate-reducing bacterium *Desulfovibrio vulgaris* Hildenborough. *Nat. Biotechnol.* **2004**, *22*, 554–559. [CrossRef]
16. Postgate, J.R.; Kent, H.M.; Robson, R.L.; Chesshyre, J.A. The genomes of *Desulfovibrio gigas* and *D. vulgaris*. *Microbiology* **1984**, *130*, 1597–1601. [CrossRef] [PubMed]
17. Chhabra, S.; He, Q.; Huang, K.; Gaucher, S.; Alm, E.; He, Z.; Hadi, M.; Hazen, T.; Wall, J.; Zhou, J. Global analysis of heat shock response in *Desulfovibrio vulgaris* Hildenborough. *J. Bacteriol.* **2006**, *188*, 1817–1828. [CrossRef]
18. Gilmour, C.C.; Elias, D.A.; Kucken, A.M.; Brown, S.D.; Palumbo, A.V.; Schadt, C.W.; Wall, J.D. Sulfate-reducing bacterium *Desulfovibrio desulfuricans* ND132 as a model for understanding bacterial mercury methylation. *Appl. Environ. Microbiol.* **2011**, *77*, 3938–3951. [CrossRef]
19. Podar, M.; Gilmour, C.C.; Brandt, C.C.; Soren, A.; Brown, S.D.; Crable, B.R.; Palumbo, A.V.; Somenahally, A.C.; Elias, D.A. Global prevalence and distribution of genes and microorganisms involved in mercury methylation. *Sci. Adv.* **2015**, *1*, e1500675. [CrossRef] [PubMed]
20. ORNL. Table of Predicted Methylators (04/25/16). Available online: https://www.esd.ornl.gov/programs/rsfa/data.shtml (accessed on 14 June 2021).
21. Samanta, S.; Biswas, P.; Banerjee, A.; Bose, A.; Siddiqui, N.; Nambi, S.; Saini, D.K.; Visweswariah, S.S. A universal stress protein in *Mycobacterium smegmatis* sequesters the cAMP-regulated lysine acyltransferase and is essential for biofilm formation. *J. Biol. Chem.* **2020**, *295*, 1500–1516. [CrossRef] [PubMed]
22. Chen, W.; Honma, K.; Sharma, A.; Kuramitsu, H.K. A universal stress protein of *Porphyromonas gingivalis* is involved in stress responses and biofilm formation. *FEMS Microbiol. Lett.* **2006**, *264*, 15–21. [CrossRef] [PubMed]
23. Nachin, L.; Nannmark, U.; Nyström, T. Differential roles of the universal stress proteins of *Escherichia coli* in oxidative stress resistance, adhesion, and motility. *J. Bacteriol.* **2005**, *187*, 6265–6272. [CrossRef]
24. Vollmer, A.C.; Bark, S.J. Twenty-five years of investigating the universal stress protein: Function, structure, and applications. *Adv. Appl. Microbiol.* **2018**, *102*, 1–36.
25. Kvint, K.; Nachin, L.; Diez, A.; Nyström, T. The bacterial universal stress protein: Function and regulation. *Curr. Opin. Microbiol.* **2003**, *6*, 140–145. [CrossRef]
26. Chauhan, J.S.; Mishra, N.K.; Raghava, G.P. Identification of ATP binding residues of a protein from its primary sequence. *BMC Bioinform.* **2009**, *10*, 434. [CrossRef] [PubMed]

27. Tkaczuk, K.L.; Shumilin, I.A.; Chruszcz, M.; Evdokimova, E.; Savchenko, A.; Minor, W. Structural and functional insight into the universal stress protein family. *Evol. Appl.* **2013**, *6*, 434–449. [CrossRef] [PubMed]
28. Drumm, J.E.; Mi, K.; Bilder, P.; Sun, M.; Lim, J.; Bielefeldt-Ohmann, H.; Basaraba, R.; So, M.; Zhu, G.; Tufariello, J.A.M.; et al. *Mycobacterium tuberculosis* universal stress protein Rv2623 regulates bacillary growth by ATP-Binding: Requirement for establishing chronic persistent infection. *PLoS Pathog.* **2009**, *5*, e1000460. [CrossRef]
29. Khelaifia, S.; Fardeau, M.-L.; Pradel, N.; Aussignargues, C.; Garel, M.; Tamburini, C.; Cayol, J.-L.; Gaudron, S.; Gaill, F.; Ollivier, B. *Desulfovibrio piezophilus* sp. nov., a piezophilic, sulfate-reducing bacterium isolated from wood falls in the Mediterranean Sea. *Int. J. Syst. Evol. Microbiol.* **2011**, *61*, 2706–2711. [CrossRef]
30. Silva, G.; Rodrigues-Pousada, C. A 6940 bp DNA fragment from *Desulfovibrio gigas* contains genes coding for lipoproteins, universal stress response and transcriptional regulator protein homologues. *DNA Seq.* **2001**, *12*, 229–238. [CrossRef]
31. Williamson, A.J.; Carlson, H.K.; Kuehl, J.V.; Huang, L.L.; Iavarone, A.T.; Deutschbauer, A.; Coates, J.D. Dissimilatory sulfate reduction under high pressure by *Desulfovibrio alaskensis* G20. *Front. Microbiol.* **2018**, *9*, 1465. [CrossRef]
32. Zhang, W.; Gritsenko, M.A.; Moore, R.J.; Culley, D.E.; Nie, L.; Petritis, K.; Strittmatter, E.F.; Camp, D.G.; Smith, R.D.; Brockman, F.J. A proteomic view of *Desulfovibrio vulgaris* metabolism as determined by liquid chromatography coupled with tandem mass spectrometry. *Proteomics* **2006**, *6*, 4286–4299. [CrossRef] [PubMed]
33. Chen, I.M.A.; Markowitz, V.M.; Chu, K.; Palaniappan, K.; Szeto, E.; Pillay, M.; Ratner, A.; Huang, J.; Andersen, E.; Huntemann, M.; et al. IMG/M: Integrated genome and metagenome comparative data analysis system. *Nucleic Acids Res.* **2017**, *45*. [CrossRef]
34. Marchler-Bauer, A.; Bryant, S.H. CD-Search: Protein domain annotations on the fly. *Nucleic Acids Res.* **2004**, *32*, 327–331. [CrossRef] [PubMed]
35. Isokpehi, R.D.; Simmons, S.S.; Johnson, M.O.; Payton, M. Genomic evidence for bacterial determinants influencing obesity development. *Int. J. Environ. Res. Public Health* **2017**, *14*, 345. [CrossRef]
36. Sacha, D.; Stoffel, A.; Stoffel, F.; Kwon, B.C.; Ellis, G.; Keim, D.A. Knowledge generation model for visual analytics. *IEEE Trans. Vis. Comput. Graph.* **2014**, *20*, 1604–1613. [CrossRef]
37. Sedig, K.; Parsons, P. Interaction design for complex cognitive activities with visual representations: A pattern-based approach. *AIS Trans. Hum. Comput. Interact.* **2013**, *5*, 84–133. [CrossRef]
38. Zarembinski, T.I.; Hung, L.-W.; Mueller-Dieckmann, H.-J.; Kim, K.-K.; Yokota, H.; Kim, R.; Kim, S.-H. Structure-based assignment of the biochemical function of a hypothetical protein: A test case of structural genomics. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 15189–15193. [CrossRef] [PubMed]
39. LeDuc, R.D.; Doak, T.; Wu, L.-S.; Blood, P.D.; Ganote, C.L.; Vaughn, M. National Center for Genome Analysis support leverages XSEDE to support life science research. In Proceedings of the Extreme Science and Engineering Discovery Environment, San Diego, CA, USA, 22–25 July 2013.
40. Clark, M.E.; He, Z.; Redding, A.M.; Joachimiak, M.P.; Keasling, J.D.; Zhou, J.Z.; Arkin, A.P.; Mukhopadhyay, A.; Fields, M.W. Transcriptomic and proteomic analyses of *Desulfovibrio vulgaris* biofilms: Carbon and energy flow contribute to the distinct biofilm growth state. *BMC Genom.* **2012**, *13*, 138. [CrossRef]
41. Isokpehi, R.D.; Mahmud, O.; Mbah, A.N.; Simmons, S.S.; Avelar, L.; Rajnarayanan, R.V.; Udensi, U.K.; Ayensu, W.K.; Cohly, H.H.; Brown, S.D. Developmental regulation of genes encoding universal stress proteins in *Schistosoma mansoni*. *Gene Regul. Syst. Bio.* **2011**, *5*. [CrossRef] [PubMed]
42. Isokpehi, R.D.; Simmons, S.S.; Cohly, H.H.; Ekunwe, S.I.; Begonia, G.B.; Ayensu, W.K. Identification of drought-responsive universal stress proteins in viridiplantae. *Bioinform. Biol. Insights* **2011**, *5*. [CrossRef]
43. Isokpehi, R.D.; Wootson, K.M.; Smith-McInnis, D.R.; Simmons, S.S. Interactive analytics for complex cognitive activities on information from annotations of prokaryotic genomes. *J. Comput. Sci. Educ.* **2017**, *8*, 29–36. [CrossRef]
44. Kashim, Z.A. *Genomic Context Analytics of Genes for Universal Stress Proteins from Petroleum-Degrading Alcanivorax*; University of South Africa: Pretoria, South Africa, 2016.
45. Mbah, A.N.; Mahmud, O.; Awofolu, O.R.; Isokpehi, R.D. Inferences on the biochemical and environmental regulation of universal stress proteins from Schistosomiasis parasites. *Adv. Appl. Bioinform. Chem.* **2013**, *6*, 15–27. [CrossRef]
46. Williams, B.S.; Isokpehi, R.D.; Mbah, A.N.; Hollman, A.L.; Bernard, C.O.; Simmons, S.S.; Ayensu, W.K.; Garner, B.L. Functional annotation analytics of *Bacillus* genomes reveals stress responsive acetate utilization and sulfate uptake in the biotechnologically relevant *Bacillus megaterium*. *Bioinform. Biol. Insights* **2012**, *6*. [CrossRef] [PubMed]
47. Hingley-Wilson, S.; Lougheed, K.; Ferguson, K.; Leiva, S.; Williams, H. Individual *Mycobacterium tuberculosis* universal stress protein homologues are dispensable in vitro. *Tuberculosis* **2010**, *90*, 236–244. [CrossRef]
48. Shivani, Y.; Subhash, Y.; Sasikala, C.; Ramana, C.V. *Halodesulfovibrio spirochaetisodalis* gen. nov. sp. nov. and reclassification of four *Desulfovibrio* spp. *Int. J. Syst. Evol. Microbiol.* **2017**, *67*, 87–93. [CrossRef] [PubMed]
49. Ranchou-Peyruse, M.; Goñi-Urriza, M.; Guignard, M.; Goas, M.; Ranchou-Peyruse, A.; Guyoneaud, R. *Pseudodesulfovibrio hydrargyri* sp. nov., a mercury-methylating bacterium isolated from a brackish sediment. *Int. J. Syst. Evol. Microbiol.* **2018**, *68*, 1461–1466. [CrossRef] [PubMed]
50. Goñi-Urriza, M.; Klopp, C.; Ranchou-Peyruse, M.; Ranchou-Peyruse, A.; Monperrus, M.; Khalfaoui-Hassani, B.; Guyoneaud, R. Genome insights of mercury methylation among *Desulfovibrio* and *Pseudodesulfovibrio* strains. *Res. Microbiol.* **2020**, *171*, 3–12. [CrossRef]

51. Gilmour, C.C.; Soren, A.B.; Gionfriddo, C.M.; Podar, M.; Wall, J.D.; Brown, S.D.; Michener, J.K.; Urriza, M.S.G.; Elias, D.A. *Pseudodesulfovibrio mercurii* sp. nov., a mercury-methylating bacterium isolated from sediment. *Int. J. Syst. Evol. Microbiol.* **2021**, *71*, 004697.
52. Jia, Q.; Hu, X.; Shi, D.; Zhang, Y.; Sun, M.; Wang, J.; Mi, K.; Zhu, G. Universal stress protein Rv2624c alters abundance of arginine and enhances intracellular survival by ATP binding in mycobacteria. *Sci. Rep.* **2016**, *6*, 35462. [CrossRef]
53. Tremonte, P.; Succi, M.; Coppola, R.; Sorrentino, E.; Tipaldi, L.; Picariello, G.; Pannella, G.; Fraternali, F. Homology-based modeling of universal stress protein from Listeria innocua up-regulated under acid stress conditions. *Front. Microbiol.* **2016**, *7*, 1998. [CrossRef]
54. Cellier, M.F. Nramp: From sequence to structure and mechanism of divalent metal import. *Curr. Top. Membr.* **2012**, *69*, 249–293.
55. Qian, C.; Chen, H.; Johs, A.; Lu, X.; An, J.; Pierce, E.M.; Parks, J.M.; Elias, D.A.; Hettich, R.L.; Gu, B. Quantitative proteomic analysis of biological processes and responses of the bacterium *Desulfovibrio desulfuricans* ND132 upon deletion of its mercury methylation genes. *Proteomics* **2018**, *18*. [CrossRef]
56. Vázquez, M.; Vélez, D.; Devesa, V.; Puig, S. Participation of divalent cation transporter DMT1 in the uptake of inorganic mercury. *Toxicology* **2015**, *331*, 119–124. [CrossRef]
57. Isaure, M.-P.; Albertelli, M.; Kieffer, I.; Tucoulou, R.; Petrel, M.; Gontier, E.; Tessier, E.; Monperrus, M.; Goñi-Urriza, M. Relationship between Hg speciation and Hg methylation/demethylation processes in the sulfate-reducing bacterium *Pseudodesulfovibrio hydrargyri*: Evidences from HERFD-XANES and nano-XRF. *Front. Microbiol.* **2020**, *11*, 2506. [CrossRef]
58. Schaefer, J.K.; Szczuka, A.; Morel, F.o.M. Effect of divalent metals on Hg (II) uptake and methylation by bacteria. *Environ. Sci. Technol.* **2014**, *48*, 3007–3013. [CrossRef] [PubMed]
59. An, J.; Zhang, L.; Lu, X.; Pelletier, D.A.; Pierce, E.M.; Johs, A.; Parks, J.M.; Gu, B. Mercury uptake by *Desulfovibrio desulfuricans* ND132: Passive or active? *Environ. Sci. Technol.* **2019**, *53*, 6264–6272. [CrossRef] [PubMed]
60. Thomas, S.A.; Mishra, B.; Myneni, S.C. Cellular mercury coordination environment, and not cell surface ligands, influence bacterial methylmercury production. *Environ. Sci. Technol.* **2020**, *54*, 3960–3968. [CrossRef]
61. Waite, D.W.; Chuvochina, M.; Pelikan, C.; Parks, D.H.; Yilmaz, P.; Wagner, M.; Loy, A.; Naganuma, T.; Nakai, R.; Whitman, W.B. Proposal to reclassify the proteobacterial classes Deltaproteobacteria and Oligoflexia, and the phylum Thermodesulfobacteria into four phyla reflecting major functional capabilities. *Int. J. Syst. Evol. Microbiol.* **2020**, *70*, 5972–6016. [CrossRef] [PubMed]
62. Parte, A.C.; Carbasse, J.S.; Meier-Kolthoff, J.P.; Reimer, L.C.; Göker, M. List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *Int. J. Syst. Evol. Microbiol.* **2020**, *70*, 5607. [CrossRef] [PubMed]
63. Isokpehi, R.D.; Krejci, S.E.; Johnson, M.O.; Mapp, B.W. *RAPID: Educational Interventions for Undergraduate Students and Informal Learners for Robust Learning of COVID-19 Knowledge (Award #2029363)*; National Science Foundation, Bethune-Cookman University: Daytona Beach, FL, USA, 2020.
64. Isokpehi, R.D.; Kim, Y. *Excellence in Research: Microbiome of the Eastern Oyster (Crassostrea virginica) and Its Denitrification Potential in Benthic Systems (Award #1901377)*; National Science Foundation, Bethune-Cookman University: Daytona Beach, FL, USA, 2019.