




Article

Insights into Functions of Universal Stress Proteins Encoded by Genomes of Gastric Cancer Pathogen *Helicobacter pylori* and Related Bacteria

Raphael D. Isokpehi ^{1,*}, Shaneka S. Simmons ², Angela U. Makolo ³, Antoinisha L. Hollman ², Solayide A. Adesida ⁴, Olabisi O. Ojo ^{5,*} and Amos O. Abioye ⁶

¹ Transdisciplinary Data Scholars Development Program, Bethune-Cookman University, Daytona Beach, FL 32114, USA

² Division of Arts and Sciences, Jarvis Christian University, Hawkins, TX 75765, USA

³ University of Ibadan Bioinformatics Group, Department of Computer Science, University of Ibadan, Ibadan 200005, Oyo State, Nigeria

⁴ Department of Microbiology, Faculty of Science, University of Lagos, Akoka 101017, Lagos State, Nigeria

⁵ Department of Natural Sciences, Albany State University, Albany, GA 31721, USA

⁶ College of Pharmacy & Health Sciences, Belmont University, Nashville, TN 37212, USA; amos.abioye@belmont.edu

* Correspondence: raphael.isokpehi@outlook.com (R.D.I.); olabisi.ojo@asurams.edu (O.O.O.)

[†] Current address: Office of Data Science Strategy, National Institutes of Health, Bethesda, MD 20892, USA.

Abstract: The genes that encode the universal stress protein (USP) family domain (pfam00582) aid the survival of bacteria in specific host or habitat-induced stress conditions. Genome sequencing revealed that the genome of *Helicobacter pylori*, a gastric cancer pathogen, typically contains one USP gene, while related helicobacters have one or two distinct USP genes. However, insights into the functions of *Helicobacteraceae* (*Helicobacter* and *Wolinella*) USP genes are still limited to inferences from large-scale genome sequencing. Thus, we have combined bioinformatics and visual analytics approaches to conduct a more comprehensive data investigation of a set of 1045 universal stress protein sequences encoded in 1014 genomes including 785 *Helicobacter pylori* genomes. The study generated a representative set of 183 USP sequences consisting of 180 *Helicobacter* sequences, two *Wolinella succinogenes* sequences, and a sequence from a related campylobacteria. We used the amino acid residues and positions of the 12 possible functional sites in 1030 sequences to identify 25 functional sites patterns for guiding studies on functional interactions of *Helicobacteraceae* USPs with ATP and other molecules. Genomic context searches and analysis identified USP genes of gastric and enterohepatic helicobacters that are adjacent or in operons with genes for proteins responsive to DNA-damaging oxidative stress (ATP-dependent proteases: ClpS and ClpA); and DNA uptake proteins (natural competence for transformation proteins: ComB6, ComB7, ComB8, ComB9, ComB10, ComBE, and conjugative transfer signal peptidase TraF). Since transcriptomic evidence indicates that oxidative stress and the presence of virulence-associated genes regulate the transcription of *H. pylori* USP gene, we recommend further research on *Helicobacter* USP genes and their neighboring genes in oxidative stress response and virulence of helicobacters. To facilitate the reuse of data and research, we produced interactive analytics resources of a dataset composed of values for variables including phylogeography of *H. pylori* strains, protein sequence features, and gene neighborhood.

Keywords: ATP-binding; bioinformatics; DNA damage; DNA uptake; environmental stress response; gastric cancer; genomic context; *Helicobacteraceae*; *Helicobacter pylori*; natural transformation competence; universal stress proteins; virulence; *Wolinella*



Received: 5 January 2025

Revised: 9 March 2025

Accepted: 11 March 2025

Published: 13 March 2025

Citation: Isokpehi, R.D.; Simmons, S.S.; Makolo, A.U.; Hollman, A.L.; Adesida, S.A.; Ojo, O.O.; Abioye, A.O. Insights into Functions of Universal Stress Proteins Encoded by Genomes of Gastric Cancer Pathogen *Helicobacter pylori* and Related Bacteria. *Pathogens* **2025**, *14*, 275.

<https://doi.org/10.3390/pathogens14030275>

Copyright: © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license

(<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The genera in the Gram-negative bacteria family named *Helicobacteraceae* currently includes *Helicobacter*, with 55 species, and *Wolinella*, with one species [1,2]. The *Helicobacter* species are commonly associated with the gastrointestinal system of animal or human hosts, where they may be responsible for acute infections or chronic infections that could lead to malignancies including gastric cancer [3]. *Helicobacters* have been broadly grouped by anatomic niches as gastric (stomach) and enterohepatic (intestinal tract and liver) [4]. The zoonotic potential of avian *Helicobacter* species, such as *H. pullorum* and *H. canadensis*, is of public health concern [5]. *Helicobacter* species that infect the human gastrointestinal system must adapt and survive extreme conditions such as acidic gastric juice, antimicrobial compounds, bile, low pH, and oxidative stress from host immune response [6,7]. *Helicobacter pylori* is a microaerophilic, neutrophilic, and flagellated pathogen that infects the highly acidic human stomach [8].

We have previously investigated the functions of genes that encode the universal stress protein (Usp) domain (Pfam protein family: pfam00582 or PF00582) in bacteria and eukaryotes [9–15]. The universal stress proteins (USPs) are broadly classified by the presence of amino acid residues that bind to adenosine triphosphate (ATP) [16–18]. In both pathogenic and non-pathogenic bacteria, the USPs enhance adaptive survival in diverse unfavorable conditions including high osmolarity [19], iron scavenging [20], acid fluctuations [21], nutrient starvation [22,23], and antibiotic resistance [24,25]. ATP binding of mycobacteria universal stress proteins influences chronic persistent infection and intracellular survival [26,27].

The spectrum of clinical manifestations of infection by *Helicobacteraceae* presents a need for a greater understanding of the biological processes, including response to stress by this bacteria family. However, the functions of genes for *Helicobacteraceae* universal stress proteins are still limited to data in the Supplementary Files on genome-enabled investigations of *Helicobacter pylori* strains [28,29]. For example, a study of the primary transcriptome of *H. pylori* 26695 included a supplementary operon map with the USP gene (HP0031) in a 15-gene operon [28]. Thus, the goal of our data investigation project was to obtain insights into the functional associations of *Helicobacteraceae* universal stress proteins in host organisms. To achieve this goal, we designed four data investigation objectives.

Our first objective was to construct a representative set of universal stress protein sequences predicted from a collection of *Helicobacteraceae* genomes. This representative set of sequences from multiple species/strains of *Helicobacter* and *Wolinella* would provide a reference set for further sequence and phylogenetic analysis. Our second objective was to group the *Helicobacteraceae* species according to the patterns of functional sites in the *Helicobacteraceae* USPs. This objective is significant, because identical patterns of functional sites of USPs could suggest common ancestry of *Helicobacteraceae* species. In addition, amino acid residues involved in ATP-binding could infer ATP-dependent biological processes such as growth regulation, as observed with *Mycobacterium tuberculosis* [26,27,30]. Our third objective was to determine the transcription direction and functions of genes adjacent to *Helicobacteraceae* USP genes. Findings from this objective could guide further research on *Helicobacteraceae* USPs that regulate biological processes encoded by adjacent genes. In *Mycobacterium smegmatis*, an operon containing a USP and gene for acyltransferase was implicated in regulating biofilm formation [31]. Our fourth objective was to collect transcriptome (RNA-seq) and interactome (protein–protein interactions) evidence for biological activity of *H. pylori* USP gene. This last objective will help verify or validate findings from the data investigations. Altogether, results from our objectives provided insights into functions, directions for further research, and potential applications of the universal stress proteins of *H. pylori* and related bacteria.

2. Methods

2.1. Retrieval of *Helicobacteraceae* Genomes Dataset and Genes Encoding Universal Stress Protein Domain Datasets

In the first stage, we retrieved a dataset of *Helicobacteraceae* genomes that included genome identifiers (such as Genome ID and Genome Name) from the Integrated Microbial Genomes & Microbiomes Expert Review (IMG/M-ER) website [32] using a uniform resource locator (URL) search script that included the bacteria family name. The Data Availability section of this report has details of the URL search script. The dataset of genomes was stored in the IMG/M-ER Genome Cart. In the second stage, we used the IMG/MER Function Search tool with Pfam list filter to retrieve a dataset of genes annotated with the universal stress protein domain (pfam00582) for the *Helicobacteraceae* genomes retrieved into the Genome Cart in the first stage. The genome and gene datasets were saved as tab-delimited files for interactive data visualization to identify subgroups and patterns [33] in Tableau Desktop and Tableau Public (Seattle, WA, USA), a visual analytics software [34].

2.2. Grouping of *Helicobacteraceae* Species by Patterns of Functional Sites in Universal Stress Protein Sequences

We retrieved the FASTA formatted amino acid sequences for the predicted *Helicobacteraceae* universal stress proteins from the IMG/M-ER bioinformatics resource. In order to predict the functional sites, the collection of sequences was submitted to the National Center for Biotechnology Information (NCBI) Batch Web Conserved Domain (CD) Search Tool [35]. The results of the NCBI CD-Search tool for protein domain features include amino acid residue letters and positions of the functional sites (ligands). The expected count of functional sites for USPs is 12 [17]. An example of functional sites coordinates with 12 functional sites is A8, T9, D10, V38, V100, G101, T103, G104, G115, S116, V117, and T118. We constructed a protein functional sites dataset for investigations in visual analytics software [34] by deleting the comment section of the protein domain features output file from NCBI CD-Search tool. We then used visual analytics software to construct calculated fields such as patterns of the sequence of the amino acid residues of functional sites and patterns of the sequence of the amino acid positions of functional sites. A combination of (1) residue at third position of the functional sites coordinate; (2) the patterns of functional sites; and (3) protein sequence length allowed us to group *Helicobacteraceae* species. We selected the third position of the functional sites pattern because of the effect on ATP binding. In *Mycobacterium tuberculosis*, a change of aspartate (D) to glutamate (E) in the third position of the functional sites pattern of two universal stress proteins (Rv2623 and Rv2624c) abrogated ATP binding [26,27].

2.3. Similarity of *Helicobacteraceae* Universal Stress Proteins Sequences

We used the Cluster Database at High Identity with Tolerance (CD-HIT) biological sequence clustering software [36], implemented by the Galaxy@Pasteur platform [37], to cluster (at 100% identity) a set of *Helicobacteraceae* universal stress protein sequences. We constructed a dataset from the list of CD-HIT clusters of sequences with spreadsheet software, Microsoft Excel, and added strain phylogeographic annotations [38]. The non-redundant set of protein sequences produced by CD-HIT were aligned on the MAFFT multiple sequence alignment software [39]. We viewed and interacted with the multiple sequence alignment using the NCBI Multiple Sequence Alignment Viewer (version 1.25.0) [40]. In addition, phylogenetic trees visualizing the multiple sequence alignment of the representative *Helicobacteraceae* USP sequences were constructed with the Interactive Tree Of Life (iTOL) online tool for phylogenetic tree display and annotation [41] and the Environment for Tree Exploration (ETE) [42].

2.4. Genomic Context of Genes for Universal Stress Proteins in *Helicobacteraceae* Genomes

Genomic context is “information gleaned from the genes surrounding a gene of interest in the chromosome” [43]. The transcription direction, start and end coordinates, as well as functional annotations of genes surrounding a gene of interest are examples of genomic context information provided by genomic sequencing projects [44]. Additionally, an operon “is a collection of genes that are co-transcribed to form a single mRNA molecule in at least some set of biologically relevant conditions [45]”. Genes in bacterial genomes with similar function or co-expression can be clustered in a genomic region and arranged as operons [46,47]. Therefore, we obtained information on the genomic context of *Helicobacteraceae* universal stress protein genes by using genomic context tools available from microbial web portals [48].

Datasets retrieved from IMG/M by providing a list of gene identifiers (Gene ID) were the source of transcription direction (inferred from genome strand); base pair distances between the USP gene and adjacent gene; and protein family (pfam domain) annotation for the genes adjacent to the USP gene. We constructed a three-digit binary number to represent the genome strand annotation (+ or −) relative of the upstream and downstream genes to the direction of the USP gene. We assigned “1” to USP, and if the adjacent gene to the USP gene is in the same transcription direction, “1” represents the transcription direction relative to the USP gene, otherwise “0” is assigned. The possible three-digit representations and transcription direction using greater than and less than symbols are “010” (< > <), “011” (< > >), “110” (> > <), and “111” (> > >). Therefore, “111” represents that the USP gene and the two adjacent genes are in the same direction; while “010” represents that the USP gene is in a different transcription direction relative to the two adjacent genes. This representation could identify USP genes that have a different gene adjacency pattern from the majority of genomes from the same species. We designed visual analytics worksheets to integrate and group the USP genes according to gene transcription direction, location coordinates (start and stop positions), and pfam annotations of adjacent genes.

The protein function annotations using Cluster of Orthologous Groups (COGs) of proteins [49] and protein family (pfam) were obtained for adjacent genes. We used the “Conserved Neighborhood” tool of IMG to verify the gene transcription direction and the pfam annotation of the gene neighborhood group (010, 011, 110, or 111) assigned in the visual analytics worksheets. The web pages of a USP gene in BioCyc [50] were used to determine the presence of the USP gene in a putative operon.

2.5. Transcriptome and Interactome Evidence for *Helicobacter pylori* Universal Stress Protein

We searched scholarly literature databases such as Google Scholar and PubMed for studies on RNA-Sequencing (RNA-seq), transcriptome profiling, and protein–protein interaction maps of *Helicobacter pylori* that have supplementary data that included the universal stress protein HP0031. For RNA-seq studies, we prioritize studies with fold change and statistical significance data in the Supplementary Files.

3. Results

3.1. Universal Stress Protein Proteins Encoded by *Helicobacteraceae* Genomes

Genomes in the Dataset: Our query for genomes annotated as *Helicobacteraceae* in the Integrated Microbial Genome & Microbiome (IMG/M) system in January 2024 retrieved 1014 genomes from 999 unique strains. Nine *Helicobacter pylori* strains and one *Helicobacter suis* had two genome sequences in the dataset. One strain each of *Helicobacter canadensis*, *Helicobacter cinaedi*, and *Helicobacter heilmannii* also had two genomes. *Helicobacter bilis* ATCC 51630 had three genomes. Two strains (CG1_02_36_14 and CG2_30_36_10) had the annotation of unclassified *Helicobacteraceae* bacterium, obtained from a groundwater

metagenome. We retained the two genomes in the data investigation to determine the features of the universal stress proteins and possibly infer the taxonomic association. The final dataset consisted of 1045 USP genes from 1011 genomes of *Helicobacter*, two genomes of *Wolinella*, and one genome of strain CG1_02_36_10. There were 47 *Helicobacter* (33 enterohepatic, one enterohepatic and gastric, and 13 gastric) with species name in the dataset (Table 1). In addition, 26 *Helicobacter* genomes with only strain identifiers (Supplementary Table S1) were grouped by anatomic niche and host organisms criteria as described in publications (such as Mannon et al. [51], O'Toole et al. [52], and Gilbert et al. [53]) as well as genome pages of the NCBI's BioProject [54] and IMG/M system [32].

Table 1. Grouping of *Helicobacter* genomes with species name encoding universal stress proteins by anatomic niche and host organism.

<i>Helicobacter</i> Group by Anatomic Niche	Host Organism ¹	<i>Helicobacter</i> Species
Enterohepatic	Bird	<i>H. anatolicus</i> , <i>H. anseris</i> , <i>H. pullorum</i> , <i>H. valdiviensis</i>
	Dog	<i>H. canis</i>
	Hamster	<i>H. aurati</i> , <i>H. mesocricetorum</i>
	Horse	<i>H. equorum</i>
	Human	<i>H. bilis</i> , <i>H. burdigaliensis</i> , <i>H. canadensis</i> , <i>H. cinaedi</i> , <i>H. fennelliae</i> , <i>H. labetoulli</i> , <i>H. rappini</i> , <i>H. trogontum</i> , <i>H. winghamensis</i>
	Marmot	<i>H. marmotae</i> , <i>H. himalayensis</i>
	Marmoset	<i>H. jaachi</i>
	Monkey	<i>H. macacae</i> , <i>H. saguini</i>
	Mouse	<i>H. apodemus</i> , <i>H. ganmani</i> , <i>H. hepaticus</i> , <i>H. japonicas</i> , <i>H. magdeburgensis</i> , <i>H. rodentium</i> , <i>H. typhlonius</i>
	Opossum	<i>H. didelphidarum</i>
	Pig	<i>H. colisuus</i>
Enterohepatic and Gastric	Rat	<i>H. muridarum</i>
	Squirrel	<i>H. turcicus</i>
Gastric	Ferret	<i>H. mustelae</i>
	Cat	<i>H. ailurogastricus</i> , <i>H. baculiformis</i> , <i>H. felis</i> , <i>H. heilmannii</i>
	Cheetah, Lion, Tiger	<i>H. acinonychis</i>
	Dog	<i>H. bizzozeronii</i> , <i>H. cynogastricus</i> , <i>H. salomonis</i>
	Dolphin	<i>H. cetorum</i>
	Fox	<i>H. mehlei</i> , <i>H. vulpis</i>
	Human	<i>H. pylori</i>
	Pig	<i>H. suis</i>

¹ Some *Helicobacter* species have been associated with more than one host organism.

Universal Stress Protein Gene Count per Genome: In the dataset collected, there were 1045 predicted genes with the universal stress protein (USP) domain annotations from 1014 *Helicobacteraceae* genomes. In addition, 984 *Helicobacteraceae* genomes had one USP gene annotation while 29 *Helicobacteraceae* genomes had two annotations for the USP gene. The genome of *Helicobacter fennelliae* NCTC 13102 had three annotations for USP genes. In the case of *H. pylori*, 781 (99%) of the 785 genomes had annotations for one USP each. Three *H. pylori* strains with two USP genes per genome are UM163 (IMG/M Genome ID: 2639762898), UM171S (IMG/M Genome ID: 2648501688), and UM276S (IMG/M Genome

ID: 2651869863). The genome sequencing status of all the *H. pylori* strains with two USPs is permanent drafts. Strain UM037 (IMG/M Genome ID: 2545824636) genome had three genes annotated with the Usp (PF00582) domain. Of the two strains of the unclassified *Helicobacteraceae* bacterium, strain CG2_30_36_10 had annotation for one USP gene. The total count of *H. pylori* USP sequences retrieved was 789.

Lengths of Universal Stress Protein Sequences: The 24 observed protein lengths in amino acids (aa) were 66, 68, 72, 78, 88, 89, 99, 104, 111, 121, 126, 136, 137, 138, 139, 140, 141, 146, 147, 273, 274, 278, 279, and 285 (Supplementary Table S2, Figure S1). Among the 1045 USP sequences, the top two frequent protein sequence lengths were 137 aa (785 sequences) and 138 (182 sequences). The *Helicobacteraceae* USP sequence lengths observed are in two categories, from 66 aa to 147 aa and 273 aa to 285 aa. One USP gene encoding 279 aa protein sequence was predicted from the genome of unclassified *Helicobacteraceae* strain CG2_30_36_10. The protein lengths less than 137 aa were from one 88 aa USP sequence from *H. bilis* ATCC 51630, and 14 USP genes from 12 *H. pylori* genomes: A45, CPY6081, Hp238, HPJP26, HPKX_438_AG0C1, UM137R, UM137S, UM163 (2 USP genes), UM171S, UM229R, UM276R, and UM276S (2 USP genes). The 137 aa sequences were predicted from only the gastric helicobacters species (*H. acinonychis*, *H. cetorum* and *H. pylori*). Additional description of the USP sequence length is available in the footer section Table S2.

3.2. Grouping of *Helicobacteraceae* Species by Patterns of Functional Sites in Universal Stress Protein Sequences

The NCBI Batch Web Conserved Domain (CD) Search Tool predicted functional sites for 1030 (98%) of the 1045 universal stress protein sequences submitted. The output of the search tool includes coordinates (e.g., G6, I7, S8, V36, I105, G106, S108, E109, S119, H120, and Q121), the complete size of the ligand (e.g., 12), and the mapped size (e.g., 11). We constructed an Amino Acid Pattern by combining the amino acids symbols (e.g., GISVIGSE-SHQ) and the Amino Acid Position Pattern by combining numeric position of the amino acid residue separated by an underscore (e.g., 6_7_8_36_105_106_108_109_119_120_121). In the third position of the USP functional sites patterns, there were three types of residues: aspartate (D), histidine (H) and serine (S). The residue in the third position was selected because changes in the amino acid affected ATP-binding [26,27]. The design of the grouping approach used the residue in the third position, which has the highest level of the grouping, followed by the amino acid pattern and the amino acid position pattern. We used the 25 amino acid patterns and the 14 amino acid position patterns to group 47 named species of *Helicobacter*, 26 species of *Helicobacter* without assigned names (labeled sp.), *Wolinella succinogenes* and species of an unclassified related campylobacteria strain (labeled bacterium in the species column) (Figure 1). The protein sequence length associated with the functional site patterns provided an additional feature in Figure 1 to facilitate interpretations and further investigations of taxa groups.

A use of our grouping approach is for identifying species with unique or identical functional sites patterns in *Helicobacteraceae* USP sequences. All the USPs with lengths 273 aa, 274 aa, 279 aa, and 285 aa have aspartate as residue in the third position. The 138 aa *Helicobacteraceae* USPs group into those with third position residue as aspartate (D) or serine (S). Among USP sequences with aspartate (D) in the third position of functional site patterns, the 273 aa of *H. turcicus* and *H. winghamensis* have an identical amino acid pattern of functional sites (ATDIIGTGGSVA). Other examples are (1) 273 aa USPs of *H. apodemus* and *H. mesocricetorum* with pattern ATDIVGAGGSTA; and (2) 278 aa USPs of *H. burdigaliensis* and *H. valdiviensis* with pattern ATDVIGRGGSVA. The functional sites patterns of the 138 aa USP of *H. valdiviensis* is AIDVIGSESSNQ, which is distinct from AVDVIGSESSNQ from the 138 USPs of *H. burdigaliensis*, *H. turcicus*, and *H. winghamensis*. The only two 139 aa USPs with histidine (H) in the third position of functional site patterns

were encoded in the genomes of *H. anatolicus* and three genomes of *H. mustelae*. The groupings by this approach are consistent with prior studies using other methods: *H. turcicus* and *H. winghamensis* [1]; *H. apodemus* and *H. mesocricetorum* [55]; *H. burdigaliensis* and *H. valdiviensis* [56]; and *H. anatolicus* and *H. mustelae* [1].

Third Functional Site	Amino Acid Pattern	Amino Acid Position Pattern	Species	Universal Stress Protein Sequence Length											
				136	137	138	139	140	141	146	147	273	274	278	279
D	AIDIGSASSNQ	7 8 9 38 107 108 110 111 120 121 122 123	sp.			1									
	AIDVIGSESHQ	7 8 9 38 107 108 110 111 121 122 123	succinogenes			2									
	AIDVIGSESSNQ	7 8 9 38 107 108 110 111 120 121 122 123	valdiviensis			1									
	AIDVIGSETNQ	7 8 9 38 107 108 110 111 121 122 123	ganmani			1									
			sp.			1									
	AIDVIGSKSSNQ	7 8 9 38 107 108 110 111 120 121 122 123	mesocricetorum			1									
	ATDIIGAGGSVA	8 9 10 38 100 101 103 104 115 116 117 118	sp.									1			
	ATDIIGTGSVA	8 9 10 38 100 101 103 104 115 116 117 118	turcicus									1			
			winghamensis									1			
	ATDIVAGGSTA	8 9 10 38 100 101 103 104 115 116 117 118	apodemus										2		
			mesocricetorum										1		
	ATDIVGTGGSVT	8 9 10 38 100 101 103 104 115 116 117 118	pullorum										9		
	ATDVIGRGSVA	8 9 10 38 100 101 103 104 115 116 117 118	burdigaliensis											1	
			valdiviensis											1	
	ATDVIGSGGSHC	11 12 13 41 104 105 107 108 119 120 121 122	succinogenes												2
	ATDVIGTGGSVT	8 9 10 38 100 101 103 104 115 116 117 118	canadensis										3		
	ATDVVGTGGSVT	8 9 10 38 100 101 103 104 115 116 117 118	colisuis											1	
			ganmani										1		
			rodentium										1		
			sp.										1	1	
	AVDVIGSESSNQ	7 8 9 38 107 108 110 111 120 121 122 123	burdigaliensis			1									
			turcicus			1									
			winghamensis			1									
	AVDVIGSETDFS	7 8 9 38 107 108 110 111 121 125 126 127	sp.			1									
	AVDVIGSETNQ	7 8 9 38 107 108 110 111 121 122 123	canadensis			3									
			colisuis			1									
			pullorum			9									
			rodentium			1									
			sp.			1									
	AYDVMGYSGSFT	165 166 167 195 247 248 250 251 261 262 263 264	bacterium												1
H	GVRVIGSENSHQ	7 8 9 39 108 109 111 112 121 122 123 124	anatolicus				1								
			mustelae					3							
S	GISVIGSESHQ	6 7 8 37 106 107 109 110 119 120 121 122	cetorum			3									
		7 8 9 38 107 108 110 111 120 121 122 123	ailurogastricus				6								
			baculiformis				1								
			bizzozzeronii				9								
			cynogastricus				1								
			felis				22								
			heilmannii				9								
			mehlei				2								
			salomonis				6								
			sp.				2								
			suis				17								
			vulpis				1								
		9 10 11 40 109 110 112 113 122 123 124 125	suis					2							
	GISVIGSESHQ	6 7 8 36 105 106 108 109 119 120 121	pylori			1									
		6 7 8 37 106 107 109 110 120 121 122	pylori				771								
		16 17 18 47 116 117 119 120 130 131 132	pylori									1			
	GISVIGSESSNQ	6 7 8 37 106 107 109 110 119 120 121 122	acinonychis			8									
			pylori			1									
	GVSIGSENSHQ	7 8 9 38 107 108 110 111 120 121 122 123	anseris				1								
			cinaedi				37								
			fennelliae				6								
			hepaticus				1								
			jaachi				1								
			japonicus				1								
			labetoulli				1								
			macacae				2								
			magdeburgensis				1								
			marmotae				2								
			sp.				14								
			typhlonius				2								
		9 10 11 40 109 110 112 113 122 123 124 125	saguini					5							
		10 11 12 41 110 111 113 114 123 124 125 126	aurati						2						
			billis						9						
			muridarum						5						
			rappini						1						
			trogontum						5						
		15 16 17 46 115 116 118 119 128 129 130 131	didelphidarum							1					
	GVSIGSESHQ	6 7 8 37 106 107 109 110 120 121 122	pylori			1									
	GVSIGSESSHQ	7 8 9 38 107 108 110 111 120 121 122 123	sp.				2								
	GVSIGSESSHQ	7 8 9 38 107 108 110 111 120 121 122 123	canis				2								
			equorum				2								
			himalayensis				1								
			sp.				4								

Figure 1. Grouping of species of *Helicobacter* and related taxa according to functional sites patterns in 1030 universal stress protein sequences defined by length. The three amino acids in the third position of the amino acid patterns, 25 amino acid patterns and 14 amino acid position patterns provided an approach to group 47 named species of *Helicobacter*, 26 species of *Helicobacter* without assigned species name (labeled sp.), *Wolinella succinogenes* and species of an unclassified related campylobacteria strain (labeled bacterium).

The 774 *H. pylori* USPs (136 aa, 137 aa, and 147 aa lengths) shared the same functional sites pattern of GISVIGSESHQ, where the residue in the third position is serine (S). A 137 aa USP encoded by *H. pylori* Aklavik86 and eight *H. acinonychis* shared functional sites coordinates with 12 positions (G6, I7, S8, V37, I106, G107, S109, E110, S119, S120, N121, and Q122). *H. pylori* Aklavik86 and *H. acinonychis* have shared genomic features [57]. Another use of the grouping approach is to predict the taxonomic or population identity of strains without a species name. *H. ganmani* USPs with sequence lengths 138 aa and 273 aa have aspartate in the third position and grouped with USPs of the same lengths, respectively, from *Helicobacter* sp. MIT 05-5294. Phylogenetic analysis clustered *H. ganmani* and strain MIT 05-5294 in the Enterohepatic *Helicobacter* Species clade 1 (EHS 1) [51]. Furthermore, the pattern GISVIGSESHQ is common to the 138 aa USP from 11 gastric helicobacters (*H. ceterum*, *H. ailurogstricus*, *H. baculiformis*, *H. bizzozeronii*, *H. cynogastricus*, *H. felis*, *H. heilmannii*, *H. mehlei*, *H. salmonis*, *H. suis*, and *H. vulpis*) as well as *Helicobacter* sp. L8 and *Helicobacter* sp. NHP19-012. According to entries in the Genomes OnLine Database [58], strain L8 and strain NHP19-012 are gastric helicobacters.

3.3. Similarity of Helicobacteraceae Universal Stress Proteins Sequences

We used the CD-HIT biological sequence clustering software [36] to cluster 1045 *Helicobacteraceae* USP sequences into 183 representative sequences (181 *Helicobacter*, 1 *Wolinella* and 1 unclassified *Helicobacteraceae*). Consequently, our dataset of protein sequences contained 182 representative sequences from the taxonomically classified *Helicobacteraceae* species. The initial set of protein sequences and the representative sequences are available as Supplementary Files S1 and S2, respectively. The multiple sequence alignment in fasta format, guide tree in Newick format, multiple sequence alignment view, and iTOL constructed phylogenetic tree for the 183 representative USP sequence are available as Supplementary Files S3, S4, S5 and S6, respectively. The clustering process assigned the 789 *Helicobacter pylori* USP sequences to 83 clusters with 54 clusters having one USP sequence. A view of the multiple sequence alignment with the NCBI Multiple Sequence Alignment Viewer, Version 1.25.0 identified stretches (without gaps) of highly conserved and lower conserved positions in the representative USP sequences (Figure 2). A consensus motif of LLHVS (Leucine–Leucine–Histidine–Valine–Serine) was identified from the alignment.



Figure 2. A section of multiple sequence alignment of 183 representative *Helicobacteraceae* universal stress proteins. The sequences, irrespective of their length, contained two sequence regions with highly conserved (red) positions and lower conserved positions (blue). The image was constructed using the NCBI Multiple Sequence Alignment Viewer.

The clustering of the 1045 USP sequences into 183 clusters also allowed us to determine the count of USP sequence clusters per group of strains (strain group). A strain group has strains, which share features such as a genome sequencing project or publication. For example, *H. pylori* strains with “SA” as the first two letters of strain name were isolated in South Africa [59]. We present in Table 2 the count of universal stress protein sequences

and corresponding count of sequence clusters for selected strain groups with isolates from Gambia, South Africa, China, and Malaysia. In addition, we designed a worksheet in Tableau to determine unique and shared clusters among USP sequences in a strain group. The visual design of the worksheet had the following order of the fields: cluster, protein length, amino acid pattern (functional sites), genome name, and IMG Gene ID for USP gene (Figure 3 for MIT strain group and Supplementary Figure S2 for *H. pylori* GAM strain group). The MIT strain group, which is typically a *Helicobacter* species other than *H. pylori*, had 24 strains with strain name having the MIT-XX-XXXX suffix, 27 USP sequence clusters, and 31 USP sequences (Figure 3). We also observed that the IMG genome names of four MIT strains (05-5293, 11-8110, 15-1451, and 16-1353) did not include the suffix MIT. Furthermore, the full scientific name for strain 15-1451 is now *Helicobacter monodelphidis* MIT 15-1451 [60].

Table 2. Counts of universal stress protein sequences and unique sequence clusters for selected strain groups of *Helicobacter pylori*.

Strain Group ¹	Country of Isolation	Count of Strains	Count of Universal Stress Protein Sequences	Count of Sequence Clusters	Reference for Strain Group
GAM	Gambia	45	45	4	
SA	South Africa	97	97	11	[59,61]
Hpfe	China	95	95	21	[62]
UM	Malaysia	61	68	17	[63,64]
CPY	Japan	9	9	5	[65]
PUNO	Peru	9	9	4	[66]
Hp	USA	74	74	11	[65]

¹ Prefix associated with the strain designation.

Cluster	Protein Length	Amino Acid Pattern	Genome Name	Gene ID USP
>Cluster 4	274	ATDVIGTGGSVT	Helicobacter canadensis MIT 98-5491, ATCC 700968	643921339 2807233395
>Cluster 5	274	ATDIVGTGGSVT	Helicobacter pullorum MIT 98-5489	643925486
>Cluster 9	274	ATDIVGAGGSTA	Helicobacter apodemus MIT-03-7007	2661413769
>Cluster 18	273	ATDVVGTGGSVT	Helicobacter sp. MIT 05-5294	2665177338
>Cluster 19	273	ATDIIGAGGSVA	Helicobacter sp. MIT 11-5569	2702849913
>Cluster 20	273	ATDVVGTGGSVT	Helicobacter ganmani MIT 99-5101	8032104691
>Cluster 23	146	GVSIVIGSENSHQ	Helicobacter didelphidarum MIT 17-337	2996945265
>Cluster 27	141	GVSIVIGSENSHQ	Helicobacter aurati MIT 97-5075	2996965264
>Cluster 30	140	GVSIVIGSENSHQ	Helicobacter saquinii MIT 97-6194	2702845445
>Cluster 36	138	AVDVIGSETNQ	Helicobacter canadensis MIT 98-5491, ATCC 700968	643920184 2807232627
>Cluster 37	138	GVSIVIGSENSHQ	Helicobacter magdeburgensis MIT 96-1001	2702851514
>Cluster 38	138	AVDVIGSETNQ	Helicobacter pullorum MIT 98-5489	643925527
>Cluster 43	138	GVSIVIGSENSHQ	Helicobacter macacae MIT 99-5501	2559319580
			Helicobacter sp. MIT 01-3238	8018515419
>Cluster 51	138	GVSIVIGSENSHQ	Helicobacter sp. MIT 03-1616	2683189729
>Cluster 52	138	AIDVIGSETNQ	Helicobacter sp. MIT 05-5294	2665178997
>Cluster 53	138	GVSIVIGSENSHQ	Helicobacter jaachi MIT 09-6949	2665304058
>Cluster 55	138	GVSIVIGSENSHQ	Helicobacter sp. MIT 03-1614	2702847725
>Cluster 56	138	AVDVIGSETDFS	Helicobacter sp. MIT 11-5569	2702849689
>Cluster 57	138	GVSIVIGSENSHQ	Helicobacter japonicus MIT 01-6451	2793530219
>Cluster 68	138	GVSIVIGSENSHQ	Helicobacter marmotae MIT 98-6070	2916539665
>Cluster 69	138	GVSIVIGSENSHQ	Helicobacter anseris MIT 04-9362	2916541842
>Cluster 78	138	GVSIVIGSESSHQ	Helicobacter sp. MIT 00-7814	2996969329
			Helicobacter sp. MIT 99-10781	8018510255
>Cluster 87	138	GVSIVIGSENSHQ	Helicobacter sp. MIT 14-3879	8018517860
>Cluster 88	138	GVSIVIGSESSHQ	Helicobacter equorum MIT 12-6600	8032085575
>Cluster 90	138	AIDVIGSETNQ	Helicobacter ganmani MIT 99-5101	8032105084
>Cluster 110	137	GISVIGSEGSHQ	Helicobacter cetorum MIT 99-5656	2514075592
>Cluster 111	137	GISVIGSEGSHQ	Helicobacter cetorum MIT 00-7128	2514124516

Figure 3. Clusters, sequence lengths, and amino acid patterns of functional sites for 31 universal stress proteins in 27 clusters of 24 strains of the *Helicobacter* MIT strain group. The visualization contains multiple findings. For example, the two genomes of *Helicobacter canadensis*, MIT 98-5491 and ATCC 700968, each encode two genes for universal stress proteins with amino acid lengths of 274 and 138. In addition, strain MIT 01-3238 in Cluster 43 could be a strain of *Helicobacter macacae*. Strains MIT 00-7814 and MIT 99-10781 in Cluster 78 could be strains of the same *Helicobacter* species.

3.4. Predicted Protein Functions for Adjacent Genes of Universal Stress Protein Genes in *Helicobacteraceae* Genomes

We observed 36 predicted protein domains (pfam families) annotated for genes adjacent and in the same transcription direction with *Helicobacteraceae* USP genes. The encoded proteins have amino acid sequence lengths of 137 aa, 138 aa, 139 aa, 140 aa, 141 aa, 146 aa, 147 aa, 273 aa, 274 aa, 278 aa, and 285 aa (Supplementary Figure S3). Some proteins encoded by genes adjacent to USP genes had more than one protein domain. In addition, some adjacent genes in the same transcription direction did not have predicted protein functions. We used the four three-digit binary numbers (010, 011, 110, and 111) to analyze the gene adjacency pattern of each USP gene. Our visual analytics filtering procedure assigned 1009 *Helicobacteraceae* USP genes, located in the same IMG genome scaffold source as their adjacent genes, into one of four possible three-digit binary numbers representing gene transcription direction: 010 (11 USP genes), 110 (320 USP genes), 011 (599 USP genes), or 111 (119 USP genes) (Table S3). In Supplementary Figure S4, we present examples of *Helicobacteraceae* USP genes in the four patterns of gene transcription direction obtained from the IMG/M system for three USP sequence lengths: 137 aa, 138 aa, and 273 aa. The genome of *Helicobacter winthamensis* ATCC BAA-430 encodes USP genes with gene transcription pattern 010 (138 aa) and 110 (273 aa). A complete figure with 23 USP example genes according to Table S3 is available as Supplementary File S7.

Among the USP genes of the *Helicobacter* genus, we prioritized six categories of biological processes for which the protein encoded by adjacent genes could function (Table 3). Genes encoding DNA uptake were adjacent to USP genes in the two categories of length (conjugal transfer protein TraF for < 200 aa and ComEC/Rec2 family competence protein for > 200 aa) (Figure 4). Five other biological process categories that we prioritized were (1) amino acid transport and metabolism; (2) coenzyme transport and metabolism; (3) membrane transport; (3) posttranslational modification, protein turnover, chaperones; (4) replication, recombination and repair; and (5) mediation of protein–protein interactions.

Table 3. Biological process categories of predicted proteins encoded by adjacent genes in the same transcription direction as universal stress protein (USP) genes of *Helicobacter* species.

Universal Stress Protein Sequence Length Category	Biological Process Category of Predicted Function of Adjacent Gene to Universal Stress Protein Gene	<i>Helicobacter</i> Species
<200 aa	Energy production and conversion	<i>H. aurati</i> , <i>H. muridarum</i>
	Amino acid transport and metabolism	<i>H. bilis</i> , <i>H. canadensis</i> , <i>H. colisuus</i> , <i>H. pullorum</i> , <i>H. macacae</i>
	Coenzyme transport and metabolism	<i>H. anatolicus</i> , <i>H. cinaedi</i> , <i>H. equorum</i> , <i>H. hepaticus</i> , <i>H. himalayensis</i> , <i>H. jaachi</i> , <i>H. japonicus</i> , <i>H. labetoulli</i> , <i>H. magdeburgensis</i> , <i>H. marmotae</i> , <i>H. mesocricetorum</i> , <i>H. typhlonius</i> , <i>H. valdiviensis</i>
	Replication, recombination, and repair	<i>H. anatolicus</i> , <i>H. mustelae</i>
	Posttranslational modification, protein turnover, and chaperones	<i>H. acinonychis</i> , <i>H. ailurogastricus</i> , <i>H. baculiformis</i> , <i>H. bizzozeronii</i> , <i>H. cetorum</i> , <i>H. cynogastricus</i> , <i>H. felis</i> , <i>H. heilmannii</i> , <i>H. mehlei</i> , <i>H. mustelae</i> , <i>H. pylori</i> , <i>H. salomonis</i> , <i>H. suis</i> , <i>H. vulpis</i>
	DNA Uptake	<i>H. canis</i> , <i>H. cinaedi</i> , <i>H. fennelliae</i> , <i>H. hepaticus</i> , <i>H. jaachi</i> , <i>H. japonicus</i> , <i>H. labetoulli</i> , <i>macacae</i> , <i>H. magdeburgensis</i> , <i>H. marmotae</i> , <i>H. typhlonius</i>
	Mediation of protein–protein interactions	<i>H. ganmani</i>
>200 aa	DNA Uptake	<i>H. canadensis</i> , <i>H. colisuus</i> , <i>H. ganmani</i> , <i>H. pullorum</i> , <i>H. rodentium</i> , <i>H. turcicus</i> , <i>H. valdiviensis</i> , <i>H. winthamensis</i>
	Membrane transport	<i>H. ganmani</i> , <i>H. rodentium</i>

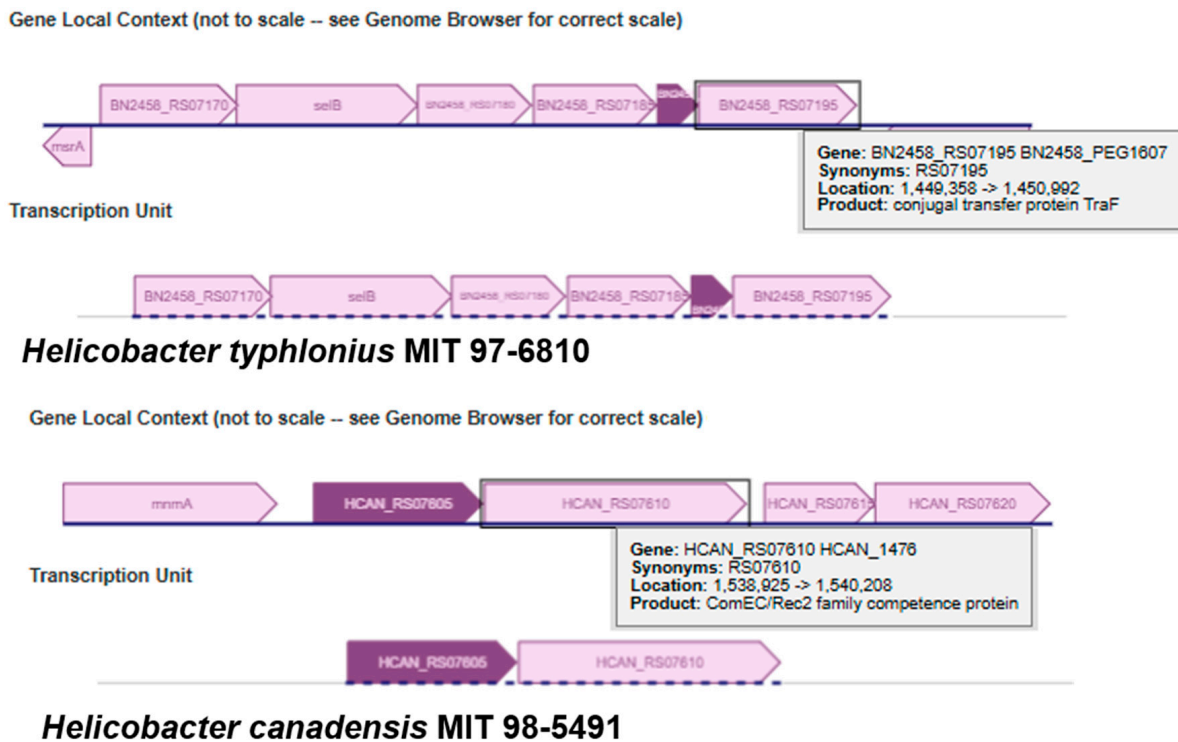


Figure 4. BioCyc diagrams of gene local context and transcription unit for genes for universal stress proteins adjacent to genes for proteins involved in DNA uptake. The universal stress protein (USP) gene image is darker than neighboring genes. The transcription unit indicates genes that could be in an operon. According to BioCyc Operons tab in the gene page, a dashed baseline indicates that there is no high-quality evidence to confirm the extent of this transcription unit. *Top diagram:* Gene encoding a 138 aa universal stress protein (USP) in the genome of *Helicobacter typhlonius* MIT 97-6810. *Bottom diagram:* Gene encoding a 274 aa USP in the genome of *Helicobacter canadensis* MIT 98-5491. Interactive diagrams can be accessed at the BioCyc website.

We observed that the adjacent genes annotated with “posttranslational modification, protein turnover, chaperones” biological process category occurred exclusively in gastric helicobacters (Tables 1 and 3). Furthermore, within the biological process category, there were three ATP-dependent protein families: (1) ATP-dependent Clp protease ATP-binding subunit ClpA [COG0542]; (2) ATP-dependent serine protease [COG1066]; and (3) ATP-dependent Clp protease adapter protein ClpS [COG2127] (Figure S4). The second protein family was encoded by an adjacent gene to the USP gene of *H. mustelae*. The helicobacters of three species (*H. acinonychis*, *H. cetorum* and *H. pylori*) had the gene for ClpS adjacent to the USP gene (Figure S5). An operon could contain one or both of the adjacent genes to a *Helicobacter* USP gene in our prioritized biological processes (Figures 5 and S5). The number of genes predicted to be in the operon with the *H. pylori* USP gene was seven or 11 and included the genes for five DNA uptake competence proteins (ComB6, ComB7, ComB8, ComB9 and ComB10) (Figure 5). The gene for ComB7 is 117 bp and encodes a 35 aa protein. Finally, in the case of taxonomically unclassified *Helicobacteraceae* strain CG2_30_36_10, a search for gene neighborhood similarity in the IMG/M system identified similar regions in campylobacterial genomes of *Sulfurospirillum arcachonense* and *Arcobacter* species. The gene for sulfate transport is adjacent to the USP gene of strain CG2_30_36_10.



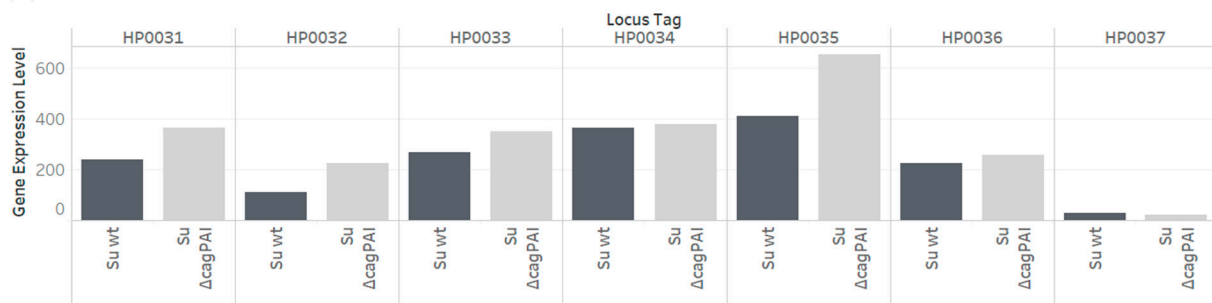
Figure 5. BioCyc multi-genome alignment of the genomic region of nine *Helicobacter pylori* genomes that includes the gene for universal stress protein. The *H. pylori* gene for universal stress protein (USP) image has diagonal lines. The predicted operon (grey filled rectangle) containing the USP gene in all strains except strain 26695 consists of 11 genes starting with the USP gene (HP0031) and ending with the DNA type IV secretion protein ComB10 (HP0041). Other predicted proteins in the operon are (1) HP0032: ATP-dependent Clp protease adapter protein ClpS; (2) HP0033: ATP-dependent Clp protease ATP-binding subunit ClpA; (3) HP0034: L-aspartate 1-decarboxylase; (4) HP0035: nucleoid-associated protein EbfC; (5) HP0036: PDZ domain protein; (6) HP0037: ComB6 competence protein; (7) ComB7; (8) HP0038: ComB8; (9) HP0040: ComB9; and (10): HP0041: ComB10. Interactive diagrams can be accessed at the BioCyc website.

3.5. Transcriptome and Interactome Evidence for *Helicobacter pylori* Universal Stress Protein

We identified two relevant genome-wide gene expression datasets to identify biological processes that regulate the gene for *H. pylori* USP. The datasets were the National Center for Biotechnology Institute (NCBI) Gene Expression Omnibus (GEO) GSE227450 [67] and the European Bioinformatics Institute (EBI) ArrayExpress E-MTAB-13025 [68]. The publications associated with the studies included spreadsheet files with data on fold change and statistical significance for *H. pylori* genes including USP and its neighboring genes such as the ClpS gene.

The GSE227450 dataset includes an RNA-sequencing (RNA-Seq) comparison of gene transcription between parental strain (Su wt) of *H. pylori* and a strain of *H. pylori* with a deficient cytotoxic-associated gene-pathogenicity island (Su Δ cagPAI). The ArrayExpress E-MTAB-13025 dataset includes an RNA-Seq comparison of gene transcription between *H. pylori* incubated in microaerophilic (optimal, WT) conditions and *H. pylori* incubated in aerobic (oxidative stress, WTS) conditions. In both experiments, the genes for USP (HP0031) and ClpS (HP0032) had statistically significant fold change values (Figure 6).

(a) Virulence



(b) Oxidative Stress

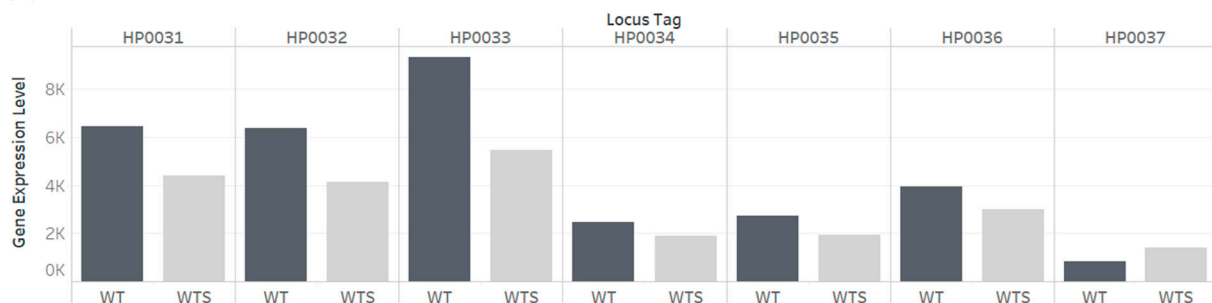


Figure 6. Evidence of expression of genes for *Helicobacter pylori* universal stress protein (HP0031) and ATP-dependent Clp protease adapter protein ClpS (HP0032) and other neighboring genes. (a) Gene expression of wild type strain (Su wt) and strain deficient in the cytotoxic-associated gene-pathogenicity island (Su Δ cagPAI). *Helicobacter pylori* Su isolate is from Northeast Africa, Sudan population [69]. (b) Gene expression of wild type strain (WT, incubated with 2% oxygen) and strain (WTS, incubated with 21% oxygen). (1) HP0033: ATP-dependent Clp protease ATP-binding subunit ClpA; (2) HP0034: L-aspartate 1-decarboxylase; (3) HP0035: nucleoid-associated protein EbfC; (4) HP0036: PDZ domain protein; (5) HP0037: ComB6 competence protein.

We identified a protein–protein interaction dataset of *H. pylori* that included interaction partners for universal stress protein, HP0031 [70]. Using a high-throughput strategy of the yeast two-hybrid assay, seven proteins encoded by genes with locus tags HP0006, HP0066, HP0281, HP1041, HP1513, and HP1567 interacted with *H. pylori* USP (HP0031) (Table 4).

Table 4. Proteins with evidence of interaction with *Helicobacter pylori* universal stress protein (HP0031).

Locus Tag *	Gene Symbol	Protein Name	Function Category
HP0006	<i>panC</i>	Pantoate—beta-alanine ligase	Coenzyme transport and metabolism
HP0066		DNA translocase FtsK	Cell cycle control, cell division, and chromosome partitioning
HP0281	<i>tgt</i>	tRNA guanosine(34) transglycosylase Tgt	Nucleotide transport and metabolism
HP1041	<i>flhA</i>	Flagellar biosynthesis protein FlhA	Cell motility
HP1513		L-seryl-tRNA(Sec) selenium transferase	Translation, ribosomal structure, and biogenesis
HP1567		Ribosome biogenesis GTP-binding protein YihA/YsxC	Cell cycle control, cell division, and chromosome partitioning

* Data source is Supplementary Data from protein–protein interaction map of *H. pylori* [70].

3.6. Interactive Analytics Resources for Investigating Dataset on *Helicobacteraceae* Genomes and Universal Stress Protein Genes

We constructed a dataset of 1045 USP genes encoded by 1011 *Helicobacter* genomes, two *Wolinella*, and an unclassified *Helicobacteraceae* bacterium CG2_30_36-10. The dataset has 46 data columns that are in five categories, and the count of data columns is Gene Annotation (18), Gene Adjacency (7), Genome Annotation (9), Phylogeography of *H. pylori* (5), and Protein Sequence (7). This dataset and description of each data column is available as Supplementary File S8. We designed the dataset and web resource for researchers to use the USP sequence clusters and gene adjacency distances to group *H. pylori* populations according to geographic or *H. pylori* population factors. A subset of 773 *H. pylori* USPs were identified to be adjacent to the gene for ClpS and have a protein length of 137 aa. The sequence analysis and gene adjacency procedures respectively clustered the USPs by sequence similarity to 78 clusters and grouped the USPs by the 24 distances between the genes for ClpS and USP. A cluster of 474 USP sequences (Cluster 22) was the largest cluster, followed by Cluster 102 with 119 USP sequences.

Combining the two grouping procedures with the distance pattern as the first level and sequence similarity as the second level revealed groups with multiple *H. pylori* populations or from a specific *H. pylori* population (examples in Table 5). The distances between the USP gene and adjacent genes did not change in experiments that determined changes such as mutation to genomic regions of input and output strains, as observed with BCM-300 and HE labeled strains as well as J166 and J166output strains.

We developed an online (Tableau Public) web resource consisting of interactive analytics worksheets and dashboards that allow for further data investigations and research on several aspects of *Helicobacteraceae* universal stress proteins. The website address to the resource is provided in the Data Availability section. The online resource includes interactive versions of figures in this report. In addition, static images of all the figures in this report are also available for viewing in the web resource. We present an example of interactive analytics worksheet with filters, named “usp_seq_clusters_hyplori_pop” (Figure 7). This interactive worksheet displays values of variables for 84 USP genes by categories of gene annotation, genome annotation, protein sequence features, gene adjacency, and phylogeography of genes for universal stress proteins encoded by *H. pylori* genomes.

The visual revealed the subgroups within clusters of USP genes according to the gene adjacency distances (base pairs between the USP gene and each adjacent gene), population ancestry of *H. pylori* isolate, pattern of amino acid functional sites, and the binary number of the transcription direction of the USP gene and adjacent genes. Among isolates with assigned population ancestry, the 11 distance (bp) values observed between the USP gene and the gene for ClpS (ATP-dependent caseinolytic protease (Clp) adaptor protein) were 2, 7, 8, 15, 20, 22, 23, 31, 32, 36, and 53. The grouping by gene adjacency distances revealed USP

sequences from different *H. pylori* populations that share values of gene adjacency distances. For example, hspAfrica1WAfrica (GAM101Bv and GAM115Ai) and hspEAsia (HLJHP253) have the 2-bp between the USP and ClpS genes. We also observed findings unique to strain groups. All the 45 GAM (isolates from The Gambia, West Africa) genomes have the 2-bp distance. Among 97 South Africa (SA) *H. pylori* genomes, four base pair distances (2, 15, 32, and 44) occurred between the genes for USP and ClpS. The gene adjacency (−104:136) and transcription direction pattern (010) for strain Sat464 (*H. pylori* population: hspIndigenousSAmerica) are atypical compared with other *H. pylori* strains because in the IMG/M genome annotation there is an adjacent upstream gene in opposite genome strand (IMG/M Gene ID: 648251455) to the USP Gene (IMG/M Gene ID: 648251454). The ClpS gene is on the same strand as the USP gene and has the IMG Gene ID of 648251456.

Table 5. Groupings of *Helicobacter pylori* populations by combining gene adjacency distances and sequence similarity of amino acid sequences for *H. pylori* universal stress protein.

Gene Adjacency Distances *	Sequence Cluster	<i>H. pylori</i> Population or County and Strains in Sequence Cluster	Notes and References
8:171	Cluster 22	hpEurope: BCM-300, HE101/09, HE132/09, HE136/09, HE141/09, HE142/09, HE143/09, HE147/09, HE170/09, HE171/09, HE178/09	HE labeled strains are re-isolates from human volunteers experimentally challenged with strain BCM-300 [71].
10:111	Cluster 22	hpEurope: J166, J166output_1moA, J166output_1moB, J166output_1moC, J166output_1wkA, J166output_1wkB, J166output_1wkC, J166output_2moA, J166output_2moB, J166output_2moC, J166output_6moA, J166output_6moB, J166output_6moC	J166 output isolates are from time points after experimental infection of a rhesus macaque with J166 [72].
31:109	Cluster 22	hpEurope: 26695, 26695 dRdM2addM2, 26695-1, 26695-1CH, 26695-1CL, 26695-1MET, 26695-dR, 26695-dRdM1dM2, 26695-dRdM2, dRdM1, G27, HP2RS, Rif1, Rif2	The strains are derived from or closely related to strain 26695 [73]. HP2RS indicates <i>H. pylori</i> 26695-related sequence [74]. G27 and 26695 have the same fundamental structure of the lipopolysaccharide of the outer membrane protein, a key factor for colonization and persistence in gastric niche [75].
31:109	Cluster 102	Australia: JCM 12093, CCUG 17874 hpEAsia: Hpfe0001	JCM 12093, CCUG 17874, NCTC 11637, and ATCC 43504 are equivalent strains [58]. CCUG 17874 is a CagA and VacA producing strain [76,77]. There is genomic sequence evidence that Hpfe0001 is closely related to CCUG 17874 and ATCC 43504 [78].
44:126	Cluster 103	hpAfrica2: SA47A, SA47C	<i>H. pylori</i> isolates from atrium (SA47A) and corpus (SA47C) stomach regions of same individual [59,61].

* First number is the distance (base pairs) between the ClpS gene and the USP gene. The second number is the distance between the USP gene and the other adjacent gene.

helicobacters_usp_genes by geubic

introduction

data_column_descriptions

usp_per_genome

genome_specific_ecosystem

genome_groups_hosts

genome_usp_seq_length

usp_sequence_clusters

usp_seq_clusters_hpylori_pop

function_adjacent_genes

usp_functional_sites

strain_group_counts

strain_usp_site_pattern

Integration of Gene Annotations, Genome Annotations, Protein Sequence Features, Gene Adjacency, and Phylogeography of Genes for Universal Stress Proteins Encoded by *Helicobacteraceae* Genomes

The default view is a result of filtering by universal stress protein (USP) genes from genomes with assigned *Helicobacter pylori* (Hp) population. Change values of filters to display a new list of USP genes.

Cluster	Protein Length	Gene Adjacency Distances	Hp population	Amino Acid Pattern	Transcription Direction	Gene ID USP	Genome Name	Genome ID	Pfam Before	Pfam After	
>Cluster 22	137	GA_2:108	hspAfrica1WAfrica	GISVIGSESHQ	TD_011	2530669389	Helicobacter pylori GAM1018Iv	2529292961	Null	pfam02617===ClpS	+
					TD_110	2530661628	Helicobacter pylori GAM115Ai	2529292956	pfam02617===ClpS	Null	+
			hspEAsia	GISVIGSESHQ	TD_110	2552274823	Helicobacter pylori HLJHP253	2551306198	pfam02617===ClpS	Null	+
		GA_2:111	hspAfrica1WAfrica	GISVIGSESHQ	TD_011	2530668125	Helicobacter pylori GAM103Bi	2529292960	Null	pfam02617===ClpS	+
						2530671137	Helicobacter pylori GAM120Ai	2529292962	Null	pfam02617===ClpS	+
						2536165951	Helicobacter pylori GAM245Ai	2534681970	Null	pfam02617===ClpS	+
						2536172020	Helicobacter pylori GAM201Ai	2534681973	Null	pfam02617===ClpS	+
						2536604317	Helicobacter pylori GAM100Ai	2534682105	Null	pfam02617===ClpS	+
					TD_110	2530659708	Helicobacter pylori GAM118Bi	2529292955	pfam02617===ClpS	Null	+
						2530673040	Helicobacter pylori GAM119Bi	2529292963	pfam02617===ClpS	Null	+
						2536163724	Helicobacter pylori GAM239Bi	2534681968	pfam02617===ClpS	Null	+
						2536170229	Helicobacter pylori GAM121Aii	2534681972	pfam02617===ClpS	Null	+
		GA_2:112	hspAfrica1LatinAmerica	GISVIGSESHQ	TD_011	2514198323	Helicobacter pylori PeCan18	2513237210	Null	pfam02617===ClpS	+
			hspAfrica1WAfrica	GISVIGSESHQ	TD_011	637022041	Helicobacter pylori J99	637000134	Null	pfam02617===ClpS	+
		GA_2:145	hspEAsia	GISVIGSESHQ	TD_110	643093196	Helicobacter pylori 98-10	642979330	pfam02617===ClpS	Null	+
						2552273256	Helicobacter pylori HLJHP271	2551306197	pfam02617===ClpS	Null	+
		GA_2:147	hspEAsia	GISVIGSESHQ	TD_011	2552275757	Helicobacter pylori HLJHP256	2551306199	Null	pfam02617===ClpS	+
		GA_2:149	hspEAsia	GISVIGSESHQ	TD_011	2551942886	Helicobacter pylori HLJHP193	2551306138	Null	pfam02617===ClpS	+
		GA_8:107	hspSahul	GISVIGSESHQ	TD_011	2720368837	Helicobacter pylori PNG64A	2718218166	Null	pfam02617===ClpS	+
		GA_15:109	hspNEurope	GISVIGSESHQ	TD_011	638019854	Helicobacter pylori HPAG1	637000133	Null	pfam02617===ClpS	+
		GA_15:179	hspIndigenousSAmerica	GISVIGSESHQ	TD_011	2514126282	Helicobacter pylori Shi112	2513237188	Null	pfam02617===ClpS	+
		GA_20:481	hspEAsia	GISVIGSESHQ	TD_110	646899833	Helicobacter pylori 51	646862330	pfam02617===ClpS	Null	+
		GA_22:109	hspEAsia	GISVIGSESHQ	TD_110	8007016336	Helicobacter pylori 52	8007015780	pfam02617===ClpS	Null	+
		GA_23:108	hspEAsia	GISVIGSESHQ	TD_110	646940281	Helicobacter pylori 52	646862329	pfam02617===ClpS	Null	+
		GA_31:86	hspIndigenousNAmerica	GISVIGSESHQ	TD_011	2522041472	Helicobacter pylori Aklavik117	2521172715	Null	pfam02617===ClpS	+
		GA_31:109	hspNEurope	GISVIGSESHQ	TD_011	637017765	Helicobacter pylori 26695	637000132	Null	pfam02617===ClpS	+
						2523150328	Helicobacter pylori 26695	2522572170	Null	pfam02617===ClpS	+
			hspSEurope	GISVIGSESHQ	TD_011	643382409	Helicobacter pylori G27	643348558	Null	pfam02617===ClpS	+
		GA_32:110	hspAfrica1SAfrica	GISVIGSESHQ	TD_011	2620889972	Helicobacter pylori SA220C	2619619171	Null	pfam02617===ClpS	+
		GA_32:111	hspAfrica1SAfrica	GISVIGSESHQ	TD_011	2612511065	Helicobacter pylori SA156C	2609460339	Null	pfam02617===ClpS	+
						2619009843	Helicobacter pylori SA168C	2617271177	Null	pfam02617===ClpS	+
						2619039467	Helicobacter pylori SA170C	2617271187	Null	pfam02617===ClpS	+
						2620906112	Helicobacter pylori SA227C	2619619180	Null	pfam02617===ClpS	+
						2620939277	Helicobacter pylori SA45C	2619619201	Null	pfam02617===ClpS	+
						2620942458	Helicobacter pylori SA46C	2619619203	Null	pfam02617===ClpS	+
					TD_110	2618965305	Helicobacter pylori SA157A	2617271158	pfam02617===ClpS	Null	+
						2618969757	Helicobacter pylori SA158C	2617271161	pfam02617===ClpS	Null	+
						2620867908	Helicobacter pylori SA146C	2619619157	pfam02617===ClpS	Null	+
						2620901959	Helicobacter pylori SA226A	2619619178	pfam02617===ClpS	Null	+
						2620915102	Helicobacter pylori SA252C	2619619186	pfam02617===ClpS	Null	+
						2620926634	Helicobacter pylori SA30C	2619619193	pfam02617===ClpS	Null	+
						2620928519	Helicobacter pylori SA35C	2619619194	pfam02617===ClpS	Null	+
		GA_32:112	hspAfrica1WAfrica	GISVIGSESHQ	TD_110	2672758665	Helicobacter pylori J99	2671180434	pfam02617===ClpS	Null	+
		GA_-104:136	hspIndigenousSAmerica	GISVIGSESHQ	TD_010	648251454	Helicobacter pylori Sat464	648231710	Null	Null	+
>Cluster 97	137	GA_8:179	hspIndigenousSAmerica	GISVIGSESHQ	TD_011	642635679	Helicobacter pylori Shi470	642555131	Null	pfam02617===ClpS	+
>Cluster 98	137	GA_36:111	hspSEurope	GISVIGSESHQ	TD_011	643415016	Helicobacter pylori P12	643348559	Null	pfam02617===ClpS	+

Hp population

(Multiple values)

Cluster

(All)

Helicobacter Group

+ Gastric

Helicobacter Group

(All)

Genome Name

(All)

Species

(All)

Transcription Pattern

(All)

Pfam Before

/

Pfam After

/

Gene Adjacency Dist...

(All)

Strain Group

(All)

Gene Adjacency Dist...

/

Protein Length

(All)

Figure 7. A section of an image from the web resource for interacting with the dataset composed of values for annotations of universal stress protein sequences encoded by *Helicobacter pylori* genomes with assigned population. The categories of annotations in the visual are (1) Gene Annotations (Gene ID USP); Genome Annotations (Genome ID, Genome Name, Helicobacter Group, Species, and Strain Group); Protein Sequence Features (amino Acid Pattern, Cluster, and Protein Length), Gene Adjacency (Gene Adjacency Distances, Pfam Before, Pfam After, and Transcription Direction); and Phylogeography of *H. pylori* isolates (Hp population). The insert rectangle shows filters that can be modified to change the list of genes displayed. The population ancestry was obtained from The *Helicobacter pylori* Genome Project [38]. We obtained the additional annotations from the Integrated Microbial Genomes/Microbiomes (IMG/M) system or derived them with bioinformatics/visual analytics procedures.

4. Discussion

We have conducted an in-depth data investigation of universal stress proteins (USPs) encoded by genomes of *Helicobacter pylori* and related *Helicobacteraceae* bacteria. To facilitate reuse of data, research, and applications, we produced interactive analytics resources of a dataset composed of values for variables on *Helicobacteraceae* universal stress protein, including phylogeography of *H. pylori* strains, protein sequence features, and gene neighborhood. We combined bioinformatics and visual analytics approaches to investigate 1045 universal stress protein sequences encoded in 1014 *Helicobacteraceae* genomes, including 785 *Helicobacter pylori* genomes. The study generated a representative set of 183 universal stress protein (USP) sequences consisting of 180 *Helicobacter* sequences, two *Wolinella succinogenes* sequences, and a sequence from a related campylobacteria, which could serve as an outgroup sequence. This non-redundant collection of 183 USP sequences and the initial 1045 USP sequences are available as Supplementary Files. These sequences can be used for structural, evolutionary, and functional studies of *Helicobacteraceae* USPs. Such studies have helped to deepen the knowledge of universal stress proteins in different organisms including pathogens [17,26,30,79–81].

Genomic context searches and analysis identified USP genes of gastric and enterohepatic helicobacters that are in adjacency or operonic arrangements with genes for oxidative stress response proteins (ATP-dependent proteases: ClpS and ClpA); and DNA uptake proteins (natural competence for transformation proteins: ComB6, ComB7, ComB8, ComB9, ComB10, ComBE, and conjugative transfer signal peptidase: TraF). The ComB6, ComB7, ComB9, and ComB10 are necessary for the first step of DNA uptake in *H. pylori* [68,82]. In *Escherichia coli*, the genes for ClpS and ClpA occur as an operon and are induced by environmental hydrogen peroxide that can diffuse into the cell [83]. Furthermore, ClpS interacts with ClpA in an ATP-dependent manner [84]. The *E. coli* protein complex, ClpSAP (ClpS, ClpA, and ClpP) balances the protection against DNA damage (due to hydroxyl radicals produced from the reaction of hydrogen peroxide with ferrous cation) and the availability of iron required for iron-dependent enzymes for DNA repair pathways and cell viability [83]. A mechanism for the development of gastric cancer is that severe oxidative stress in the gastric mucosa in response to *H. pylori* colonization leads to increased protein expression of virulence factors (CagA, VacA, and AlpA) that cause repeated inflammation of gastric epithelial cells and increased production of antioxidant enzymes (AhpC, KatA, and HtrA) to protect *H. pylori* [85]. A search with the chromosomal cassette search tool in the IMG/M database for genomes that contain genes for USP, ClpS, and ClpA in same chromosomal cassette revealed that the genomic context where gene for USP is adjacent downstream to gene for ClpS is unique to the gastric helicobacters (*H. acinonychis*, *H. ceterum*, and *H. pylori*). Experimental studies will help verify genomic context and elucidate functions of the gene neighborhood of *Helicobacteraceae* USP genes in oxidative stress response and virulence.

Our analysis of datasets from transcript profiling experiments verified that *H. pylori* USP and ClpS genes are expressed in wild-type strains and are regulated by the cagPAI region and oxidative stress condition [67,68] (Figure 6). The cagPAI genomic region encodes a type IV secretion system (T4SS), which is a virulence factor that is involved in carcinogenesis [86]. Furthermore, we provided protein–protein interaction evidence for the interaction of *H. pylori* USP with seven proteins, including (1) DNA translocase FtsK with ATP-binding motif, and (2) flagellar biosynthesis protein FlhA [70]. A DNA recombination/repair protein (RecA)/FtsK-dependent regulatory pathway regulates the universal stress protein A of *Escherichia coli* [87]. The *H. pylori* FlhA is part of the type III secretion system of the flagellar basal body [88]. Future studies could determine the genes that regulate USP and ClpS genes. These studies could provide evidence for gene networks that

involve both genes in *H. pylori* and related gastric helicobacters, such as *H. acinonychis* and *H. cetorum*.

We used the functional site patterns of 1030 universal stress proteins predicted with NCBI Conserved Domain search to group *Helicobacter* species, *Wolinella succinogenes* and an unclassified related bacterium (Figure 1). The predictions revealed the presence of 11 and 12 functional sites among 25 amino acid patterns and 14 amino acid position patterns. There was independent evidence, including phylogenetic analysis, to support the use of the universal stress protein-based grouping for assigning taxonomic or population annotations to genomes. The *Helicobacteraceae* USPs with protein sequence lengths 273 aa, 274 aa, 278 aa, and 285 aa have functional sites patterns in which the three typical glycine residues are present with 10 residues between the second and third glycine residues in the known ATP-binding motif. Thus, the predicted ATP-binding motif in the *Helicobacter* and *Wolinella* USPs is G2XG10GXS compared with the typical G2XG9XG(S/T) of the ATP-binding *Methanocaldococcus jannaschii* USP [17,18]. These functional site patterns matched with protein lengths can provide a guide for planning investigations on interactions of USPs of *Helicobacter* and *Wolinella* with ATP and other molecules. Publications on *Helicobacter* genes and proteins (e.g., [89,90]) as well as USPs of other bacteria (e.g., [26,27]) can provide techniques for investigating the interaction with ATP and other molecules by *Helicobacteraceae* universal stress proteins.

The 1045 USP sequences from 1014 genomes of *Helicobacter*, *Wolinella*, and an unclassified strain revealed a remarkable diversity in protein lengths (Table S2 and Figure S1) and functional site patterns (Figure 1). The predominant USP sequence lengths of 137 aa and 138 aa, observed in gastric and non-pylori *Helicobacter* species, respectively, suggest potential functional specialization. This distinction may reflect adaptations to different host environments and stress conditions encountered by various *Helicobacter* species. Additionally, the identification of highly conserved and lower conserved regions in the multiple sequence alignment of representative USP sequences, along with the consensus motif of LLHVS (Figure 2), suggests functional importance and potential evolutionary constraints on these proteins. These conserved regions could serve as targets for future studies on USP function and as potential markers for diagnostic or therapeutic interventions [91,92]. The 15 USP genes encoding protein lengths of less than 137 aa could be further investigated for accuracy of the gene predictions. Nine of these USP genes are encoded in genomes of *H. pylori* strains isolated in Malaysia (UM137R, UM137S, UM163, UM171S, UM229R, UM276R, and UM276S). For strain UM163, the IMG/M system predicted two USP genes that code for 99 aa (2641177340) and 121 aa (2641177784). Strain UM163 is classified as a good biofilm former [93]. Since universal stress proteins can function in biofilm formation [30,31], the list of 31 *H. pylori* strains [93] of the UM strain group that are classified as poor, moderate, and good biofilm formers, could aid investigations on *Helicobacter* USP functions in biofilm formation.

The functional annotations and transcription direction of genes adjacent to USP genes have provided valuable insights into potential functional associations and operon arrangements. Noteworthy is the finding that genes involved in posttranslational modification, protein turnover, and chaperone functions are adjacent to USP genes exclusively in gastric helicobacters. The USPs in these species may function in protein synthesis quality control and stress response mechanisms specific to the gastric environment [94]. Krüger et al. proposed that the import of huge amounts of non-homologous DNA might establish a dilution effect through a reservoir of oxidizable nucleotides, thereby protecting *H. pylori*'s cytoplasmic chromosomal DNA against oxidative stress, in particular during host immune response [95]. Some bacteria USPs contribute to oxidative stress defense [20,21]. The *H. pylori* DNA, proteins and lipids are targets for host oxidative stress [96,97]. The adjacency

of DNA-uptake genes to *Helicobacter* USP genes indicates that further research is needed to determine the roles of *Helicobacter* USPs in oxidative stress defense.

The analysis of *H. pylori* strains from different geographical regions, such as Gambia, South Africa, China, and Malaysia, has revealed strain-specific variations in USP sequences and genomic contexts. These differences may reflect adaptations to diverse host populations and environmental conditions such as exposure to antibiotics. The observed variations in gene adjacency distances between USP and ClpS genes among different *H. pylori* populations highlight the dynamic nature of genome organization in this species and suggest potential functional implications of these arrangements [46]. We have developed datasets and an interactive web resource (Figure 7) to support further research on the effects of USP genomic context and protein sequence features on *H. pylori* populations. Thus, datasets on antibiotic resistance [62] and efficiency of biofilm formation [93] could be integrated into future analysis.

The ClpS is an adaptor that binds N-terminal residues (tyrosine, phenylalanine, tryptophan, and leucine) and delivers attached substrates to the AAA + ClpAP protease for degradation [98,99]. In *H. pylori* genomes, the gene for ClpS is adjacent to the gene for the ClpA chaperone of the ClpAP complex (Figures 6 and S4). These figures show that the two proteins involved in stress response are encoded in the same operon as the USP. The *H. pylori* strains with defective ClpA or ClpP (proteolytic component) of ClpAP were sensitive to some antimicrobials [100]. In *Campylobacter jejuni*, a closely related bacterium to *H. pylori*, ClpP is required for natural competence and DNA uptake [101]. The ClpS of *H. pylori* has unusual amino acid substitutions that may alter the binding specificity [102]. Since the gene for ClpS is not adjacent to the USP gene in non-*H. pylori* gastric helicobacters and in some *H. pylori* strains of West African origin [103], our findings revealed a need for research on the genomic context of the USP gene of *H. pylori* isolates, including implications on gastric cancer prevalence rates. A 2023 publication of a Kenyan eight-year study on the pattern and trends of *Helicobacter pylori* genotypes in gastric cancer concluded that “further assessment of the specific genes encoded by *H. pylori* isolates, in chronically infected persons, can aid in stratifying those at increased risk for development of gastric adenocarcinoma [104]”. Therefore, we recommend research studies on features of USP, ClpS and ClpA genes or proteins of *H. pylori* isolates from different clinical presentations of *H. pylori* infection in populations. In Africa, such studies on the universal stress proteins of *Helicobacteraceae* isolated in Africa would contribute data, information, and evidence for knowledge [105] to address identified research priorities [106,107].

The results of genomic context of USP genes revealed that the unclassified strain CG2_30_36_10, whose genome was from metagenomics studies of groundwater (Freshwater microbial communities from Crystal Geyser, Utah, USA, has USP gene adjacency to sulfate transport similar to *Arcobacter*, *Sulfurimonas*, and *Sulfurospirillum* genera. Within the phylum *Campylobacterota*, strains in the *Arcobacter*, *Sulfurimonas*, *Sulfurospirillum* genera inhabit extreme sulfidic habitats [108,109]. This observation suggests that genomic context analysis can contribute to the taxonomic identification of genomes from metagenomics studies.

This data investigation project relies on the integration of datasets from multiple sources. Consequently, changes to datasets or sources of data can influence the findings. To mitigate this limitation, we have included Supplementary Files and a web resource (Figure 7) as part of this report. In addition, where possible, we confirmed content of our dataset with published findings. For example, our dataset contains the same 97 genomes with the SA (South Africa) strain prefix, which is consistent with the genome count reported in the original publication [59]. In addition, the gene prediction algorithms could influence the prediction of genes in a genome. Thus, to confirm the presence or absence

of genes and proteins, experimental assays may be needed, or multiple sources of gene annotations from microbial web portals, such as IMG/M, BioCyc, and BV-BRC (Bacteria and Viral–Bioinformatics Resource Center) [32,48,50,110].

5. Conclusions

There has been limited knowledge on the functions of the universal stress proteins encoded in the genomes of the *Helicobacteraceae* bacterial family, which consisted of *Helicobacter* and *Wolinella* genera. We have reported a more comprehensive data investigation of *Helicobacteraceae* universal stress proteins. New insights into their functions were obtained through a combination of bioinformatics and visual analytics approaches. Our findings suggest several promising avenues and applications in the field of *Helicobacter* biology and pathogenesis. These include further investigations on the functional characterization of *Helicobacteraceae* universal stress proteins from genomic context and protein sequence perspectives. Such studies would elucidate their specific roles in stress response and adaptation mechanisms. In addition, the potential operonic association of *H. pylori* USP in oxidative stress response and DNA uptake warrants further research. Finally, findings contribute to the knowledge of *Helicobacter* biology and may inform future strategies for managing *H. pylori* infections and associated diseases, including gastric cancer.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/pathogens14030275/s1>, File S1: The 1045 *Helicobacteraceae* universal stress protein sequences in FASTA format collected from IMG/M-ER; File S2: The 183 representative *Helicobacteraceae* sequences; File S3: Multiple sequence alignment file for 183 representative universal stress protein sequences in fasta format; File S4: Guide tree for 183 representative universal stress protein sequences in Newick format; File S5: Multiple sequence alignment view of 183 representative universal stress protein sequences with NCBI Multiple Sequence Alignment Viewer; File S6: iTOL constructed phylogenetic tree of the 183 representative universal stress protein sequences; File S7: Figure of gene transcription pattern of adjacent genes of 23 *Helicobacteraceae* universal stress protein genes; File S8: Two worksheets: (1) Descriptions of 46 data columns (variables) in dataset, and (2) Dataset of values of 46 data columns (variables) for 1045 *Helicobacteraceae* USPs. The following supplementary tables and figures are available in File S9. Table S1. *Helicobacter* genomes without species name encoding universal stress proteins grouped by anatomic niche and host organism. Table S2. Lengths of universal stress proteins encoded by *Helicobacteraceae* genomes. Table S3. Counts of *Helicobacteraceae* universal stress protein sequence lengths according to gene transcription direction binary patterns. Figure S1. Protein sequence lengths of 1030 universal stress proteins encoded in 1002 genomes from species of *Helicobacter* and *Wolinella* as well as an unclassified *Helicobacteraceae* bacterium. Figure S2. Clusters, sequence lengths, and amino acid pattern of functional sites for 45 universal stress proteins in four clusters of 45 strains of the *Helicobacter pylori* GAM strain group (Isolates from The Gambia, West Africa). Figure S3. Predicted protein domains encoded by genes adjacent and in same transcription direction to genes for universal stress proteins in *Helicobacter* and *Wolinella* genomes. Figure S4. The gene transcription direction patterns (010, 011, 110, and 111) for three protein sequence lengths (137 aa, 138 aa, and 273 aa) of selected *Helicobacter* universal stress proteins (USP). Figure S5. Integration of gene neighborhood diagrams of universal stress proteins genes (with red box) in genomes of selected helicobacters with an adjacent gene for ATP-dependent protease. References [32,111] are cited in the supplementary materials.

Author Contributions: Conceptualization, R.D.I., S.S.S., A.U.M., A.L.H., S.A.A., O.O.O. and A.O.A.; validation, R.D.I., S.S.S., A.U.M., A.L.H., S.A.A., O.O.O. and A.O.A.; investigation, R.D.I., S.S.S., A.U.M., A.L.H., S.A.A., O.O.O. and A.O.A.; data curation, R.D.I., S.S.S., A.U.M., A.L.H., S.A.A., O.O.O. and A.O.A.; writing—original draft preparation, R.D.I., S.S.S., A.U.M., O.O.O. and A.O.A.; writing—review and editing, R.D.I., S.S.S., A.U.M., A.L.H., S.A.A., O.O.O. and A.O.A.; visualization,

R.D.I., S.S.S., A.U.M. and A.L.H.; funding acquisition, R.D.I., S.S.S., A.U.M., A.L.H. and O.O.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by grants from United States Department of Education Title III Program (P031B170091); National Institutes of Health (H3Africa Bioinformatics Network U41HG006941); and National Science Foundation (EES-1435186, EES-2106821, DUE-2142465, and EES-2406027).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Genomes of *Helicobacteraceae* can be retrieved from the IMG/M website: <https://img.jgi.doe.gov/cgi-bin/m/main.cgi?section=TaxonList&page=lineageMicrobes&domain=Bacteria&family=Helicobacteraceae> (accessed on 8 March 2025). The dataset on genes annotated with the pfam domain (pfam00582) for universal stress proteins can be retrieved from IMG/M <https://img.jgi.doe.gov> (accessed on 8 March 2025). The web resource with interactive analytics resources is available at Tableau Public by searching for *helicobacter_usp_genes* or at this web link: https://public.tableau.com/app/profile/qeubic/viz/helicobacters_usp_genes/introduction (accessed on 8 March 2025).

Acknowledgments: Transdisciplinary Data Scholars Development Program at Bethune-Cookman University, United States and the Train-the-Trainer Project of the H3Africa Bioinformatics Network at the University of Ibadan, Nigeria. Mentorship of the late Akitoye O. Coker, who led the *Campylobacter* Research Laboratory at College of Medicine, University of Lagos, Nigeria.

Conflicts of Interest: The authors declare no conflicts 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.

Abbreviations

The following abbreviations are used in this manuscript:

AAA+	ATPases associated with various cellular activities
AlpA	Adherence-associated lipoprotein A
AhpC	Alkyl hydroperoxide reductase
ATP	Adenosine Tri-Phosphate
CagA	Cytotoxin-associated gene A
DNA	Deoxyribonucleic acid
HtrA	High temperature requirement A protease
KatA	Catalase
NCBI	National Center for Biotechnology Information
VacA	Vacuolating cytotoxin A
USP	Universal Stress Protein

References

1. Aydin, F.; Saticioglu, I.B.; Ay, H.; Kayman, T.; Karakaya, E.; Abay, S. Description of the two novel species of the genus *Helicobacter*: *Helicobacter anatolicus* sp. nov., and *Helicobacter kayseriensis* sp. nov., isolated from feces of urban wild birds. *Syst. Appl. Microbiol.* **2022**, *45*, 126326. [CrossRef] [PubMed]
2. On, S.L.; Miller, W.G.; Houf, K.; Fox, J.G.; Vandamme, P. Minimal standards for describing new species belonging to the families *Campylobacteraceae* and *Helicobacteraceae*: *Campylobacter*, *Arcobacter*, *Helicobacter* and *Wolinella* spp. *Int. J. Syst. Evol. Microbiol.* **2017**, *67*, 5296. [CrossRef] [PubMed]
3. Varon, C.; Azzi-Martin, L.; Khalid, S.; Seeneevassen, L.; Ménard, A.; Spuul, P. *Helicobacters* and cancer, not only gastric cancer? *Semin. Cancer Biol.* **2022**, *86*, 1138–1154. [CrossRef] [PubMed]
4. Solnick, J.V.; Schauer, D.B. Emergence of diverse *Helicobacter* species in the pathogenesis of gastric and enterohepatic diseases. *Clin. Microbiol. Rev.* **2001**, *14*, 59–97. [CrossRef]

5. Mladenova-Hristova, I.; Grekova, O.; Patel, A. Zoonotic potential of *Helicobacter* spp. *J. Microbiol. Immunol. Infect.* **2017**, *50*, 265–269. [[CrossRef](#)]
6. Okoli, A.S.; Wilkins, M.R.; Raftery, M.J.; Mendz, G.L. Response of *Helicobacter hepaticus* to bovine bile. *J. Proteome Res.* **2010**, *9*, 1374–1384. [[CrossRef](#)]
7. Karkhah, A.; Ebrahimpour, S.; Rostamtabar, M.; Koppolu, V.; Darvish, S.; Vasigala, V.K.R.; Validi, M.; Nouri, H.R. *Helicobacter pylori* evasion strategies of the host innate and adaptive immune responses to survive and develop gastrointestinal diseases. *Microbiol. Res.* **2019**, *218*, 49–57. [[CrossRef](#)]
8. Marcus, E.A.; Sachs, G.; Scott, D.R. Colloidal bismuth subcitrate impedes proton entry into *Helicobacter pylori* and increases the efficacy of growth-dependent antibiotics. *Aliment. Pharmacology Ther.* **2015**, *42*, 922–933. [[CrossRef](#)]
9. 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. [[CrossRef](#)]
10. 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*, 275–286. [[CrossRef](#)]
11. 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*, 41–58. [[CrossRef](#)]
12. 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](#)] [[PubMed](#)]
13. 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. Biol.* **2011**, *5*, 61–74. [[CrossRef](#)] [[PubMed](#)]
14. 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. [[PubMed](#)]
15. Makolo, A.U. Transcription direction patterns of adjacent genes in *Mycobacterium tuberculosis* using GENAVIS. *Comput. Biol. Bioinform.* **2019**, *7*, 1–4. [[CrossRef](#)]
16. Sousa, M.C.; McKay, D.B. Structure of the universal stress protein of *Haemophilus influenzae*. *Structure* **2001**, *9*, 1135–1141. [[CrossRef](#)]
17. Tkaczuk, K.L.; I, A.S.; 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](#)]
18. 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](#)]
19. Schweikhard, E.S.; Kuhlmann, S.I.; Kunte, H.J.; Grammann, K.; Ziegler, C.M. Structure and function of the universal stress protein TeaD and its role in regulating the ectoine transporter TeaABC of *Halomonas elongata* DSM 2581(T). *Biochemistry* **2010**, *49*, 2194–2204. [[CrossRef](#)]
20. Nachin, L.; Nannmark, U.; Nystrom, 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](#)]
21. Seifart Gomes, C.; Izar, B.; Pazan, F.; Mohamed, W.; Mraheil, M.A.; Mukherjee, K.; Billion, A.; Aharonowitz, Y.; Chakraborty, T.; Hain, T. Universal stress proteins are important for oxidative and acid stress resistance and growth of *Listeria monocytogenes* EGD-e in vitro and in vivo. *PLoS ONE* **2011**, *6*, e24965. [[CrossRef](#)]
22. Nystrom, T.; Gustavsson, N. Maintenance energy requirement: What is required for stasis survival of *Escherichia coli*? *Biochim. Biophys. Acta* **1998**, *1365*, 225–231. [[CrossRef](#)] [[PubMed](#)]
23. Nystrom, T.; Neidhardt, F.C. Expression and role of the universal stress protein, UspA, of *Escherichia coli* during growth arrest. *Mol. Microbiol.* **1994**, *11*, 537–544. [[CrossRef](#)] [[PubMed](#)]
24. Gustavsson, N.; Diez, A.; Nystrom, T. The universal stress protein paralogues of *Escherichia coli* are co-ordinately regulated and co-operate in the defence against DNA damage. *Mol. Microbiol.* **2002**, *43*, 107–117. [[CrossRef](#)] [[PubMed](#)]
25. Heermann, R.; Lippert, M.L.; Jung, K. Domain swapping reveals that the N-terminal domain of the sensor kinase KdpD in *Escherichia coli* is important for signaling. *BMC Microbiol.* **2009**, *9*, 133. [[CrossRef](#)]
26. Drumm, J.E.; Mi, K.; Bilder, P.; Sun, M.; Lim, J.; Bielefeldt-Ohmann, H.; Basaraba, R.; So, M.; Zhu, G.; Tufariello, J.M. *Mycobacterium tuberculosis* universal stress protein Rv2623 regulates bacillary growth by ATP-Binding: Requirement for establishing chronic persistent infection. *PLoS Pathog.* **2009**, *5*, e1000460. [[CrossRef](#)]
27. 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](#)]

28. Sharma, C.M.; Hoffmann, S.; Darfeuille, F.; Reignier, J.; Findeiß, S.; Sittka, A.; Chabas, S.; Reiche, K.; Hackermüller, J.; Reinhardt, R. The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature* **2010**, *464*, 250–255. [\[CrossRef\]](#)
29. Singh, S.; Guttula, P.K.; Guruprasad, L. Structure based annotation of *Helicobacter pylori* strain 26695 proteome. *PLoS ONE* **2014**, *9*, e115020. [\[CrossRef\]](#)
30. Luo, D.; Wu, Z.; Bai, Q.; Zhang, Y.; Huang, M.; Huang, Y.; Li, X. Universal stress proteins: From gene to function. *Int. J. Mol. Sci.* **2023**, *24*, 4725. [\[CrossRef\]](#)
31. 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. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Chen, I.-M.A.; Chu, K.; Palaniappan, K.; Ratner, A.; Huang, J.; Huntemann, M.; Hajek, P.; Ritter, S.J.; Webb, C.; Wu, D. The IMG/M data management and analysis system v. 7: Content updates and new features. *Nucleic Acids Res.* **2023**, *51*, D723–D732. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Kehrner, J.; Hauser, H. Visualization and visual analysis of multifaceted scientific data: A survey. *IEEE Trans. Vis. Comput. Graph.* **2012**, *19*, 495–513. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Beard, L.; Aghassibake, N. Tableau (version 2020.3). *J. Med. Libr. Assoc.* **2021**, *109*, 159. [\[CrossRef\]](#)
35. Wang, J.; Chitsaz, F.; Derbyshire, M.K.; Gonzales, N.R.; Gwadz, M.; Lu, S.; Marchler, G.H.; Song, J.S.; Thanki, N.; Yamashita, R.A. The conserved domain database in 2023. *Nucleic Acids Res.* **2023**, *51*, D384–D388. [\[CrossRef\]](#)
36. Li, W.; Godzik, A. CD-HIT: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **2006**, *22*, 1658–1659. [\[CrossRef\]](#)
37. The Galaxy Community. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Res.* **2022**, *50*, W345–W351. [\[CrossRef\]](#)
38. Thorell, K.; Muñoz-Ramírez, Z.Y.; Wang, D.; Sandoval-Motta, S.; Boscolo Agostini, R.; Ghirotto, S.; Torres, R.C.; Falush, D.; Camargo, M.C. The *Helicobacter pylori* Genome Project: Insights into *H. pylori* population structure from analysis of a worldwide collection of complete genomes. *Nat. Commun.* **2023**, *14*, 8184. [\[CrossRef\]](#)
39. Rozewicki, J.; Li, S.; Amada, K.M.; Standley, D.M.; Katoh, K. MAFFT-DASH: Integrated protein sequence and structural alignment. *Nucleic Acids Res.* **2019**, *47*, W5–W10. [\[CrossRef\]](#)
40. Sayers, E.W.; Beck, J.; Bolton, E.E.; Bourexis, D.; Brister, J.R.; Canese, K.; Comeau, D.C.; Funk, K.; Kim, S.; Klimke, W. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2021**, *49*, D10. [\[CrossRef\]](#)
41. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **2021**, *49*, W293–W296. [\[CrossRef\]](#)
42. Huerta-Cepas, J.; Serra, F.; Bork, P. ETE 3: Reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.* **2016**, *33*, 1635–1638. [\[CrossRef\]](#)
43. Kountz, D.J.; Balskus, E.P. Leveraging microbial genomes and genomic context for chemical discovery. *Acc. Chem. Res.* **2021**, *54*, 2788–2797. [\[CrossRef\]](#)
44. Mavromatis, K.; Chu, K.; Ivanova, N.; Hooper, S.D.; Markowitz, V.M.; Kyrpides, N.C. Gene context analysis in the Integrated Microbial Genomes (IMG) data management system. *PLoS ONE* **2009**, *4*, e7979. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Seitzer, P.; Yao, A.I.; Cisneros, A.; Facciotti, M.T. The exploration of novel regulatory relationships drives Haloarchaeal operon-like structural dynamics over short evolutionary distances. *Microorganisms* **2020**, *8*, 1900. [\[CrossRef\]](#)
46. Lim, H.N.; Lee, Y.; Hussein, R. Fundamental relationship between operon organization and gene expression. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 10626–10631. [\[CrossRef\]](#)
47. Li, X.-Q.; Du, D. Gene direction in living organisms. *Sci. Rep.* **2012**, *2*, 982. [\[CrossRef\]](#)
48. Karp, P.D.; Ivanova, N.; Krummenacker, M.; Kyrpides, N.; Latendresse, M.; Midford, P.; Ong, W.K.; Paley, S.; Seshadri, R. A comparison of microbial genome web portals. *Front. Microbiol.* **2019**, *10*, 208. [\[CrossRef\]](#) [\[PubMed\]](#)
49. Galperin, M.Y.; Makarova, K.S.; Wolf, Y.I.; Koonin, E.V. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res.* **2015**, *43*, D261–D269. [\[CrossRef\]](#)
50. Karp, P.D.; Billington, R.; Caspi, R.; Fulcher, C.A.; Latendresse, M.; Kothari, A.; Keseler, I.M.; Krummenacker, M.; Midford, P.E.; Ong, Q. The BioCyc collection of microbial genomes and metabolic pathways. *Brief. Bioinform.* **2019**, *20*, 1085–1093. [\[CrossRef\]](#)
51. Mannion, A.; Shen, Z.; Fox, J.G. Comparative genomics analysis to differentiate metabolic and virulence gene potential in gastric versus enterohepatic *Helicobacter* species. *BMC Genom.* **2018**, *19*, 830. [\[CrossRef\]](#)
52. O'Toole, P.W.; Snelling, W.J.; Canchaya, C.; Forde, B.M.; Hardie, K.R.; Josenhans, C.; Graham, R.L.; McMullan, G.; Parkhill, J.; Belda, E. Comparative genomics and proteomics of *Helicobacter mustelae*, an ulcerogenic and carcinogenic gastric pathogen. *BMC Genom.* **2010**, *11*, 164. [\[CrossRef\]](#)
53. Gilbert, M.J.; Duim, B.; Timmerman, A.J.; Zomer, A.L.; Wagenaar, J.A. Whole genome-based phylogeny of reptile-associated *Helicobacter* indicates independent niche adaptation followed by diversification in a poikilothermic host. *Sci. Rep.* **2017**, *7*, 8387. [\[CrossRef\]](#) [\[PubMed\]](#)

54. Barrett, T.; Clark, K.; Gevorgyan, R.; Gorelenkov, V.; Gribov, E.; Karsch-Mizrachi, I.; Kimelman, M.; Pruitt, K.D.; Resenchuk, S.; Tatusova, T. BioProject and BioSample databases at NCBI: Facilitating capture and organization of metadata. *Nucleic Acids Res.* **2012**, *40*, D57–D63. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Jeon, W.J.; Dong, H.-J.; Shin, J.H.; Kim, I.Y.; Ho, H.; Oh, S.H.; Yoon, Y.M.; Choi, Y.-K.; Suh, J.G.; Nam, K.-H. *Helicobacter apodemus* sp. nov., a new *Helicobacter* species identified from the gastrointestinal tract of striped field mice in Korea. *J. Vet. Sci.* **2015**, *16*, 475–481. [\[CrossRef\]](#) [\[PubMed\]](#)
56. Lopez-Cantillo, M.; Vidal-Veuthey, B.; Mella, A.; de la Haba, R.R.; Collado, L. *Helicobacter ibis* sp. nov., isolated from faecal droppings of black-faced ibis (*Theristicus melanopis*). *Int. J. Syst. Evol. Microbiol.* **2023**, *73*, 005983. [\[CrossRef\]](#)
57. Kersulyte, D.; Bertoli, M.T.; Tamma, S.; Keelan, M.; Munday, R.; Geary, J.; Veldhuyzen van Zanten, S.; Goodman, K.J.; Berg, D.E. Complete genome sequences of two *Helicobacter pylori* strains from a Canadian Arctic Aboriginal community. *Genome Announc.* **2015**, *3*, e00209-15. [\[CrossRef\]](#)
58. Mukherjee, S.; Stamatis, D.; Li, C.T.; Ovchinnikova, G.; Bertsch, J.; Sundaramurthi, J.C.; Kandimalla, M.; Nicolopoulos, P.A.; Favognano, A.; Chen, I.-M.A. Twenty-five years of Genomes OnLine Database (GOLD): Data updates and new features in v. 9. *Nucleic Acids Res.* **2023**, *51*, D957–D963. [\[CrossRef\]](#)
59. Didelot, X.; Nell, S.; Yang, I.; Woltemate, S.; Van der Merwe, S.; Suerbaum, S. Genomic evolution and transmission of *Helicobacter pylori* in two South African families. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 13880–13885. [\[CrossRef\]](#)
60. Shen, Z.; Mannion, A.; Lin, M.; Esmail, M.; Bakthavatchalu, V.; Yang, S.; Ho, C.; Feng, Y.; Smith, B.; Elliott, J. *Helicobacter monodelphidis* sp. nov. and *Helicobacter didelphidarum* sp. nov., isolated from grey short-tailed opossums (*Monodelphis domestica*) with endemic cloacal prolapses. *Int. J. Syst. Evol. Microbiol.* **2020**, *70*, 6032–6043. [\[CrossRef\]](#)
61. Schwarz, S.; Morelli, G.; Kusecek, B.; Manica, A.; Balloux, F.; Owen, R.J.; Graham, D.Y.; Van der Merwe, S.; Achtman, M.; Suerbaum, S. Horizontal versus familial transmission of *Helicobacter pylori*. *PLoS Pathog.* **2008**, *4*, e1000180. [\[CrossRef\]](#)
62. Hu, L.; Zeng, X.; Ai, Q.; Liu, C.; Zhang, X.; Chen, Y.; Liu, L.; Li, G.-Q. Long-read-and short-read-based whole-genome sequencing reveals the antibiotic resistance pattern of *Helicobacter pylori*. *Microbiol. Spectr.* **2023**, *11*, e04522. [\[CrossRef\]](#) [\[PubMed\]](#)
63. Gunaletchumy, S.P.; Teh, X.; Khosravi, Y.; Ramli, N.S.K.; Chua, E.G.; Kavitha, T.; Mason, J.N.; Lee, H.T.; Alias, H.; Zaidan, N.Z. Draft genome sequences of *Helicobacter pylori* isolates from Malaysia, cultured from patients with functional dyspepsia and gastric cancer. *J. Bacteriol.* **2012**, *194*, 5695–5696. [\[CrossRef\]](#)
64. Rehvathy, V.; Tan, M.H.; Gunaletchumy, S.P.; Teh, X.; Wang, S.; Baybayan, P.; Singh, S.; Ashby, M.; Kaakoush, N.O.; Mitchell, H.M.; et al. Multiple genome sequences of *Helicobacter pylori* strains of diverse disease and antibiotic resistance backgrounds from Malaysia. *Genome Announc.* **2013**, *1*, e00687-13. [\[CrossRef\]](#)
65. Blanchard, T.G.; Czinn, S.J.; Correa, P.; Nakazawa, T.; Keelan, M.; Morningstar, L.; Santana-Cruz, I.; Maroo, A.; McCracken, C.; Shefchek, K. Genome sequences of 65 *Helicobacter pylori* strains isolated from asymptomatic individuals and patients with gastric cancer, peptic ulcer disease, or gastritis. *Pathog. Dis.* **2013**, *68*, 39–43. [\[CrossRef\]](#)
66. Gutiérrez-Escobar, A.J.; Velapatiño, B.; Borda, V.; Rabkin, C.S. Identification of new *Helicobacter pylori* subpopulations in native Americans and mestizos from Peru. *Front. Microbiol.* **2020**, *11*, 601839. [\[CrossRef\]](#) [\[PubMed\]](#)
67. Hauke, M.; Metz, F.; Rapp, J.; Faass, L.; Bats, S.H.; Radziej, S.; Link, H.; Eisenreich, W.; Josenhans, C. *Helicobacter pylori* modulates heptose metabolite biosynthesis and heptose-dependent innate immune host cell activation by multiple mechanisms. *Microbiol. Spectr.* **2023**, *11*, e03132-22. [\[CrossRef\]](#) [\[PubMed\]](#)
68. Noszka, M.; Strzałka, A.; Muraszko, J.; Kolenda, R.; Meng, C.; Ludwig, C.; Stingl, K.; Zawilak-Pawlik, A. Profiling of the *Helicobacter pylori* redox switch HP1021 regulon using a multi-omics approach. *Nat. Commun.* **2023**, *14*, 6715. [\[CrossRef\]](#)
69. Olbermann, P.; Josenhans, C.; Moodley, Y.; Uhr, M.; Stamer, C.; Vauterin, M.; Suerbaum, S.; Achtman, M.; Linz, B. A global overview of the genetic and functional diversity in the *Helicobacter pylori* cag pathogenicity island. *PLoS Genet.* **2010**, *6*, e1001069. [\[CrossRef\]](#)
70. Rain, J.-C.; Selig, L.; De Reuse, H.; Battaglia, V.; Reverdy, C.; Simon, S.; Lenzen, G.; Petel, F.; Wojcik, J.; Schächter, V. The protein–protein interaction map of *Helicobacter pylori*. *Nature* **2001**, *409*, 211–215. [\[CrossRef\]](#)
71. Malfertheiner, P.; Selgrad, M.; Wex, T.; Romi, B.; Borgogni, E.; Spensieri, F.; Zedda, L.; Ruggiero, P.; Pancotto, L.; Censini, S. Efficacy, immunogenicity, and safety of a parenteral vaccine against *Helicobacter pylori* in healthy volunteers challenged with a Cag-positive strain: A randomised, placebo-controlled phase 1/2 study. *Lancet Gastroenterol. Hepatol.* **2018**, *3*, 698–707. [\[CrossRef\]](#)
72. Linz, B.; Windsor, H.M.; McGraw, J.J.; Hansen, L.M.; Gajewski, J.P.; Tomsho, L.P.; Hake, C.M.; Solnick, J.V.; Schuster, S.C.; Marshall, B.J. A mutation burst during the acute phase of *Helicobacter pylori* infection in humans and rhesus macaques. *Nat. Commun.* **2014**, *5*, 4165. [\[CrossRef\]](#)
73. Kumar, S.; Karmakar, B.C.; Nagarajan, D.; Mukhopadhyay, A.K.; Morgan, R.D.; Rao, D.N. N4-cytosine DNA methylation regulates transcription and pathogenesis in *Helicobacter pylori*. *Nucleic Acids Res.* **2018**, *46*, 3429–3445. [\[CrossRef\]](#)
74. Clancy, C.D.; Forde, B.M.; Moore, S.A.; O’Toole, P.W. Draft genome sequences of *Helicobacter pylori* strains 17874 and P79. *J. Bacteriol.* **2012**, *194*, 2402. [\[CrossRef\]](#) [\[PubMed\]](#)

75. Li, H.; Yang, T.; Liao, T.; Debowski, A.W.; Nilsson, H.-O.; Fulurija, A.; Haslam, S.M.; Mulloy, B.; Dell, A.; Stubbs, K.A. The redefinition of *Helicobacter pylori* lipopolysaccharide O-antigen and core-oligosaccharide domains. *PLoS Pathog.* **2017**, *13*, e1006280. [\[CrossRef\]](#) [\[PubMed\]](#)
76. Covacci, A.; Censini, S.; Bugnoli, M.; Petracca, R.; Burrone, D.; Macchia, G.; Massone, A.; Papini, E.; Xiang, Z.; Figura, N. Molecular characterization of the 128-kDa immunodominant antigen of *Helicobacter pylori* associated with cytotoxicity and duodenal ulcer. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 5791–5795. [\[CrossRef\]](#)
77. Xiang, Z.; Censini, S.; Bayeli, P.F.; Telford, J.L.; Figura, N.; Rappuoli, R.; Covacci, A. Analysis of expression of CagA and VacA virulence factors in 43 strains of *Helicobacter pylori* reveals that clinical isolates can be divided into two major types and that CagA is not necessary for expression of the vacuolating cytotoxin. *Infect. Immun.* **1995**, *63*, 94–98. [\[CrossRef\]](#) [\[PubMed\]](#)
78. Ashaolu, J.O.; Tsai, Y.-J.; Liu, C.-C.; Ji, D.-D. Prevalence, diversity and public health implications of *Helicobacter* species in pet and stray dogs. *One Health* **2022**, *15*, 100430. [\[CrossRef\]](#)
79. Foret, S.; Seneca, F.; de Jong, D.; Bieller, A.; Hemmrich, G.; Augustin, R.; Hayward, D.C.; Ball, E.E.; Bosch, T.C.; Agata, K.; et al. Phylogenomics reveals an anomalous distribution of USP genes in metazoans. *Mol. Biol. Evol.* **2011**, *28*, 153–161. [\[CrossRef\]](#) [\[PubMed\]](#)
80. Masamba, P.; Kappo, A.P. Parasite survival and disease persistence in cystic fibrosis, schistosomiasis and pathogenic bacterial diseases: A role for universal stress proteins? *Int. J. Mol. Sci.* **2021**, *22*, 10878. [\[CrossRef\]](#)
81. Matarredona, L.; Zafrilla, B.; Rubio-Portillo, E.; Bonete, M.-J.; Esclapez, J. Deepening the knowledge of universal stress proteins in *Haloferax mediterranei*. *Appl. Microbiol. Biotechnol.* **2024**, *108*, 124. [\[CrossRef\]](#)
82. Corbinais, C.; Mathieu, A.; Damke, P.P.; Kortulewski, T.; Busso, D.; Prado-Acosta, M.; Radicella, J.P.; Marsin, S. ComB proteins expression levels determine *Helicobacter pylori* competence capacity. *Sci. Rep.* **2017**, *7*, 41495. [\[CrossRef\]](#) [\[PubMed\]](#)
83. Sen, A.; Zhou, Y.; Imlay, J.A. During oxidative stress the Clp proteins of *Escherichia coli* ensure that iron pools remain sufficient to reactivate oxidized metalloenzymes. *J. Bacteriol.* **2020**, *202*, e00235-20. [\[CrossRef\]](#)
84. Dougan, D.A.; Reid, B.G.; Horwich, A.L.; Bukau, B. ClpS, a substrate modulator of the ClpAP machine. *Mol. Cell* **2002**, *9*, 673–683. [\[CrossRef\]](#) [\[PubMed\]](#)
85. Huang, C.-H.; Chiou, S.-H. Proteomic analysis of upregulated proteins in *Helicobacter pylori* under oxidative stress induced by hydrogen peroxide. *Kaohsiung J. Med. Sci.* **2011**, *27*, 544–553. [\[CrossRef\]](#)
86. Tegtmeyer, N.; Wessler, S.; Backert, S. Role of the cag-pathogenicity island encoded type IV secretion system in *Helicobacter pylori* pathogenesis. *FEBS J.* **2011**, *278*, 1190–1202. [\[CrossRef\]](#)
87. Diez, A.; Gustavsson, N.; Nyström, T. The universal stress protein A of *Escherichia coli* is required for resistance to DNA damaging agents and is regulated by a RecA/FtsK-dependent regulatory pathway. *Mol. Microbiol.* **2000**, *36*, 1494–1503. [\[CrossRef\]](#) [\[PubMed\]](#)
88. Gu, H. Role of flagella in the pathogenesis of *Helicobacter pylori*. *Curr. Microbiol.* **2017**, *74*, 863–869. [\[CrossRef\]](#)
89. Shaik, M.M.; Cendron, L.; Percudani, R.; Zanotti, G. The structure of *Helicobacter pylori* HP0310 reveals an atypical peptidoglycan deacetylase. *PLoS ONE* **2011**, *6*, e19207. [\[CrossRef\]](#)
90. Zarzecka, U.; Matkowska, D.; Backert, S.; Skorko-Glonek, J. Importance of two PDZ domains for the proteolytic and chaperone activities of *Helicobacter pylori* serine protease HtrA. *Cell. Microbiol.* **2021**, *23*, e13299. [\[CrossRef\]](#)
91. Kumar, S.; Patel, G.K.; Ghoshal, U.C. *Helicobacter pylori*-induced inflammation: Possible factors modulating the risk of gastric cancer. *Pathogens* **2021**, *10*, 1099. [\[CrossRef\]](#)
92. Masamba, P.; Adenowo, A.F.; Oyinloye, B.E.; Kappo, A.P. Universal stress proteins as new targets for environmental and therapeutic interventions of Schistosomiasis. *Int. J. Environ. Res. Public Health* **2016**, *13*, 972. [\[CrossRef\]](#)
93. Wong, E.H.J.; Ng, C.G.; Chua, E.G.; Tay, A.C.Y.; Peters, F.; Marshall, B.J.; Ho, B.; Goh, K.L.; Vadivelu, J.; Loke, M.F. Comparative genomics revealed multiple *Helicobacter pylori* genes associated with biofilm formation in vitro. *PLoS ONE* **2016**, *11*, e0166835. [\[CrossRef\]](#)
94. Njenga, R.; Boele, J.; Öztürk, Y.; Koch, H.-G. Coping with stress: How bacteria fine-tune protein synthesis and protein transport. *J. Biol. Chem.* **2023**, *299*, 105163. [\[CrossRef\]](#) [\[PubMed\]](#)
95. Krüger, N.-J.; Knüver, M.-T.; Zawilak-Pawlik, A.; Appel, B.; Stingl, K. Genetic diversity as consequence of a microaerobic and neutrophilic lifestyle. *PLoS Pathog.* **2016**, *12*, e1005626. [\[CrossRef\]](#)
96. Dorer, M.S.; Sessler, T.H.; Salama, N.R. Recombination and DNA repair in *Helicobacter pylori*. *Annu. Rev. Microbiol.* **2011**, *65*, 329–348. [\[CrossRef\]](#) [\[PubMed\]](#)
97. O'Rourke, E.J.; Chevalier, C.; Pinto, A.V.; Thiberge, J.M.; Ielpi, L.; Labigne, A.; Radicella, J.P. Pathogen DNA as target for host-generated oxidative stress: Role for repair of bacterial DNA damage in *Helicobacter pylori* colonization. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 2789–2794. [\[CrossRef\]](#) [\[PubMed\]](#)
98. Roman-Hernandez, G.; Hou, J.Y.; Grant, R.A.; Sauer, R.T.; Baker, T.A. The ClpS adaptor mediates staged delivery of N-end rule substrates to the AAA+ ClpAP protease. *Mol. Cell* **2011**, *43*, 217–228. [\[CrossRef\]](#)
99. Kim, S.; Fei, X.; Sauer, R.T.; Baker, T.A. AAA+ protease-adaptor structures reveal altered conformations and ring specialization. *Nat. Struct. Mol. Biol.* **2022**, *29*, 1068–1079. [\[CrossRef\]](#)

100. Loughlin, M.F.; Arandhara, V.; Okolie, C.; Aldsworth, T.G.; Jenks, P.J. *Helicobacter pylori* mutants defective in the clpP ATP-dependant protease and the chaperone clpA display reduced macrophage and murine survival. *Microb. Pathog.* **2009**, *46*, 53–57. [\[CrossRef\]](#)
101. Vegge, C.S.; Brøndsted, L.; Ligowska-Marzeta, M.; Ingmer, H. Natural transformation of *Campylobacter jejuni* occurs beyond limits of growth. *PLoS ONE* **2012**, *7*, e45467. [\[CrossRef\]](#) [\[PubMed\]](#)
102. Vorn, J. Biophysical Characterization of Unusual ClpS Proteolytic Adapters. Bachelor's Thesis, University of Delaware, Newark, DE, USA, 2020.
103. Bullock, K.K.; Shaffer, C.L.; Brooks, A.W.; Secka, O.; Forsyth, M.H.; McClain, M.S.; Cover, T.L. Genetic signatures for *Helicobacter pylori* strains of West African origin. *PLoS ONE* **2017**, *12*, e0188804. [\[CrossRef\]](#)
104. Njenga, P.; Njau, A.; Moloo, Z.; Revathi, G.; Tshibangu, E.; Yamaoka, Y. Pattern and trends of *Helicobacter pylori* genotypes in gastric cancer: A Kenyan 8-year study. *Front. Med.* **2023**, *10*, 1119513. [\[CrossRef\]](#)
105. Dammann, O. Data, information, evidence, and knowledge: A proposal for health informatics and data science. *Online J. Public Health Inform.* **2018**, *10*, e224. [\[CrossRef\]](#) [\[PubMed\]](#)
106. Smith, S.I.; Ajayi, A.; Jolaiya, T.; Onyekwere, C.; Setshedi, M.; Schulz, C.; Otegbayo, J.A.; Ndip, R.; Dieye, Y.; Alborai, M. *Helicobacter pylori* infection in Africa: Update of the current situation and challenges. *Dig. Dis.* **2022**, *40*, 535–544. [\[CrossRef\]](#)
107. Smith, S.I.; Schulz, C.; Ugiagbe, R.; Ndip, R.; Dieye, Y.; Leja, M.; Onyekwere, C.; Ndububa, D.; Ajayi, A.; Jolaiya, T.F. *Helicobacter pylori* diagnosis and treatment in Africa: The First Lagos Consensus Statement of the African *Helicobacter* and Microbiota Study Group. *Dig. Dis.* **2024**, *42*, 240–256. [\[CrossRef\]](#)
108. Goffredi, S.K.; Panossian, B.; Brzechffa, C.; Field, N.; King, C.; Moggioli, G.; Rouse, G.W.; Martín-Durán, J.M.; Henry, L.M. A dynamic epibiont community associated with the bone-eating polychaete genus *Osedax*. *MBio* **2023**, *14*, e03140-22. [\[CrossRef\]](#)
109. Zhang, Y.; Zhao, Z.; Chen, C.-T.A.; Tang, K.; Su, J.; Jiao, N. Sulfur metabolizing microbes dominate microbial communities in andesite-hosted shallow-sea hydrothermal systems. *PLoS ONE* **2012**, *7*, e44593. [\[CrossRef\]](#) [\[PubMed\]](#)
110. Olson, R.D.; Assaf, R.; Brettin, T.; Conrad, N.; Cucinell, C.; Davis, J.J.; Dempsey, D.M.; Dickerman, A.; Dietrich, E.M.; Kenyon, R.W. Introducing the bacterial and viral bioinformatics resource center (BV-BRC): A resource combining PATRIC, IRD and ViPR. *Nucleic Acids Res.* **2023**, *51*, D678–D689. [\[CrossRef\]](#)
111. Paysan-Lafosse, T.; Blum, M.; Chuguransky, S.; Grego, T.; Pinto, B.L.; Salazar, G.A.; Bileschi, M.L.; Bork, P.; Bridge, A.; Colwell, L. InterPro in 2022. *Nucleic Acids Res.* **2023**, *51*, D418–D427. [\[CrossRef\]](#)

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