



Article

# Identification of a Putative CodY Regulon in the Gram-Negative Phylum Synergistetes

Jianing Geng<sup>1,†</sup> , Sainan Luo<sup>1,2,†</sup> , Hui-Ru Shieh<sup>3</sup>, Hsing-Yi Wang<sup>3</sup>, Songnian Hu<sup>1,2,\*</sup>   
and Yi-Ywan M. Chen<sup>3,4,5,\*</sup>

<sup>1</sup> State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China; gengjn@im.ac.cn (J.G.); luosainan9606@163.com (S.L.)

<sup>2</sup> University of Chinese Academy of Sciences, Beijing 100049, China

<sup>3</sup> Department of Microbiology and Immunology, College of Medicine, Chang Gung University, Taoyuan 333, Taiwan; cowcow119@hotmail.com (H.-R.S.); kenny200929@gmail.com (H.-Y.W.)

<sup>4</sup> Graduate Institute of Biomedical Sciences, College of Medicine, Chang Gung University, Taoyuan 333, Taiwan

<sup>5</sup> Molecular Infectious Disease Research Center, Chang Gung Memorial Hospital, Taoyuan 333, Taiwan

\* Correspondence: husn@im.ac.cn (S.H.); mchen@mail.cgu.edu.tw (Y.-Y.M.C.)

† These authors have contributed equally to this work.

**Abstract:** CodY is a dominant regulator in low G + C, Gram-positive Firmicutes that governs the regulation of various metabolic pathways and cellular processes. By using various bioinformatics analyses and DNA affinity precipitation assay (DAPA), this study confirmed the presence of CodY orthologues and corresponding regulons in Gram-negative Synergistetes. A novel palindromic sequence consisting of AT-rich arms separated by a spacer region of variable length and sequence was identified in the promoters of the putative *codY*-containing operons in Synergistetes. The consensus sequence from genera *Synergistes* and *Cloacibacillus* (5'-AATTTTCTTAAAATTCSCCTTGATATTTACAATTTT) contained three AT-rich regions, resulting in two palindromic sequences; one of which is identical to Firmicutes CodY box (5'-AAATTTTCWGAAAATT). The function of the consensus sequence was tested by using a recombinant CodY protein (His-CodY<sub>DSM</sub>) of *Cloacibacillus evryensis* DSM19522 in DAPA. Mutations in the central AT-rich sequence reduced significantly the binding of His-CodY<sub>DSM</sub>, whereas mutations in the 5' or 3' end AT-rich sequence slightly reduced the binding, indicating that CodY<sub>DSM</sub> could recognize both palindromic sequences. The proposed binding sequences were found in the promoters of multiple genes involved in amino acids biosynthesis, metabolism, regulation, and stress responses in Synergistetes. Thus, a CodY-like protein from Synergistetes may function similarly to Firmicutes CodY.

**Keywords:** Synergistetes; CodY orthologues; binding consensus of Synergistetes CodY; CodY regulon of Synergistetes



**Citation:** Geng, J.; Luo, S.; Shieh, H.-R.; Wang, H.-Y.; Hu, S.; Chen, Y.-Y.M. Identification of a Putative CodY Regulon in the Gram-Negative Phylum Synergistetes. *Int. J. Mol. Sci.* **2022**, *23*, 7911. <https://doi.org/10.3390/ijms23147911>

Academic Editor: Naglis Malys

Received: 16 June 2022

Accepted: 6 July 2022

Published: 18 July 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 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

Synergistetes was formally recognized as a novel bacterial phylum based on the 16S rRNA gene phylogeny tree of the bacterial domain in 2009 [1]. The members of the phylum Synergistetes are characterized as Gram-negative, anaerobic bacteria with various cell shapes [1,2]. Microbes of the phylum Synergistetes can utilize amino acids as an energy source and degrade toxic environmental compounds, sludges, and glucose for the production of renewable energy and biogas [1,3,4]. For example, the representative species *Synergistes jonesii*, isolated from goat rumen, can degrade 3-hydroxy-4(1H)-pyridone (3,4-DHP) for growth [5]. Currently, 16 genera are classified in the phylum Synergistetes: *Acetomicrobium*, *Aminiphilus*, *Aminivibrio*, *Aminobacterium*, *Aminomonas*, *Cloacibacillus*, *Dethiosulfobivrio*, *Fretibacterium*, *Jonquetella*, *Lactivibrio*, *Pacaella*, *Pyramidobacter*, *Rarimicrobium*, *Synergistes*, *Thermanaerovibrio* and *Thermovirga* [2,6–10]. Microbes of the phylum Synergistetes reside in

extended anaerobic environments, including animal and insect gastrointestinal tracts, animal wastes, wastewater treatment systems, soils, and oil wells. Moreover, several species of Synergistetes are present in the oral microbiome and are thought to be associated with periodontal diseases [11,12].

The genome sequences of 35 cultivable strains of Synergistetes, 286,997 pieces of metagenomes from the human gut microbiome, and 2870 pieces of metagenomes from various environmental microbiomes, including the metagenomes of samples from activated sludge, anaerobic digesters, and groundwater are currently available in the NCBI Entrez genomes database [13–16]. The availability of these sequences provides opportunities for the identification of global regulators that are present in bacterial groups at different taxonomic levels using bioinformatic analyses. Genomic analyses indicate that the genomes of Synergistetes are tightly arranged, with an average of approximately 1000 genes per megabase. While a few species are asaccharolytic, all members of the phylum Synergistetes are able to ferment amino acids [2]. The average percentage of genes encoding proteins associated with amino acid uptake and metabolism in a Synergistetes genome is  $11.8 \pm 0.8\%$ , representing the highest proportion among all bacterial phyla to date [17].

CodY was first discovered in *Bacillus subtilis* in 1995 as a repressor of the dipeptide permease operon (*dppABCDE*) [18]. Later studies confirmed that CodY is a master regulator for various metabolic genes, and that regulation by CodY facilitates the adaptation of bacteria to poor nutritional conditions [19]. Subsequent studies confirmed that CodY is widely distributed and highly conserved among low G + C, Gram-positive bacteria, including pathogens such as *Streptococcus pneumoniae* and *Listeria monocytogenes* [20–22], in which CodY also regulates the expression of multiple virulence genes [23]. CodY is composed of an N-terminal GAF (cGMP-stimulated phosphodiesterase, adenylate cyclase and a bacterial transcription regulator FhlA) domain and a C-terminal winged helix-turn-helix (wHTH) DNA-binding domain. The GAF domain responds to stimuli and binds with branched-chain amino acids (BCAA, e.g., isoleucine, leucine, and valine) or GTP to activate the DNA-binding activity of CodY [24,25]. Upon activation, the wHTH domain of CodY binds to a palindromic DNA sequence (5'-AATTTTCWGAAAATT) to repress the expression of the downstream gene [20,26,27]. Although GAF and wHTH domains have been described in many proteins, CodY is the only known protein that harbors both a GAF domain and a wHTH domain, and, thus, it is possible to search for putative CodY in bacteria of all phyla using this characteristic, including phyla that are evolutionarily distant from Firmicutes.

In this study, a GAF-wHTH containing, CodY-like protein and its binding consensus sequence were identified in species of the phylum Synergistetes. Additionally, the interaction between the CodY-like protein and the predicted binding sequence was demonstrated using a DNA-affinity precipitation assay (DAPA), indicating that microbes of the Synergistetes phylum likely harbor a CodY-like global regulatory protein for regulation of multiple cellular activities.

## 2. Results

### 2.1. Identification and Characterization of CodY-like Proteins in the Phylum Synergistetes

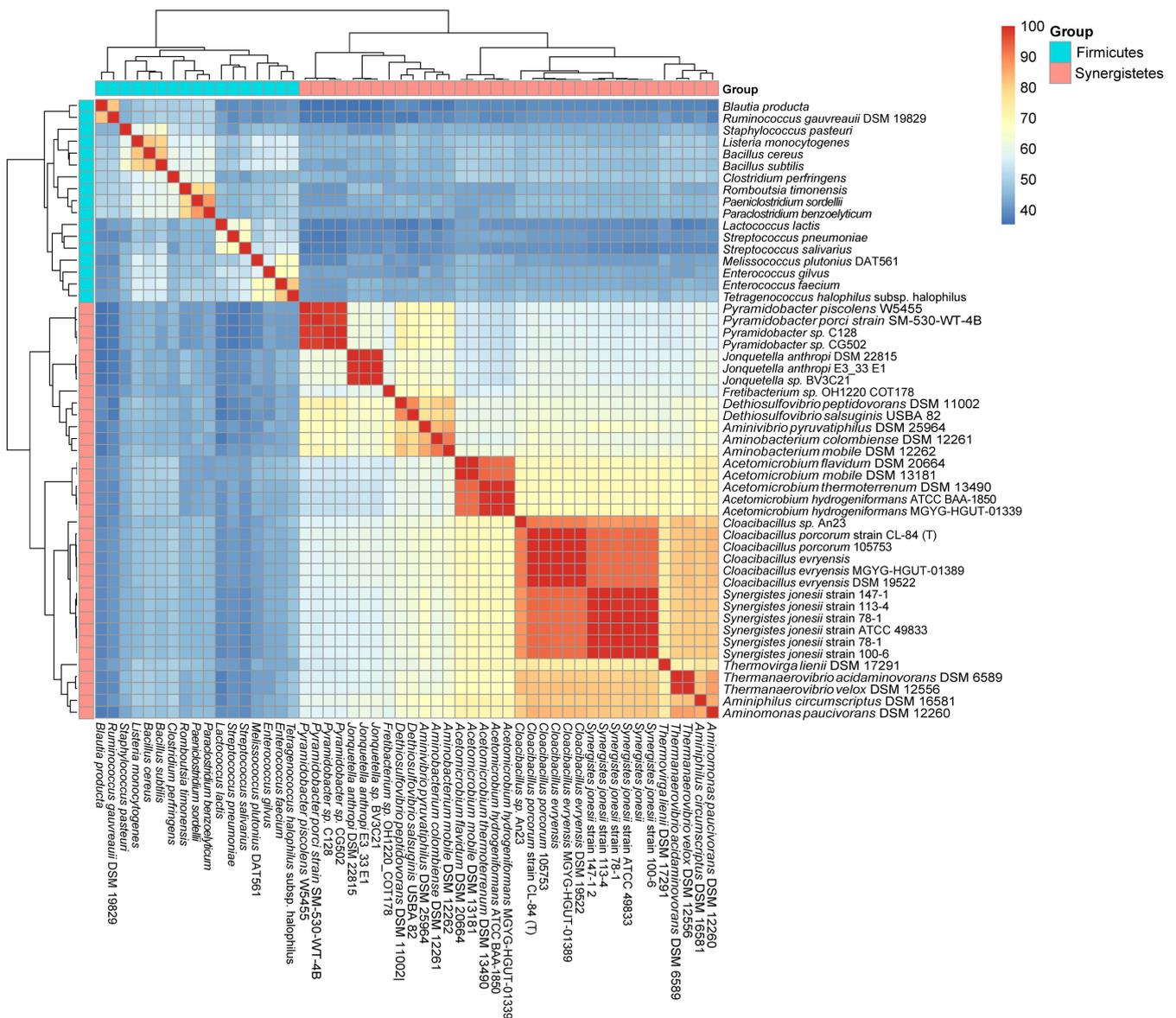
As CodY is the only known protein composed of a GAF and a wHTH domain, profile hidden Markov models (HMMs) were used to investigate whether a CodY-like protein, i.e., proteins with the same domain composition, existed outside of the phylum Firmicutes. An interrogation of bacterial genome sequences in the NCBI Reference Sequence Database (RefSeq), ecological environment metagenomes in the NCBI Entrez genomes database, and human gut microbiome datasets in the European Bioinformatics Institute database (EBI), unexpectedly revealed 107 loci encoding proteins containing both domains in species of the phylum Synergistetes. Among these, 35 loci were present in the genomes of 20 cultivable species, across 12 genera (Table 1), and 23 loci were found in the human gut microbiome datasets across four genera (Table S1). The remaining 49 loci were present in different contigs obtained from ecological samples of extreme environments, such as wastewater

treatment lagoons, anaerobic digester sludge, activated sludge enrichment, groundwater enrichment, hot springs enrichment, sulfur mats, and the deep marine trench. Thus, a CodY-like protein is widely distributed in the phylum Synergistetes, including species occupying different ecological niches.

**Table 1.** The distribution of the putative CodY-like protein among the cultivable species of the phylum Synergistetes.

Locus_Tag	RefSeq	Assembly	GC (%)	Species
BUQ78_RS04225	NZ_FSQZ01000001	GCF_900129645	48.29	<i>Acetomicrobium flavidum</i> DSM 20664
HMPREF1705_RS06035	NZ_ACJX03000001	GCF_000160455	46.53	<i>Acetomicrobium hydrogeniformans</i>
FY315_RS06110	NZ_CABKOO010000001	GCF_902374045	46.53	<i>Acetomicrobium hydrogeniformans</i> MGYG-HGUT-01339
ANAMO_RS01275	NC_018024	GCF_000266925	47.94	<i>Acetomicrobium mobile</i> DSM 13181
BLU12_RS09355	NZ_FNPD010000014	GCF_900107215	48.46	<i>Acetomicrobium thermoterrenum</i> DSM 13490
K349_RS0111180	NZ_JAFY010000007	GCF_000526375	59.16	<i>Aminiphilus circumscriptus</i> DSM 16581
C8D99_RS00920	NZ_SORI010000001	GCF_000165795	67.58	<i>Aminivibrio pyruvativiphilus</i> DSM 25964
AMICO_RS07740	NC_014011	GCF_000025885	45.31	<i>Aminobacterium colombiense</i> DSM 12261
K360_RS0102980	NZ_JAFZ010000001	GCF_000526395	43.88	<i>Aminobacterium mobile</i> DSM 12262
APAU_RS04160	NZ_CM001022	GCF_000165795	67.58	<i>Aminomonas paucivorans</i> DSM 12260
FXY42_RS11030	NZ_CABKQM010000008	GCF_902374565	56.26	<i>Cloacibacillus evryensis</i> MGYG-HGUT-01389
CLOEV_RS08630	NZ_KK073872	GCF_000585335	55.95	<i>Cloacibacillus evryensis</i> DSM 19522
HMPREF1006_RS10935	NZ_JH414697	GCF_000238615	56.26	<i>Cloacibacillus evryensis</i>
BED41_RS09070	NZ_CP016757	GCF_001701045	54.84	<i>Cloacibacillus porcorum</i> strain CL-84 (T)
HF883_RS06600	NZ_JABAGT010000007	GCF_012844265	56.42	<i>Cloacibacillus porcorum</i> 105753
B5F39_RS01840	NZ_NFJQ010000002	GCF_002159945	55.63	<i>Cloacibacillus</i> sp. An23
DPEP_RS01550	NZ_ABTR020000001	GCF_000172975	54.41	<i>Dethiosulfovibrio peptidovorans</i> DSM 11002
B9Y55_RS09045	NZ_FXBB010000024	GCF_900177735	55.24	<i>Dethiosulfovibrio salsuginis</i> USBA 82
EII26_RS10880	NZ_RQYL010000028	GCF_003860125	62.36	<i>Fretibacterium</i> sp. OH1220_COT-178
JONANDRAFT_RS04180	NZ_CM001376	GCF_000237805	59.50	<i>Jonquetella anthropic</i> DSM 22815
GCWU000246_RS04765	NZ_GG697147	GCF_000161995	59.09	<i>Jonquetella anthropic</i> E3_33 E1
HMPREF1249_RS06160	NZ_AWWC010000024	GCF_000468895	59.31	<i>Jonquetella</i> sp. BV3C21
HMPREF7215_RS03875	NZ_ADFP010000047	GCF_000177335	58.15	<i>Pyramidobacter piscicola</i> W5455
B0D78_RS09560	NZ_MUHX010000026	GCF_002007215	57.20	<i>Pyramidobacter</i> sp. C12-8
D7D26_RS08945	NZ_RAWT010000042	GCF_003612005	58.62	<i>Pyramidobacter</i> sp. CG50-2
FYJ74_RS00750	NZ_VUNH010000001	GCF_009695745	58.90	<i>Pyramidobacter porci</i> strain SM-530-WT-4B
EH55_RS06015	NZ_JMKI010000026	GCF_000712295	55.86	<i>Synergistes jonesii</i> strain 78-1
JS79_RS06185	NZ_JPZR010000022	GCF_001757565	55.84	<i>Synergistes jonesii</i> strain 147-1
JS78_RS05700	NZ_JPZS010000024	GCF_001757415	55.84	<i>Synergistes jonesii</i> strain 113-4
JS77_RS05735	NZ_JPZQ010000030	GCF_001757495	55.85	<i>Synergistes jonesii</i> strain 100-6
JS73_RS05710	NZ_JQEL010000027	GCF_001757485	55.84	<i>Synergistes jonesii</i> strain ATCC 49833
JS72_RS09655	NZ_JQEK010000046	GCF_001757465	55.83	<i>Synergistes jonesii</i>
TACI_RS06905	NC_013522	GCF_000024905	63.79	<i>Thermanaerovibrio acidaminovorans</i> DSM 6589
THEVEDRAFT_RS02110	NZ_CM001377	GCF_000237825	58.78	<i>Thermanaerovibrio velox</i> DSM 12556
TLIE_RS02745	NC_016148	GCF_000233775	47.10	<i>Thermovirga lienii</i> DSM 17291

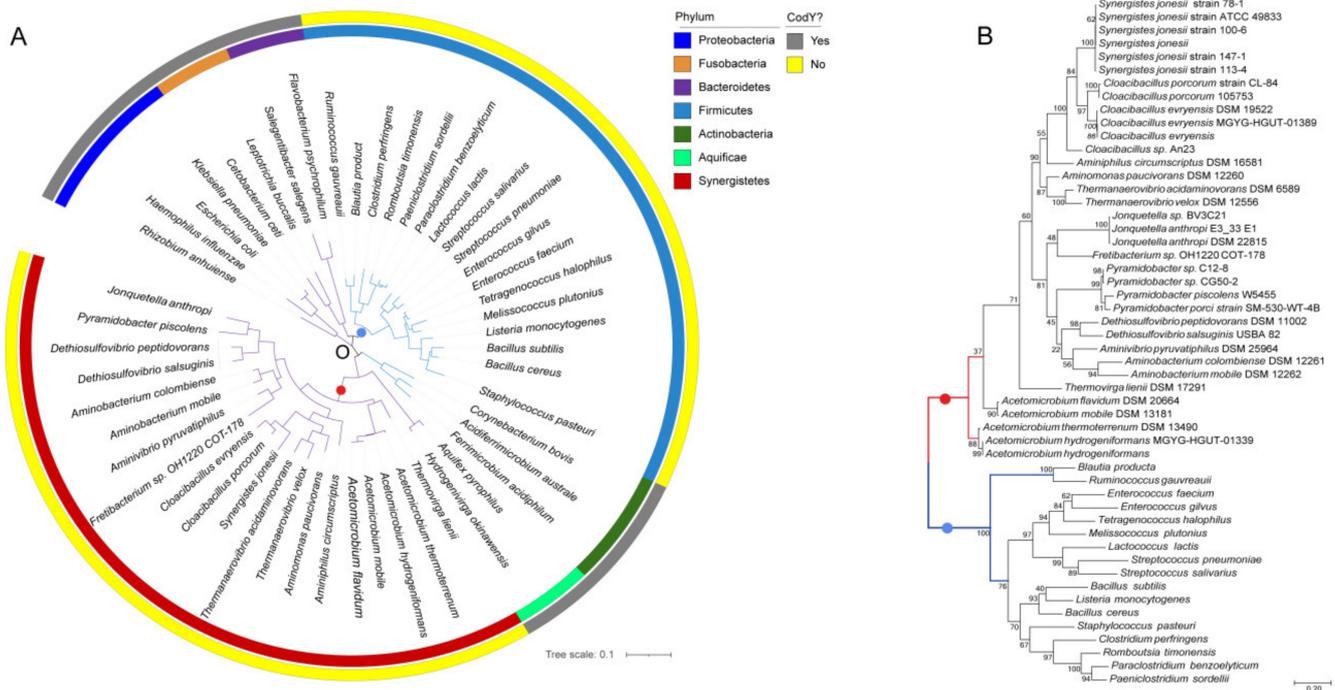
To analyze the evolutionary relationship of the newly identified CodY-like protein and CodY of Firmicutes, we first analyzed the phylogenetic relationship of all CodY proteins identified in Firmicutes (data not shown), and randomly selected one locus from each evolution branch, resulting in a total of 17 loci. The CodY-like proteins of the 35 cultivable isolates of the phylum Synergistetes were then compared with the CodY proteins of the 17 representative species of the phylum Firmicutes (Table S2) at the deduced amino acid sequence level (Figure 1). Protein identity cluster analysis revealed two main groups: one containing the CodY proteins of Firmicutes, and the other containing the CodY-like proteins of Synergistetes. Thus, CodY and the CodY-like protein of species in the same phylum are more homologous to each other and the extent of similarity decreased across phyla. The putative CodY protein of the phylum Synergistetes displays an average of 43.32% identity (35.32% to 50.20%) with the CodY protein of Firmicutes, while an average of 70.23% and 55.48% identity are observed between species of the phyla Synergistetes and Firmicutes, respectively. The high identity between species of Synergistetes (70.23%), compared to that of Firmicutes (55.48%), is likely due to the relatively low number of species in Synergistetes.



**Figure 1.** A heat map showing the identity between the CodY/CodY-like proteins from species of Firmicutes and Synergistetes. The score is the % identity of the deduced amino acid sequence of *codY* and its homologs.

## 2.2. The Origin of the CodY-like Protein of the Phylum Synergistetes

To examine whether the phylogenetic relationship of CodY and the CodY-like protein follow the same evolutionary history of the diversification of bacterial species, we constructed a 16S rRNA gene-based phylogenetic tree with 51 members of 7 phyla, including first, 5 phyla that do not possess CodY (Table S3): Proteobacteria, Fusobacteria, Bacteroidetes, Actinobacteria and Aquificae; second, the CodY-possessing Firmicutes; and third, Synergistetes. As shown in Figure 2A, species of the phyla Synergistetes and Firmicutes are located in two independent clades. An identical evolutionary relationship was found with regards to the CodY and CodY-like protein (Figure 2B), where the two proteins are also located in two independent branches. Furthermore, in both phylogenetic trees, three major groups were observed under the Synergistetes clade, indicating that the evolution of the CodY-like protein follows the evolution of the bacterial species.



**Figure 2.** The phylogenetic relationship of species and the CodY/CodY-like proteins of Firmicutes and Synergistetes. **(A)** Phylogenetic tree of the 16S rRNA of representative species from various bacterial phyla. The presence (grey) and absence (yellow) of a *codY* gene in the chromosome of a species is shown on the outer circle. Species of different phyla are color-coded and shown in the inner circle. The phylum information is shown on the right. An ML distance tree based on the 16S rRNA sequences of 51 species from different phyla is shown inside the circle. Gram-negative bacteria are indicated by purple branches and Gram-positive bacteria are indicated by blue branches. **(B)** An ML distance tree based on 52 CodY homologous protein sequences of different species in the phyla Firmicutes and Synergistetes. The phylum Synergistetes is indicated by the red branch and the phylum Firmicutes is indicated by the blue branch. The numbers on the branches indicate the percentage of calculations performed by bootstrap.

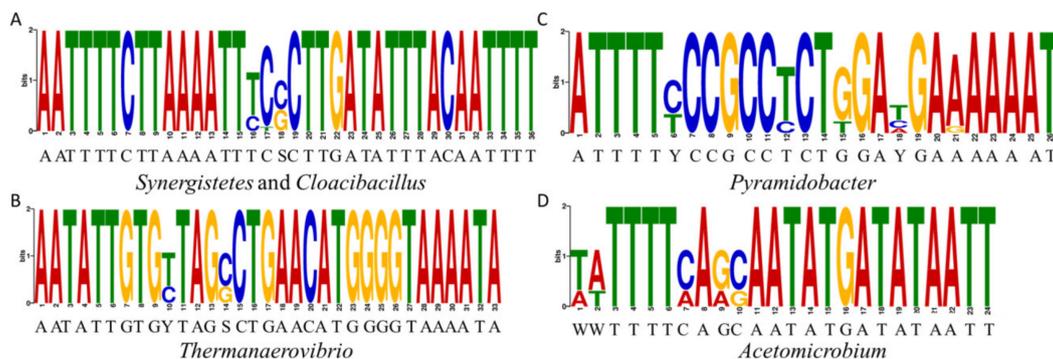
### 2.3. Analysis of the GAF-wHTH Domain

Although only approximately 40% identity was detected between CodY and the CodY-like protein, it is interesting to note that the amino acid sequence of the wHTH domain region of these proteins is relatively conserved across phyla, with an average 73.40% identity (Table S4). All conserved amino acids within the key DNA-binding region (amino acids 203–222) of CodY [28] are also present in the CodY-like proteins, except for Ile212, which is a Val residue in the CodY-like proteins. The conservation of this region suggests that the CodY of Firmicutes and the CodY-like protein of Synergistetes could recognize a similar DNA sequence. On the other hand, the GAF domains of CodY and the CodY-like protein only share an average of 33.41% identity. Nevertheless, the conserved Arg8, Gln15, Glu144 and Thr148 that are essential for the dimerization of CodY in *B. subtilis* [29] are all present in the corresponding position in the CodY-like protein of Synergistetes, suggesting that the CodY-like protein also acts as a dimer. Although the precise binding motif of BCAA to the GAF domain of the Firmicutes CodY is not well-characterized, studies have shown that amino acid residues Phe40, Arg61, Met62, Met65, Phe71, Pro72, Tyr75, Phe98 and Pro99 are essential for BCAA binding in *B. subtilis* [25,30]. However, the amino acid residues found in the corresponding positions of the CodY-like proteins were dissimilar to this set of essential residues. Furthermore, the key residues Phe40 and Phe71 are absent in all CodY-like proteins, whereas Phe98 is present only in the CodY-like proteins found in the genera *Jonquetella*, *Pyramidobacter*, *Dethiosulfovibrio*, *Aminobacterium*, *Aminivibrio* and *Fretibacterium*.

Thus, the effectors for activation of the DNA-binding activity of the CodY-like protein of Synergistetes may differ from that of the CodY from Firmicutes.

#### 2.4. The Binding Consensus of the CodY-like Protein

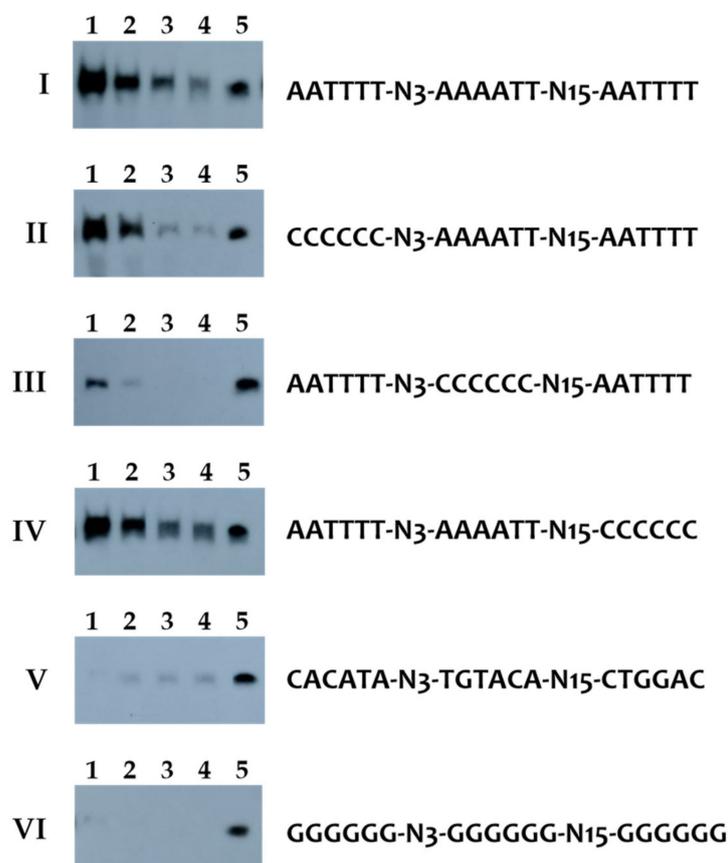
The strategy to define the binding consensus of the CodY-like protein was made based on two observations. Firstly, the similarity in the wHTH domain between CodY and the CodY-like protein would suggest that these two proteins recognize similar sequence. Secondly, the expression of *codY* is autoregulated in microbes of Firmicutes. Specifically, the CodY box is present in the promoters of *codY* of species in the genera of *Lactococcus*, *Streptococcus*, *Enterococcus*, *Staphylococcus* and *Clostridium* [20]. Furthermore, the autoregulation by CodY has been demonstrated experimentally in *Lactococcus lactis* [26], *S. pneumoniae* [21], and *Streptococcus pyogenes* [31]. Thus, it is hypothesized that the expression of the putative *codY* is autoregulated in Synergistetes and it is feasible to identify the binding consensus of the CodY-like protein in the promoters of the putative *codY*. Based on the distance between the putative *codY* and its 5' flanking loci in the genomes of Synergistetes and results from operon analysis by operon-mapper [32], the putative *codY* is likely to be a part of an operon, with *xerC* being the first gene of the operon. Thus, an unbiased search for the canonical CodY box was performed in the promoter regions of the putative *codY* operon of 35 cultivable strains in the phylum Synergistetes. A sequence of 5'-AATTTTCTTAAATT, which is identical to the CodY box of Firmicutes, was found in the promoter regions of 12 isolates of the genera *Synergistes* and *Cloacibacillus*, but not in other genera. To verify whether additional consensus sequences could be recognized by the CodY-like protein, we extracted the 5'-flanking regions of *xerC* from genomes of cultivable isolates, human gut metagenome-assembled genomes, and ecological metagenome assembly sequences in the phylum Synergistetes, and used the MEME motif-searching algorithm [33] to search for conserved palindromic sequences in each genus. As shown in Table S5 and Figure 3, four consensus sequences were discovered in the genera *Synergistes/Cloacibacillus*, *Thermanaerovibrio*, *Pyramidobacter*, and *Acetomicrobium*, respectively. No consensus motifs were discovered in other genera because less than three genomes were available for a specific genus. The consensus sequence identified in the genera *Synergistes* and *Cloacibacillus*, 5'-AATTTTCTTAAATTTCSCCTTGATATTTACAATTTT, contains a Firmicutes CodY box (underlined) and an AAAATT/AATTTT inverted repeat (double underlined). Upon close examination of the consensus found in the genera *Thermanaerovibrio* (5'-AATATT-N<sub>21</sub>-AAAATA), *Pyramidobacter* (5'-ATTTTY-N<sub>14</sub>-AAAAT), and *Acetomicrobium* (5'-WWTTTT-N<sub>12</sub>-ATAATT), it was noted that all of the sequences contained a pair of AT-rich palindromic arms separated by a spacer region of variable length and sequence, suggesting that the binding sequence of the CodY-like protein of the phylum Synergistetes is a palindromic sequence consisting of two AT-rich arms and a variable spacer region.



**Figure 3.** Motif logos representing the binding sites of the CodY-like proteins for five genera in the phylum Synergistetes. The logos were generated by the MEME function of the MEME suite. All 40 core binding sequences of the CodY-like protein listed in Table S5 were used to analyze the common motifs in *Synergistes* and *Cloacibacillus* (A), *Thermanaerovibrio* (B), *Pyramidobacter* (C) and *Acetomicrobium* (D).

### 2.5. The CodY-like Protein of *Cloacibacillus Evryensis* DSM19522 Could Interact with Both the Canonical CodY Box and an AAAATT-N<sub>15</sub>-AATTTT Palindromic Sequence

As the consensus sequence identified in the genera *Synergistes* and *Cloacibacillus* contains both a Firmicutes CodY box and an extended palindromic sequence, the binding specificity of the CodY-like protein of *C. evryensis* DSM19522 (CodY<sub>DSM</sub>) was analyzed using recombinant histidine-tagged CodY<sub>DSM</sub> (His-CodY<sub>DSM</sub>) in DAPA experiments, as detailed in Materials and Methods. His-CodY<sub>DSM</sub> was readily induced and purified from the recombinant *E. coli* strain (Figure S1), but the solubility of this protein was rather low (0.2–0.4 µg mL<sup>-1</sup>). The predicted binding sequence at the 5' flanking region of *xerC* (CLOEV\_RS08635) was extracted from the genome of *C. evryensis* DSM19522 (NZ\_KK073872.1) and used to design probes for the DAPA experiment (Table S6). The DAPA results indicated that His-CodY<sub>DSM</sub> could interact specifically with the full-length probe in a dose-dependent manner (Figure 4). When the three AT-rich sequence were substituted with GGGGGG or six random bases, the signals were abolished, indicating that AT-rich palindromic sequences are critical for the binding of the CodY-like protein. While substitutions of the 5' end AATTTT with CCCCCC reduced marginally the binding of His-CodY<sub>DSM</sub>, mutations in the 3' AATTTT did not affect the binding of His-CodY<sub>DSM</sub> significantly, indicating that His-CodY<sub>DSM</sub> could interact with both the canonical CodY box and an AAAATT-N<sub>15</sub>-AATTTT inverted repeat sequence, and possibly with a slightly higher affinity to the canonical CodY box. On the other hand, mutations of the center AAAATT significantly reduced the binding, confirming that the CodY-like protein could interact with both targets. Thus, the obtained consensus sequence contains two CodY<sub>DSM</sub> binding sites; one of which is identical to the CodY box of Firmicutes, and the other is AATTTT-N<sub>15</sub>-AAAATT.



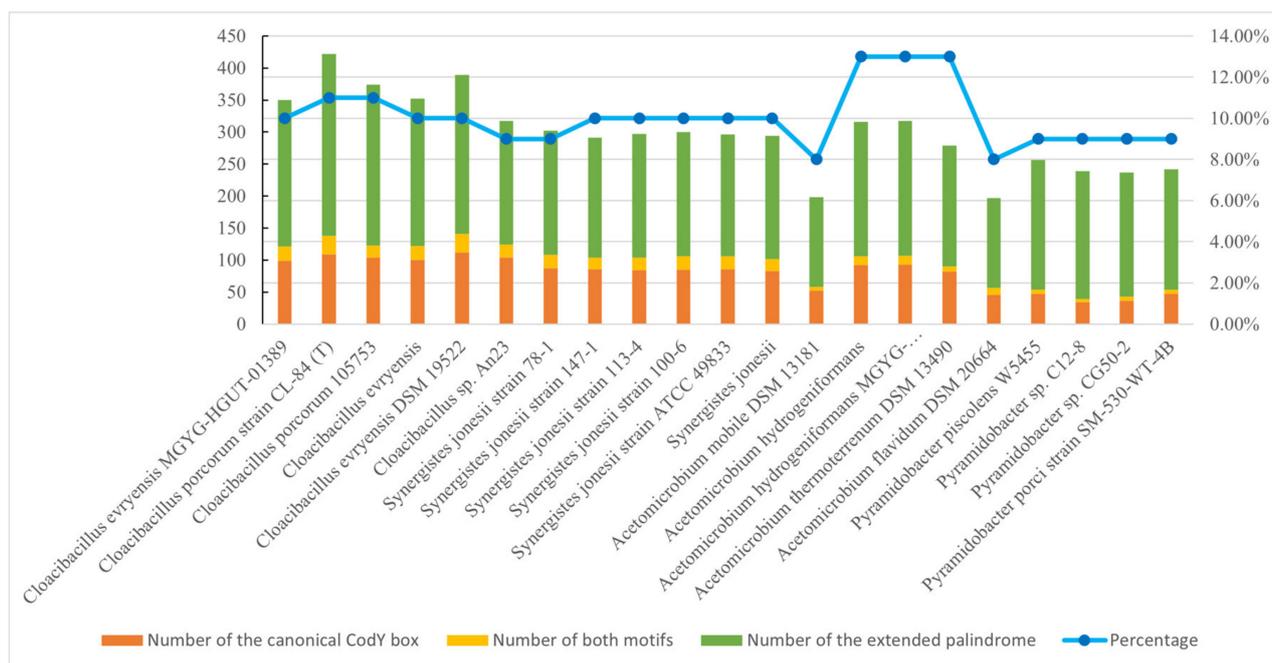
**Figure 4.** DAPA experiments showing the binding specificity of the CodY-like protein of *C. evryensis* DSM19522. Lanes 1 to 4, 2.5 µM, 0.5 µM, 0.1 µM, and 0.02 µM of the purified His-CodY<sub>DSM</sub> were used in the binding reactions, respectively. Lane 5, 1.25% of the probe used in the reaction is shown

to ensure a comparable number of labeled probes was used in each of the reactions. (I), the wild-type probe containing the predicted binding sequence 5' to *xerC* of *C. erythraea* DSM1952 was used. (II), the 5' AT-rich sequence of the probe is mutated. (III), the central AT-rich sequence of the probe is mutated. (IV), the 3' AT-rich sequence of the probe is mutated. (V) and (VI), all three AT-rich sequences were mutated. The key features of the probes used in each reaction are listed right to the images. For clarity, only the sense strand of the probe is shown.

## 2.6. Identification and Functional Prediction of the CodY-like Regulon in the Phylum Synergistetes

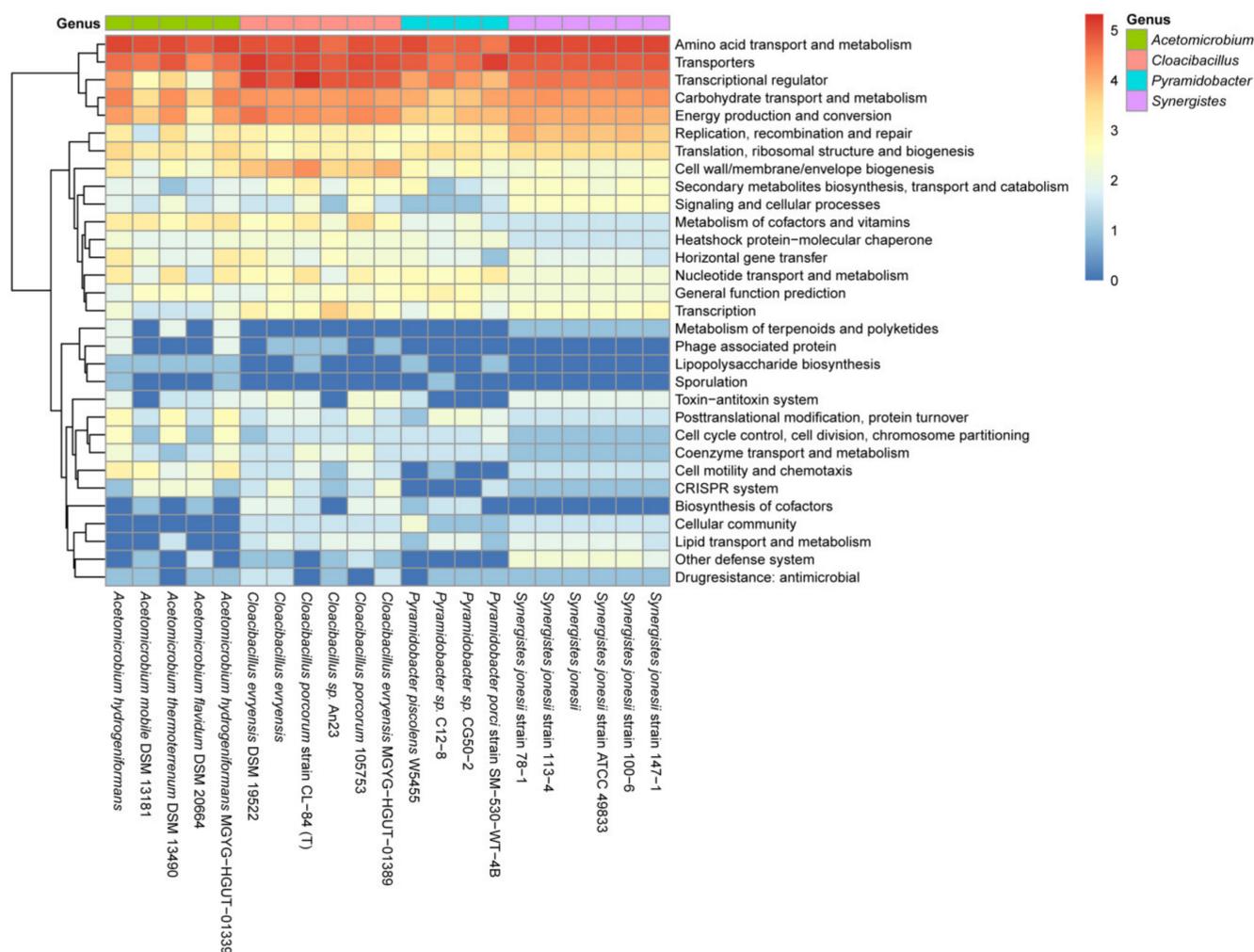
As the first step to investigate whether the CodY-like protein also acts as a global regulator in Synergistetes, a search for the proposed binding sequence of the CodY-like protein was performed in the intergenic regions of the 21 complete genomes of Synergistetes, which includes genera *Synergistetes*, *Cloacibacillus*, *Pyramidobacter* and *Acetomicrobium*. The genus *Thermanaerovibrio* was not included as only two complete genomes are available (Table 1). To be noted, since the DAPA analysis revealed that His-CodY<sub>DSM</sub> could interact with both the CodY box of Firmicutes (5'-AATTTTTCWGAAAATT) and a palindromic sequence with an extended spacer region (Figure 3 and Table S5), both motifs were used to search for potential targets in the corresponding genus. It was found that the consensus sequence was widely distributed in the intergenic regions of genomes analyzed (Table S7), accounting for approximately 8–13% of the genes in a genome (Figure 5 and Table S8). Although the percentages were comparable between the four genera, genera *Synergistetes* and *Cloacibacillus* showed higher numbers of loci that harbor both motifs compared to those of genera *Acetomicrobium* and *Pyramidobacter*. The wide distribution of this novel consensus sequence in a genome suggests that the CodY-like protein is a global regulator. Among these loci, homologs of 32 loci are parts of the CodY regulon in Firmicutes (Table S9). These loci include genes encoding proteins involved in BCAA biosynthesis (*ilvA*, *ilvN*, *ilvB*, *ilvD*, *ilvC*, *leuC*, WP\_051682752.1, WP\_200778737.1 and WP\_037975652.1), BCAA uptake (WP\_037976683.1, WP\_037975074.1, WP\_037976889.1 and WP\_009202431.1), proteolysis (*clpB*, *clpX*, *clpP*, *fstH* and *resP*), peptide degradation (WP\_037974402.1, *pepV*, *def*, WP\_051682620.1, WP\_051682707.1, WP\_078015053.1, WP\_037973903.1, WP\_111089616.1, WP\_037974668.1 and WP\_008708838.1), and peptide and amino acid transport (WP\_009201625.1, WP\_078015996.1, WP\_008709275.1 and WP\_120373146.1). This observation supports the hypothesis that the CodY-like global regulator of Synergistetes functions in parallel to Firmicutes CodY.

To further verify this hypothesis, we computed the functional enrichment of groups of genes listed in Table S7 in the genera *Acetomicrobium*, *Cloacibacillus*, *Pyramidobacter*, and *Synergistes* (Figure 6). Three main clusters emerged, with proteins in the categories of amino acid transport and metabolism, transporters, transcriptional regulators, carbohydrate transport and metabolism, and energy production and conversion being the most conserved in the four genera. The identification of genes encoding transcriptional regulators, including HrcA and regulators of the P-II family, GntR family, and CtsR family (Table S7), suggests that the regulon of the CodY-like protein would contain a set of genes indirectly regulated by the CodY-like protein. Specifically, regulators of both the PII family and the GntR family are closely associated with the regulation of nitrogen metabolism and the nitrogen-carbon interconversion in both Gram-positive and Gram-negative bacteria. HrcA, a heat-inducible transcription repressor, and CtsR, a transcriptional regulator of class III stress genes [34,35], could be involved in the stress responses. Additionally, transcriptional regulators of the LacI, LysR, TetR/AcrR, MurR/RpiR, MerR and MarR families were also found. These regulators are implicated in the regulation of many cellular processes, including metabolism, metal ion homeostasis, antibiotics resistance, virulence expression and stress responses, and thus, a regulon governed by the CodY-like protein is likely to indirectly control multiple cellular activities in Synergistetes.



**Figure 5.** The distribution of genes regulated by the CodY-like protein in the available genomes of the phylum Synergistetes. The number of loci that harbor the canonical CodY box, the newly defined extended palindromic sequence, and both motifs in the promoter regions are shown in orange, green and light orange, respectively. The percentages of total targets in the genomes are indicated.

Although the genes in the second cluster were generally less conserved, compared to the genes described above between the four genera, significant conservation within a genus was detected in genes encoding proteins involved in genetic information storage and processing, including DNA replication, recombination and repair, transcription, translation, ribosomal structure and biogenesis, nucleotide transport and metabolism. Genes encoding proteins involved in the biosynthesis of cell envelopes, secondary metabolites biosynthesis, metabolism of cofactors and vitamins, heat-shock proteins and molecular chaperones, and horizontal gene transfer were also present in the second cluster (Figure 6). Loci that were least conserved between genera were genes involved in the metabolism of terpenoids and polyketides, phage production, post-translational modification and protein turnover, the toxin-antitoxin system, cell motility and chemotaxis, CRISPR and other defense systems, and cellular community. Nevertheless, conservation among species within a genus was evident. For instance, loci associated with post-translational modification, protein turnover, cell cycle control, cell division, chromosome partition, coenzyme transport and metabolism, cell motility and chemotaxis, and the CRISPR system were enriched in the genus *Acetomicrobium*, and loci encoding proteins of the toxin-antitoxin systems, lipid transport and metabolism, and other defense systems were enriched in the genus *Synergistes*. Taken together, the loci involved in nutrient uptake, general metabolism and regulation were most conserved between the four genera, whereas loci involved in defense systems and stress responses were less conserved. The variations in the putative regulons of the CodY-like regulatory protein between genera may be associated with the natural niches of these microbes.



**Figure 6.** A pathway-enrichment heat map showing the clustering of loci regulated by the CodY-like protein. The positions of the four genera are color-coded and listed on the top of the figure. The number of loci regulated by the CodY-like protein was analyzed based on the predicted function. The number was converted using  $\log_2(N + 1)$ , where the N is the number of the identified target genes.

### 3. Discussion

CodY has been extensively studied in the Firmicutes for its dominant role in metabolism, stress responses and pathogenesis [23,36–38]. This study confirmed the presence of CodY orthologues in Synergistetes, which generally exhibits a G + C content close to, or greater than, 50% [39–43]. Thus, the distribution of CodY and the CodY-like protein in bacteria may be more widespread than previously known, and regulation by CodY and its homologs may have similar impacts on bacteria occupying different niches. The differences in the G + C content, the ecological niches and metabolic activities between Firmicutes and Synergistetes would suggest that CodY and the CodY-like protein evolved divergently. As the evolutionary relationship of these proteins follows the phylogenetic relationship of the genera in phyla Firmicutes and Synergistetes, genes encoding CodY and the CodY-like protein should originate from a common ancestor. The results of this study indicate that these two proteins are orthologues that are conserved in domain structures and functions, but distinct in their overall amino acid sequence.

Past transcriptomic studies of the CodY regulon in Firmicutes, either by microarray or RNA sequencing, often identified loci without a canonical CodY box in the promoter region. For instance, approximately 60% and 30% of the identified CodY targets in *B. subtilis* and *L. lactis*, respectively, are without a canonical CodY box [20]. Although some of these loci

are regulated by transcriptional regulators whose activity is governed by CodY, CodY likely directly regulates the rest of the targets through binding to a modified consensus motif. This hypothesis is supported by more recent studies in *B. subtilis* [44], *L. monocytogenes* [22], and *Staphylococcus aureus* [45], in which a modified CodY-binding consensus generated from two overlapping canonical CodY boxes (5'-AATTTTCWGAAWTTCWGAAAATT) was reported. Unlike the CodY box described in Firmicutes, the binding consensus in Synergistetes is more like a palindromic sequence with AT-rich arms separated by an extended spacer region, as neither the lengths nor the sequences of the identified consensus sequences fit an overlapping arrangement. Although it is possible that the targets identified in this study do not cover the full regulon, the presence of the binding consensus at the 5' region of multiple genes strongly suggests that similar to CodY in Firmicutes, this newly identified GAF-wHTH protein is a global regulatory protein in Synergistetes.

As a system for genetic manipulation has yet to be established in microbes of the phylum Synergistetes, we selected the DAPA analysis to analyze the DNA-binding activity of the CodY-like protein. Attempts were also made to use gel electrophoresis mobility shift assays (EMSA) [46] to investigate the binding specificity of the CodY-like protein; however, the solubilities of the recombinant CodY-like proteins of Synergistetes were generally lower than those required for effective analysis by EMSA with biotin-labeled probes. Using Protein-Sol (<https://protein-sol.manchester.ac.uk/>, accessed on 14 January 2022), we found that the solubility of the recombinant CodY-like protein of Synergistetes was significantly lower than CodY from *Streptococcus salivarius* 57.I [47]. Proteins with a Protein-Sol calculated value above 0.45 are considered highly soluble [48]; and, while the Protein-Sol value of CodY from *S. salivarius* 57.I is 0.719, and the value for the CodY-like protein in *C. everyensis* DSM19522 is 0.418. Similar results were also found in CodY-like proteins of two randomly selected species: *Thermanaerovibrio acidaminovorans* DSM 6589, with a value of 0.455, and *C. porcorum* CL-84, with a value of 0.426. By using Ni-NTA magnetic beads, we were able to perform DAPA experiments in a relatively large reaction volume, compared to EMSA analysis, which circumvented any limitations of the low solubility of the CodY-like protein.

Although the difference in the GAF domain between CodY and the CodY-like protein could result in a difference in the effectors, both proteins are involved in the regulation of BCAA biosynthesis. Thus, BCAA is also likely to be the effector of the CodY-like protein. On the other hand, while both BCAA and GTP are necessary to activate CodY in *B. subtilis* [49], CodY of *L. lactis* [50], *Streptococcus mutans* [51], and *S. pneumoniae* [21] is not activated by GTP, suggesting that the choice and efficiency of effector molecules to activate CodY activity could be affected by the ecological niche and physiologic activity of the bacteria. As microbes of Synergistetes could metabolize a variety of substrates, including DHP [5], plant toxin fluoroacetate [52], and mucin, it is also possible that microbes of Synergistetes could utilize different metabolites to signal intracellular nutrient status, and perhaps to activate the CodY-like protein.

As members of phylum Synergistetes inhabit a wide range of extreme environments, it is expected that they have evolved sophisticated features to cope with environmental stresses, including extreme temperature, pH, toxic compounds and challenges from other microbes or bacteriophages. As loci in motility/chemotaxis, defense mechanisms and CRISPR systems all are less conserved among the regulons governed by the CodY-like proteins in Synergistetes, it would suggest that the regulons could vary according to the natural niches of all species in the phylum. For instance, based on our analysis, the CodY-like protein could potentially regulate the chemotaxis and motility of flagella-bearing *Acetomicrobium thermoterrenum* strain DSM 13490 by regulating the expression of loci encoding the methyl-accepting chemotaxis protein (Mcp) and proteins involved in flagella biosynthesis. Thus, regulation by the CodY-like protein could potentially fine-tune the activity of chemotaxis and swimming activity. Such regulation seems necessary for bacteria that reside in constantly changing environments, and the regulation could provide certain advantages for these bacteria to thrive in harsh environments.

Taken together, although Synergistetes and Firmicutes are phylogenetically distinct, both CodY and the CodY-like protein are conserved with respect to their regulatory function. Based on the potential targets of the CodY-like protein, the regulation spectrum of the CodY-like protein overlaps with that of the Firmicutes CodY regulon. Thus, similar to Firmicutes, regulation by the CodY-like protein is likely to be essential for the microbes of Synergistetes to cope with stresses in their natural niches.

## 4. Materials and Methods

### 4.1. Bacterial Genome and Metagenome Database

The cultivable bacterial genomes were obtained from the NCBI Genbank database of sequenced bacteria (last updated August 2020) (<http://ftp.ncbi.nlm.nih.gov/refseq/release/>, accessed on 24 August 2020). Metagenomic data used in this study included genomes obtained from samples from various ecological sources (<https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Undefined&id=408169&lvl=3&keep=1&srchmode=1&unlock/>, accessed on 26 October 2020), and human gut metagenomic data ([http://ftp.ebi.ac.uk/pub/databases/metagenomics/mgnify\\_genomes/human-gut/v1.0/uhgp\\_catalogue/uhgp-100.tar.gz/](http://ftp.ebi.ac.uk/pub/databases/metagenomics/mgnify_genomes/human-gut/v1.0/uhgp_catalogue/uhgp-100.tar.gz/), accessed on 6 November 2020).

### 4.2. Identification of CodY Homologous Proteins

HMMER, version 3.1b2 [53], was used to search the GAF (PF06018) and HTH (PF08222) domains of CodY in the deduced amino acid sequence derived from the bacterial genome and metagenome databases. The significance was set at an e-value  $< 10^{-3}$ . A putative CodY was identified in proteins that contained both the GAF and HTH domains. BlastP analysis was used to calculate the average amino acid identity of the CodY-like proteins of Synergistetes (Table 1) and the CodY proteins of Firmicutes (Table S2).

### 4.3. Phylogenetic Analysis of 16S rRNA and the CodY-like/CodY Proteins from Synergistetes and Other Phyla

The 16S rRNA genes from 35 strains of the phylum Synergistetes (Table 1), 17 species of the Firmicutes (Table S2) and 12 species from the phyla Proteobacteria, Fusobacteria, Bacteroidetes, Actinobacteria and Aquificae (Table S3) were retrieved from the database. An alignment of these 16S genes was performed using ClustalW integrated in MEGA version 7 [54] with standard parameters. The sequences of the *codY* genes from species of Synergistetes and Firmicutes were retrieved from the database, and an alignment of the deduced protein sequence was prepared using MUSCLE [55] integrated into MEGA version 7 with standard parameters. The maximum likelihood algorithm (ML) with 1000 bootstrap replications was used for phylogenetic tree reconstruction. The phylogenetic tree was managed and visualized by iTOL, version 5 [56].

### 4.4. Purification of the Recombinant CodY-like Protein

A preparation of the total cellular DNA of *C. evryensis* DSM19522 was purchased from the Leibniz Institute DSMZ (Braunschweig, Germany). The coding region of the putative *codY* (CLOEV\_RS08630) was PCR amplified from chromosomal DNA by using the Blend Taq<sup>TM</sup>-Plus (Osaka, Toyobo) with primers DSM19522\_codY-SphI\_S and DSM19522\_codY-HindIII\_AS. The estimated melting temperatures of primers DSM19522\_codY-SphI\_S and DSM19522\_codY-HindIII\_AS are 60.7 °C and 61.4 °C, respectively. The annealing temperature of the PCR was set at 56 °C (30 s), and reaction was performed for 35 cycles. Restriction recognition sequences were included in the primers to facilitate cloning. The PCR product was restriction digested and cloned into pQE30 (Qiagen, Hilden, Germany) in *E. coli* M15. The identity of the recombinant plasmid was verified by sequencing analysis. The recombinant histidine-tagged CodY-like protein (His-CodY<sub>DSM</sub>) was induced and purified from the recombinant *E. coli* strain by HIS-Select Nickel Affinity Gel (Sigma, St. Louis, MO, USA) under native conditions, according to the manufacturer's recommendation. The

concentration of the purified protein was determined by Bio-Rad Protein Assay based on the method of Bradford [57]. All primers used in this study are listed in Table S6.

#### 4.5. DAPA

All oligonucleotides used in DAPA were end-labeled using the Pierce biotin 3' end DNA labeling kit (Thermo, Waltham, MA, USA). Two complementary biotin-labeled oligonucleotides containing the desired sequences were annealed and then used in DAPA reactions. Various amounts of purified His-CodY<sub>DMS</sub> was mixed with 10 µL Mag-Beads-Ni-NTA (Tools, Taiwan) in separate reactions, with constant rotation for 30 min at room temperature. At the end of the reaction, the beads were washed with 1 mL reaction buffer (10 mM MgCl<sub>2</sub>, 50 mM KCl, 10 mM Tris, [pH 7.5], 1 mM EDTA, 50 µg ml<sup>-1</sup> BSA) in a magnetic field to remove unbound proteins. The protein-DNA binding reaction was carried out in 50 µL of the reaction buffer containing 0.1 pmol biotin-labeled probe and poly[dI-dC] at 20 µg mL<sup>-1</sup>, and the reaction was incubated with tilt rotation for 30 min at room temperature. The final concentrations of His-CodY<sub>DMS</sub> in the reactions were 0.02, 0.1, 0.5 and 2.5 µM, respectively. At the end of incubation, the beads were washed with 200 µL of reaction buffer three times, and then resuspended in 20 µL TE buffer (50 mM Tris, 10 mM EDTA [pH 8]) containing proteinase K at 2 mg mL<sup>-1</sup>. The digestion was carried out at 37 °C for 30 min to release the DNA. The final product was resolved on 6% non-denaturing polyacrylamide gels, electro-transferred onto nylon membranes, and detected using a chemiluminescent nucleic acid detection module kit (Pierce). The DAPA experiments were repeated six times with three batches of purified His-CodY<sub>DMS</sub>.

#### 4.6. Binding Site Analysis and Target Gene Prediction

The sequence of a 400-bp region 5' to the *xerC* gene, the first gene of the *codY*-containing operon of species of the phylum Synergistetes were extracted from the genomes of 35 cultivatable isolates, 17 human gut metagenome-assembled genomes and 22 ecological metagenome assembly sequences. All these sequences were then used for common motif discovery using MEME motif-searching algorithm [58]. Genera with less than three genomes available were excluded from the analysis. The default settings were used with the following exceptions: the minimum length was set to 15 bases and interrogated sequences with palindromes only.

To define the possible regulon of the CodY-like protein, all promoter regions of the 21 complete genomes in the phylum Synergistetes were scanned for 5'-AATTTTCWGAAAATT and the newly identified consensus binding sequence of the CodY-like protein by FIMO [59].

#### 4.7. Functional Enrichment of the Identified Target Genes Regulated by the CodY-like Protein

The functional clustering of the target genes regulated by the CodY-like proteins was initially analyzed using Roary [60]. The identified target genes were further enriched based on predicted functions using EggNog [61], Clusters of Orthologous Groups (COGs) [62] and KEGG orthology [63]. The enrichment score was calculated using log<sub>2</sub>(N+1) (where N is the number of the identified genes). R script was used to generate the heat map showing the enrichment scores.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms23147911/s1>.

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

**Funding:** This research was supported in part by the National Key Research and Development Program of China (2021YFA0909500 to SH), the National Natural Science Foundation of China (31200968 to JG) and the Ministry of Science and Technology (MOST) of Taiwan, (108-2320-B-182-020-YM3 to YMC).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Acknowledgments:** We thank Y. Luo, S. Gao and X. Ge for technical assistance and discussion of the results. We thank R. Faustoferri for review of the manuscript.

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

## References

1. Jumas-Bilak, E.; Roudiere, L.; Marchandin, H. Description of ‘*Synergistetes*’ phyl. nov. and emended description of the phylum ‘*Deferribacteres*’ and of the family *Syntrophomonadaceae*, phylum ‘*Firmicutes*’. *Int. J. Syst. Evol. Microbiol.* **2009**, *59*, 1028–1035. [[CrossRef](#)]
2. Bhandari, V.; Gupta, R.S. Molecular signatures for the phylum *Synergistetes* and some of its subclades. *Antonie Van Leeuwenhoek* **2012**, *102*, 517–540. [[CrossRef](#)] [[PubMed](#)]
3. Maune, M.W.; Tanner, R.S. Description of *Anaerobaculum hydrogeniformans* sp. nov., an anaerobe that produces hydrogen from glucose, and emended description of the genus *Anaerobaculum*. *Int. J. Syst. Evol. Microbiol.* **2012**, *62*, 832–838. [[CrossRef](#)] [[PubMed](#)]
4. Dahle, H.; Birkeland, N.K. *Thermovirga lienii* gen. nov., sp. nov., a novel moderately thermophilic, anaerobic, amino-acid-degrading bacterium isolated from a North Sea oil well. *Int. J. Syst. Evol. Microbiol.* **2006**, *56*, 1539–1545. [[CrossRef](#)]
5. Allison, M.J.; Mayberry, W.R.; McSweeney, C.S.; Stahl, D.A. *Synergistes jonesii*, gen. nov., sp. nov.: A rumen bacterium that degrades toxic pyridinediols. *Syst. Appl. Microbiol.* **1992**, *15*, 522–529. [[CrossRef](#)]
6. Honda, T.; Fujita, T.; Tonouchi, A. *Aminivibrio pyruvatiphilus* gen. nov., sp. nov., an anaerobic, amino-acid-degrading bacterium from soil of a Japanese rice field. *Int. J. Syst. Evol. Microbiol.* **2013**, *63*, 3679–3686. [[CrossRef](#)]
7. Jumas-Bilak, E.; Bouvet, P.; Allen-Vercocoe, E.; Aujoulat, F.; Lawson, P.A.; Jean-Pierre, H.; Marchandin, H. *Rarimicrobium hominis* gen. nov., sp. nov., representing the fifth genus in the phylum *Synergistetes* that includes human clinical isolates. *Int. J. Syst. Evol. Microbiol.* **2015**, *65*, 3965–3970. [[CrossRef](#)]
8. Ndongo, S.; Dubourg, G.; Fournier, P.E.; Raoult, D. *Pacaella massiliensis* gen. nov., sp. nov., a new bacterial species isolated from the human gut. *New Microbes New Infect.* **2017**, *16*, 32–33. [[CrossRef](#)]
9. Qiu, Y.L.; Hanada, S.; Kamagata, Y.; Guo, R.B.; Sekiguchi, Y. *Lactivibrio alcoholicus* gen. nov., sp. nov., an anaerobic, mesophilic, lactate-, alcohol-, carbohydrate- and amino-acid-degrading bacterium in the phylum *Synergistetes*. *Int. J. Syst. Evol. Microbiol.* **2014**, *64*, 2137–2145. [[CrossRef](#)]
10. Vartoukian, S.R.; Downes, J.; Palmer, R.M.; Wade, W.G. *Fretibacterium fastidiosum* gen. nov., sp. nov., isolated from the human oral cavity. *Int. J. Syst. Evol. Microbiol.* **2013**, *63*, 458–463. [[CrossRef](#)] [[PubMed](#)]
11. McCracken, B.A.; Nathalia Garcia, M. Phylum *Synergistetes* in the oral cavity: A possible contributor to periodontal disease. *Anaerobe* **2021**, *68*, 102250. [[CrossRef](#)] [[PubMed](#)]
12. Wade, W.G. The oral microbiome in health and disease. *Pharmacol. Res.* **2013**, *69*, 137–143. [[CrossRef](#)]
13. Chovatia, M.; Sikorski, J.; Schroder, M.; Lapidus, A.; Nolan, M.; Tice, H.; Glavina Del Rio, T.; Copeland, A.; Cheng, J.F.; Lucas, S.; et al. Complete genome sequence of *Thermanaerovibrio acidaminovorans* type strain (Su883). *Stand. Genom. Sci.* **2009**, *1*, 254–261. [[CrossRef](#)] [[PubMed](#)]
14. Davis, C.K.; Webb, R.I.; Sly, L.I.; Denman, S.E.; McSweeney, C.S. Isolation and survey of novel fluoroacetate-degrading bacteria belonging to the phylum *Synergistetes*. *FEMS Microbiol. Ecol.* **2012**, *80*, 671–684. [[CrossRef](#)] [[PubMed](#)]
15. Schoch, C.L.; Ciufu, S.; Domrachev, M.; Hottton, C.L.; Kannan, S.; Khovanskaya, R.; Leipe, D.; McVeigh, R.; O’Neill, K.; Robertse, B.; et al. NCBI Taxonomy: A comprehensive update on curation, resources and tools. *Database* **2020**, *2020*, baaa062. [[CrossRef](#)] [[PubMed](#)]
16. Plichta, D.R.; Juncker, A.S.; Bertalan, M.; Rettedal, E.; Gautier, L.; Varela, E.; Manichanh, C.; Fouqueray, C.; Levenez, F.; Nielsen, T.; et al. Transcriptional interactions suggest niche segregation among microorganisms in the human gut. *Nat. Microbiol.* **2016**, *1*, 16152. [[CrossRef](#)]
17. Hugenholtz, P.; Hooper, S.D.; Kyrpides, N.C. Focus: *Synergistetes*. *Environ. Microbiol.* **2009**, *11*, 1327–1329. [[CrossRef](#)]
18. Slack, F.J.; Serron, P.; Joyce, E.; Sonenshein, A.L. A gene required for nutritional repression of the *Bacillus subtilis* dipeptide permease operon. *Mol. Microbiol.* **1995**, *15*, 689–702. [[CrossRef](#)]
19. Sonenshein, A.L. Control of key metabolic intersections in *Bacillus subtilis*. *Nat. Rev. Microbiol.* **2007**, *5*, 917–927. [[CrossRef](#)]
20. Guedon, E.; Sperandio, B.; Pons, N.; Ehrlich, S.D.; Renault, P. Overall control of nitrogen metabolism in *Lactococcus lactis* by CodY, and possible models for CodY regulation in Firmicutes. *Microbiology* **2005**, *151*, 3895–3909. [[CrossRef](#)]
21. Hendriksen, W.T.; Bootsma, H.J.; Estevao, S.; Hoogenboezem, T.; de Jong, A.; de Groot, R.; Kuipers, O.P.; Hermans, P.W. CodY of *Streptococcus pneumoniae*: Link between nutritional gene regulation and colonization. *J. Bacteriol.* **2008**, *190*, 590–601. [[CrossRef](#)] [[PubMed](#)]

22. Biswas, R.; Sonenshein, A.L.; Belitsky, B.R. Genome-wide identification of *Listeria monocytogenes* CodY-binding sites. *Mol. Microbiol.* **2020**, *113*, 841–858. [[CrossRef](#)] [[PubMed](#)]
23. Stenz, L.; Francois, P.; Whiteson, K.; Wolz, C.; Linder, P.; Schrenzel, J. The CodY pleiotropic repressor controls virulence in gram-positive pathogens. *FEMS Immunol. Med. Microbiol.* **2011**, *62*, 123–139. [[CrossRef](#)] [[PubMed](#)]
24. Han, A.R.; Kang, H.R.; Son, J.; Kwon, D.H.; Kim, S.; Lee, W.C.; Song, H.K.; Song, M.J.; Hwang, K.Y. The structure of the pleiotropic transcription regulator CodY provides insight into its GTP-sensing mechanism. *Nucleic Acids Res.* **2016**, *44*, 9483–9493. [[CrossRef](#)] [[PubMed](#)]
25. Levdivikov, V.M.; Blagova, E.; Colledge, V.L.; Lebedev, A.A.; Williamson, D.C.; Sonenshein, A.L.; Wilkinson, A.J. Structural rearrangement accompanying ligand binding in the GAF domain of CodY from *Bacillus subtilis*. *J. Mol. Biol.* **2009**, *390*, 1007–1018. [[CrossRef](#)]
26. den Hengst, C.D.; van Hijum, S.A.; Geurts, J.M.; Nauta, A.; Kok, J.; Kuipers, O.P. The *Lactococcus lactis* CodY regulon: Identification of a conserved *cis*-regulatory element. *J. Biol. Chem.* **2005**, *280*, 34332–34342. [[CrossRef](#)]
27. Belitsky, B.R.; Sonenshein, A.L. Genetic and biochemical analysis of CodY-binding sites in *Bacillus subtilis*. *J. Bacteriol.* **2008**, *190*, 1224–1236. [[CrossRef](#)]
28. Joseph, P.; Ratnayake-Lecamwasam, M.; Sonenshein, A.L. A region of *Bacillus subtilis* CodY protein required for interaction with DNA. *J. Bacteriol.* **2005**, *187*, 4127–4139. [[CrossRef](#)]
29. Levdivikov, V.M.; Blagova, E.; Joseph, P.; Sonenshein, A.L.; Wilkinson, A.J. The structure of CodY, a GTP- and isoleucine-responsive regulator of stationary phase and virulence in Gram-positive bacteria. *J. Biol. Chem.* **2006**, *281*, 11366–11373. [[CrossRef](#)]
30. Villapakkam, A.C.; Handke, L.D.; Belitsky, B.R.; Levdivikov, V.M.; Wilkinson, A.J.; Sonenshein, A.L. Genetic and biochemical analysis of the interaction of *Bacillus subtilis* CodY with branched-chain amino acids. *J. Bacteriol.* **2009**, *191*, 6865–6876. [[CrossRef](#)]
31. Kreth, J.; Chen, Z.; Ferretti, J.; Malke, H. Counteractive balancing of transcriptome expression involving CodY and CovRS in *Streptococcus pyogenes*. *J. Bacteriol.* **2011**, *193*, 4153–4165. [[CrossRef](#)] [[PubMed](#)]
32. Taboada, B.; Estrada, K.; Ciria, R.; Merino, E. Operon-mapper: A web server for precise operon identification in bacterial and archaeal genomes. *Bioinformatics* **2018**, *34*, 4118–4120. [[CrossRef](#)] [[PubMed](#)]
33. Bailey, T.L.; Boden, M.; Buske, F.A.; Frith, M.; Grant, C.E.; Clementi, L.; Ren, J.; Li, W.W.; Noble, W.S. MEME SUITE: Tools for motif discovery and searching. *Nucleic Acids Res.* **2009**, *37*, W202–W208. [[CrossRef](#)] [[PubMed](#)]
34. Elsholz, A.K.; Michalik, S.; Zuhlke, D.; Hecker, M.; Gerth, U. CtsR, the Gram-positive master regulator of protein quality control, feels the heat. *EMBO J.* **2010**, *29*, 3621–3629. [[CrossRef](#)]
35. Kruger, E.; Hecker, M. The first gene of the *Bacillus subtilis* *clpC* operon, *ctsR*, encodes a negative regulator of its own operon and other class III heat shock genes. *J. Bacteriol.* **1998**, *180*, 6681–6688. [[CrossRef](#)] [[PubMed](#)]
36. Brinsmade, S.R. CodY, a master integrator of metabolism and virulence in Gram-positive bacteria. *Curr. Genet.* **2017**, *63*, 417–425. [[CrossRef](#)]
37. Geng, J.; Huang, S.C.; Chen, Y.Y.; Chiu, C.H.; Hu, S.; Chen, Y.M. Impact of growth pH and glucose concentrations on the CodY regulatory network in *Streptococcus salivarius*. *BMC Genom.* **2018**, *19*, 386. [[CrossRef](#)]
38. Richardson, A.R.; Somerville, G.A.; Sonenshein, A.L. Regulating the intersection of metabolism and pathogenesis in Gram-positive bacteria. *Microbiol. Spectr.* **2015**, *3*. [[CrossRef](#)]
39. Menes, R.J.; Muxi, L. *Anaerobaculum mobile* sp. nov., a novel anaerobic, moderately thermophilic, peptide-fermenting bacterium that uses crotonate as an electron acceptor, and emended description of the genus *Anaerobaculum*. *Int. J. Syst. Evol. Microbiol.* **2002**, *52*, 157–164. [[CrossRef](#)]
40. Baena, S.; Fardeau, M.L.; Ollivier, B.; Labat, M.; Thomas, P.; Garcia, J.L.; Patel, B.K. *Aminomonas paucivorans* gen. nov., sp. nov., a mesophilic, anaerobic, amino-acid-utilizing bacterium. *Int. J. Syst. Bacteriol.* **1999**, *49 Pt 3*, 975–982. [[CrossRef](#)]
41. Sekiguchi, Y.; Kamagata, Y.; Nakamura, K.; Ohashi, A.; Harada, H. *Syntrophothermus lipocalidus* gen. nov., sp. nov., a novel thermophilic, syntrophic, fatty-acid-oxidizing anaerobe which utilizes isobutyrate. *Int. J. Syst. Evol. Microbiol.* **2000**, *50 Pt 2*, 771–779. [[CrossRef](#)] [[PubMed](#)]
42. Diaz, C.; Baena, S.; Fardeau, M.L.; Patel, B.K.C. *Aminiphilus circumscriptus* gen. nov., sp. nov., an anaerobic amino-acid-degrading bacterium from an upflow anaerobic sludge reactor. *Int. J. Syst. Evol. Microbiol.* **2007**, *57*, 1914–1918. [[CrossRef](#)] [[PubMed](#)]
43. Surkov, A.V.; Dubinina, G.A.; Lysenko, A.M.; Glockner, F.O.; Kuever, J. *Dethiosulfovibrio russensis* sp. nov., *Dethiosulfovibrio marinus* sp. nov. and *Dethiosulfovibrio acidaminovorans* sp. nov., novel anaerobic, thiosulfate- and sulfur-reducing bacteria isolated from ‘Thiodendron’ sulfur mats in different saline environments. *Int. J. Syst. Evol. Microbiol.* **2001**, *51*, 327–337. [[CrossRef](#)] [[PubMed](#)]
44. Belitsky, B.R.; Sonenshein, A.L. Genome-wide identification of *Bacillus subtilis* CodY-binding sites at single-nucleotide resolution. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 7026–7031. [[CrossRef](#)]
45. Majerczyk, C.D.; Dunman, P.M.; Luong, T.T.; Lee, C.Y.; Sadykov, M.R.; Somerville, G.A.; Bodi, K.; Sonenshein, A.L. Direct targets of CodY in *Staphylococcus aureus*. *J. Bacteriol.* **2010**, *192*, 2861–2877. [[CrossRef](#)]
46. Hellman, L.M.; Fried, M.G. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat. Protoc.* **2007**, *2*, 1849–1861. [[CrossRef](#)]
47. Huang, S.C.; Burne, R.A.; Chen, Y.Y. The pH-dependent expression of the urease operon in *Streptococcus salivarius* is mediated by CodY. *Appl. Environ. Microbiol.* **2014**, *80*, 5386–5393. [[CrossRef](#)]
48. Hebditch, M.; Carballo-Amador, M.A.; Charonis, S.; Curtis, R.; Warwicker, J. Protein-Sol: A web tool for predicting protein solubility from sequence. *Bioinformatics* **2017**, *33*, 3098–3100. [[CrossRef](#)]

49. Shivers, R.P.; Sonenshein, A.L. Activation of the *Bacillus subtilis* global regulator CodY by direct interaction with branched-chain amino acids. *Mol. Microbiol.* **2004**, *53*, 599–611. [[CrossRef](#)]
50. den Hengst, C.D.; Curley, P.; Larsen, R.; Buist, G.; Nauta, A.; van Sinderen, D.; Kuipers, O.P.; Kok, J. Probing direct interactions between CodY and the *oppD* promoter of *Lactococcus lactis*. *J. Bacteriol.* **2005**, *187*, 512–521. [[CrossRef](#)]
51. Lemos, J.A.; Nascimento, M.M.; Lin, V.K.; Abranches, J.; Burne, R.A. Global regulation by (p)ppGpp and CodY in *Streptococcus mutans*. *J. Bacteriol.* **2008**, *190*, 5291–5299. [[CrossRef](#)]
52. Kang, S.; Khan, S.; Webb, R.; Denman, S.; McSweeney, C. Characterization and survey in cattle of a rumen *Pyrimadobacter* sp. which degrades the plant toxin fluoroacetate. *FEMS Microbiol. Ecol.* **2020**, *96*, fiae077. [[CrossRef](#)] [[PubMed](#)]
53. Finn, R.D.; Clements, J.; Eddy, S.R. HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* **2011**, *39*, W29–W37. [[CrossRef](#)] [[PubMed](#)]
54. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [[CrossRef](#)] [[PubMed](#)]
55. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [[CrossRef](#)] [[PubMed](#)]
56. 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](#)] [[PubMed](#)]
57. Bradford, M.M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **1976**, *72*, 248–254. [[CrossRef](#)]
58. Bailey, T.L.; Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **1994**, *2*, 28–36.
59. Grant, C.E.; Bailey, T.L.; Noble, W.S. FIMO: Scanning for occurrences of a given motif. *Bioinformatics* **2011**, *27*, 1017–1018. [[CrossRef](#)]
60. Page, A.J.; Cummins, C.A.; Hunt, M.; Wong, V.K.; Reuter, S.; Holden, M.T.; Fookes, M.; Falush, D.; Keane, J.A.; Parkhill, J. Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **2015**, *31*, 3691–3693. [[CrossRef](#)]
61. Huerta-Cepas, J.; Szklarczyk, D.; Heller, D.; Hernandez-Plaza, A.; Forslund, S.K.; Cook, H.; Mende, D.R.; Letunic, I.; Rattei, T.; Jensen, L.J.; et al. eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **2019**, *47*, D309–D314. [[CrossRef](#)] [[PubMed](#)]
62. Galperin, M.Y.; Wolf, Y.I.; Makarova, K.S.; Vera Alvarez, R.; Landsman, D.; Koonin, E.V. COG database update: Focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res.* **2021**, *49*, D274–D281. [[CrossRef](#)] [[PubMed](#)]
63. Kanehisa, M.; Sato, Y.; Kawashima, M.; Furumichi, M.; Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **2016**, *44*, D457–D462. [[CrossRef](#)] [[PubMed](#)]