Food Research International 166 (2023) 112605

Contents lists available at ScienceDirect

Food Research International

journal homepage: www.elsevier.com/locate/foodres

Comparative genome identification of accessory genes associated with strong biofilm formation in *Vibrio parahaemolyticus*

Dan Wang [a], Graham C. Fletcher [b], Dragana Gagic [c], Stephen L.W. On [d], Jon S. Palmer [a], Steve H. Flint [a,*]

[a] School of Food and Advanced Technology, Massey University, Private Bag 11222, Palmerston North, New Zealand
[b] The New Zealand Institute for Plant & Food Research Limited, Private Bag 92169, Auckland 1142, New Zealand
[c] School of Fundamental Sciences, Massey University, Private Bag 11222, Palmerston North, New Zealand
[d] Faculty of Agriculture and Life Sciences, Lincoln University, Private Bag 85084, Canterbury, New Zealand

ARTICLE INFO

Keywords:
whole genome sequencing (WGS)
Comparative genome
Cellulose
Horizontal gene transfer (HGT)
Seafood safety

ABSTRACT

Vibrio parahaemolyticus biofilms on the seafood processing plant surfaces are a potential source of seafood contamination and subsequent food poisoning. Strains differ in their ability to form biofilm, but little is known about the genetic characteristics responsible for biofilm development. In this study, pangenome and comparative genome analysis of *V. parahaemolyticus* strains reveals genetic attributes and gene repertoire that contribute to robust biofilm formation. The study identified 136 accessory genes that were exclusively present in strong biofilm forming strains and these were functionally assigned to the Gene Ontology (GO) pathways of cellulose biosynthesis, rhamnose metabolic and catabolic processes, UDP-glucose processes and O antigen biosynthesis ($p < 0.05$). Strategies of CRISPR-Cas defence and MSHA pilus-led attachment were implicated via Kyoto Encyclopedia of Genes and Genomes (KEGG) annotation. Higher levels of horizontal gene transfer (HGT) were inferred to confer more putatively novel properties on biofilm-forming *V. parahaemolyticus*. Furthermore, cellulose biosynthesis, a neglected potential virulence factor, was identified as being acquired from within the order *Vibrionales*. The cellulose synthase operons in *V. parahaemolyticus* were examined for their prevalence (22/138, 15.94 %) and were found to consist of the genes *bcsG*, *bcsE*, *bcsQ*, *bcsA*, *bcsB*, *bcsZ*, *bcsC*. This study provides insights into robust biofilm formation of *V. parahaemolyticus* at the genomic level and facilitates: identification of key attributes for robust biofilm formation, elucidation of biofilm formation mechanisms and development of potential targets for novel control strategies of persistent *V. parahaemolyticus*.

1. Introduction

Microbial cells can colonize biotic surfaces, equipment (e.g. conveyor belts, stainless steel processing bench, pipes), or packaging materials (e.g., glass, polystyrene), forming complex community matrices covered by extracellular polymeric substances (EPSs) (Lianou, Nychas, & Koutsoumanis, 2020). Biofilms enable increased resistance by shielding pathogens and spoilage bacteria from environmental stresses, acting as hot spots for horizontal gene transfer (HGT) of virulence genes, and transforming previously benign strains into pathogens (Stalder & Top, 2016). The enhanced resistance of biofilms poses challenges for hygienic treatments in the food industry, thereby increasing the risks of cross-contamination and foodborne illness outbreaks (Simões et al., 2010).

Vibrio parahaemolyticus is a marine oriented, gram-negative microorganism, that causes food poisoning through consumption of raw or undercooked seafood (Martinez-Urtaza et al., 2010). It is causing concerns across an expanding geographical range, with an increase in the frequency of infections and outbreaks in nations with no or limited prior occurrence (FAO, 2021). *V. parahaemolyticus* can enter and persist in seafood processing plants despite cleaning and sanitation (Lei et al., 2020; Roy et al., 2021; Wang et al., 2022). Sodium hypochlorite at concentrations that are effective against *V. parahaemolyticus* planktonic cells, can be ineffective at controlling biofilms. For instance, Chen et al. (2016) reported that 81 mg/L (pH 10.8, 3 min) of sodium hypochlorite eliminates 7.85 log CFU/mL of planktonic *V. parahaemolyticus* but 200 mg/L only reduced initial concentrations of 6.87 and 7.37 $\log_{10}$ CFU/cm^2 *V. parahaemolyticus* biofilm cells by 1.83 and 1.63 $\log_{10}$ CFU/cm^2 on

* Corresponding author.
E-mail address: s.h.flint@massey.ac.nz (S.H. Flint).

https://doi.org/10.1016/j.foodres.2023.112605
Received 30 October 2022; Received in revised form 4 February 2023; Accepted 14 February 2023
Available online 16 February 2023

shrimp and crab surfaces, respectively (Roy et al., 2021). The authors did not mention the pH or temperature of this treatment. Because normal hygienic treatment cannot eradicate *V. parahaemolyticus* biofilm, it is necessary to understand key molecular mechanisms and devise novel control strategies accordingly.

V. parahaemolyticus uses polar and lateral flagella for motility. Polar flagella (driven by sodium ions) are used for swimming, and lateral flagella (driven by protons) for swarming (Kim & McCarter, 2000; McCarter, 2001). The polar flagella function as a mechano-sensor, resulting in a reduction in flagellar rotation and activation of the lateral flagella expression in response to growth on a surface (Verstraeten et al., 2008). Mature *V. parahaemolyticus* biofilm formation requires mannose-sensitive hemagglutinin (MSHA) pili, virulence-associated toxin co-regulated pili (TCP) and chitin-regulated pili (ChiRP) to aggregate, attach and promote EPS synthesis. *cpsA-J* (*vpa1403-vpa1412*) is required for the synthesis of capsular polysaccharide A (CPSA), a major component of the *V. parahaemolyticus* biofilm (Güvener & McCarter, 2003). *vp1476-vp1458* is a conserved ortholog of the *syp* locus in *Vibrio fischeri* which is responsible for wrinkled colonies, pellicle formation and matrix production (Yildiz & Visick, 2009). *vp0190-vp0214* regulates the synthesis of lipid A, the core component of lipopolysaccharide (LPS) (Chen et al., 2010). *csrA* plays critical roles in carbohydrate metabolism and switching motility-sessility functions (Wang et al., 2021). However, it has not been determined which of these or other genotypes are responsible for significant biofilm formation and sodium hypochlorite resistance in this bacterium.

Therefore, the current study evaluated genomic features of *V. parahaemolyticus* with weak, intermediate, and strong biofilm forming abilities, conducted comparative genome analysis and deciphered the exclusive molecular mechanisms of persistent *V. parahaemolyticus* on seafood contact surface. This provides insights for the design of novel biofilm control strategies.

2. Methods

2.1. Strains and growth condition

Ten *V. parahaemolyticus* strains were chosen for this investigation (Table 1), seven of them were isolated from shellfish by Plant and Food Research ltd. (New Zealand), another two were clinical isolates provided by the Institute of Environmental Science and Research ltd. (ESR, New Zealand), and the *V. parahaemolyticus* genome reference strain RIMD 2210633 kindly donated by Dr. Tetsuya Iida was also included. According to previous studies, two strains (PFR30J09 and PFR34B02) were putative strong biofilm formers and least susceptible to chlorine treatment, whereas PFR21C03 and PFR37D08 were weak biofilm forming strains (Wang et al., 2023). Isolates from a −80 ℃ bead storage system were recovered by shaking incubation at 37 ℃, 120 rpm using 3 % NaCl tryptic soy broth (TSB). Cells were centrifuged (11200 × g, 5 min) to obtain the cell pellet. This was washed, suspended in sterile phosphate-buffered saline (PBS, pH 7.4) and the cell concentration adjusted to 7 log colony forming units (CFU)/mL.

2.2. Genome assembly and annotation

Sequencing of the PFR isolates (Table 1) was carried out by Cefas (Weymouth, U.K.). DNA was extracted from overnight grown cultures using the DNeasy Blood & Tissue Kit and the genomes were sequenced by MiSeq with a coverage of 40-120x. Raw reads were processed to trim adapters, as well as clean low-quality reads and low-quality sequences. *De novo* assembly using the assemblers SPAdes (version 3.15.2) and Velvet (version 1.2.10) were generated for each genome. SPAdes used k-mers (21, 33, 55) to build *de Brujin* graphs and generate contigs, the mismatch careful mode and error correction procedure was used to improve the assembly (Bankevich et al., 2012). *De novo* assembly was also performed via *de Brujin*-based Velvet and Velvet Optimizer, the k-mer sizes applied were of 31, 51, 71, 91, and 121 (Zerbino & Birney, 2008). The quality of the draft assembly was evaluated using QUAST and CheckM (Gurevich et al., 2013; Parks et al., 2015). The GC depth analysis was used to assess the potential contamination and the coverage of the assembly, and the completeness of genome assembly to examine the genome integrity. Predicted coding sequences from bacterial genome assemblies were generated by Prokka (version: 1.14.5) with default parameters, together with translated coding genes, genomic features and GFF version annotations (Seemann, 2014).

2.3. ANI and phylogeny analysis

Considering potentially bio-accumulated bacterium in molluscan shellfish and interfering seafood safety, published genomes of bacteria *Escherichia coli* K-12 ER3413, *Listeria monocytogenes* ATCC 19117, *V. parahaemolyticus* RIMD 2210633, *V. parahaemolyticus* BB22OP, *Vibrio cholerae* ATCC 14035, *Vibrio fischeri* ES114 and *Vibrio vulnificus* ATCC 27562 were included with those of our 9 isolates for ANI and phylogeny

Table 1
Strains used in this study.

No.	Strain	Source	Collection Date	Country	Biofilm formation abilities*	Sodium hypochlorite resistance (biofilms on stainless steel surfaces)**
1	PFR21C03	Pacific oyster	16/02/2009	New Zealand	W	+
2	PFR24B07	Greenshell™ mussel	2/03/2010	New Zealand	M	+
3	PFR29A04	Pacific oyster	22/11/2010	New Zealand	M	+
4	PFR30G02	Pacific oyster	8/03/2011	New Zealand	M	+
5	PFR30J09	Pacific oyster	21/03/2011	New Zealand	S	++
6	PFR34B02	Pacific oyster	27/03/2012	New Zealand	S	++
7	PFR37C06	Pacific oyster	17/01/2013	New Zealand	M	+
8	PFR37D08	Clinical	2013	New Zealand	W	+
9	PFR37E03	Clinical NZRM 3391	1975	New Zealand	M	+
10	RIMD2210633	Clinical	1996	Japan	M	+

* W indicated weak biofilm producer, M indicated moderate biofilm producer, S indicated strong biofilm producer.
** + indicated sodium hypochlorite resistance by *V. parahaemolyticus* biofilm cells, ++ indicated increased sodium hypochlorite resistance.

analysis, to obtain a more complete view. Average nucleotide identity (ANI) was analyzed using pyANI (version 0.2.12) to measure the similarity among genomes by pairwise comparison based on nucleotide sequences (Pritchard et al., 2016). To estimate the evolutionary relationship of *V. parahaemolyticus* genomes, phylogeny analysis was conducted based on 16 s rRNA and ortholog clustering. DNA sequences were loaded into Barrnap (https://github.com/tseemann/barrnap) that applies HMMER 3.1 to search for the locate ribosomal RNA genes. The 16 s rRNA sequences were collected and aligned via MEGAX (version 10.2.6) software. A Maximum Likelihood phylogenetic tree was constructed using the nearest neighbor interchange model. Ortholog clustering analysis was performed via OrthoFinder2 (version: 2.5.4) using protein-coding genes. The orthogroups had single-copy genes aligned via MAFFT. The phylogenetic tree based on ML of replicates was built by the IQ-TREE with 1000 bootstraps.

2.4. Homology clustering and pangenome analysis

Pangenome analysis to reveal core, dispensable and unique content was conducted using Roary and Anvi'o (Eren et al., 2015; Page et al., 2015). According to Roary (version 3.11.2), protein sequences with annotation were loaded and an all-against-all BLASTP was used to cluster proteins. The sequence identity of over 95 % was set as the threshold for clustering protein homologues. Anvi'o clustered homologues based on the similarity of amino acid sequences, the interactive visualization of results was generated by the anvio-display-pan function.

2.5. Comparative genome analysis

Four genomes, two from strong biofilm formers (PFR30J09 and PFR34B02) and two from weak biofilm formers (PFR21C03 and PFR37D08) were compared. Orthologous gene families and gene duplication events were analyzed via OrthoFinder2 and compared with a protein similarity search using DIAMOND (version 0.9.18). Another approach to predict orthologous gene clusters was performed using OrthoVenn2, which is based on all-to-all protein similarity comparisons and orthologous clusters using the Markov Cluster algorithm. The E-value and inflation values were set as 1e-5 and 1.5, respectively.

Functional assignment and the DAVID database (https://david.ncifcrf.gov/) was used to annotate gene sets for GO terms. DAVID enables visualization of many-genes-to-many-terms relationships and clusters genes into groups. This was followed by a Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis using BlastKOALA (https://www.kegg.jp/blastkoala/). A false discovery rate (FDR) value smaller than 0.05 was used as the cut-off for significance.

2.6. Identification of potential horizontal transfer genes

HGTector was used to predict the presence of horizontal gene transfer among genomes with BLASTP parameter thresholds of 60 % identity, 60 % coverage and an E-value of 1e-5. HGTector differentiates self (rank, genus; taxid: 662), close (rank, order; taxid: 135623) and distal gene groups by the gene hit bitscore (Zhu et al., 2014). Cut-offs were determined by taking the midpoint between the first peak and valley from the distribution of self, close and distal groups. HGT genes were identified with a close weight smaller than the close cut-off but a distal weight greater than the distal cut-off (Yang et al., 2019).

2.7. Congo red indicator assay

Expression of extracellular polymeric cellulose was evaluated using the Congo red indicator method reported by Fang et al. (2022) with minor modification. Overnight culture was streaked onto agar plates and incubated at 37 ℃ for 48 h. Colonies of red, brown, pink and white indicated the production of curli and cellulose, curli, cellulose, and none, respectively.

2.8. Anthrone absorbance assay

Cellulose determination was based on a previous study with minor modifications (Anriany et al., 2006). Briefly, 3 g cells (wet weight) of each sample were scratched from agar plates, mixed with 5 mL of an 8:2:1 acetic acid: nitric acid: distilled water mix, boiled for 30 min and centrifuged at 4480 × g for 5 min. The pellet was washed with 1 mL distilled water then 1 mL acetone and left to dry overnight. The dried sample was dissolved in 1 mL H_2SO_4 (95 %). Next, 0.1 mL was mixed with 0.5 mL anthrone solution (0.2 g in 100 mL H_2SO_4), and the absorbance was determined at 620 nm. Crystalline cellulose (25, 50 and 100 mg/mL) were used as absorbance standards.

2.9. In silico cellulose synthase operon sequence analysis

The cellulose synthase operon genes of *V. parahaemolyticus* were searched using BlastP against 138 *V. parahaemolyticus* reference genomes. The cellulose synthase genes of each genome were aligned using ClustalW. The accession numbers for nucleotide sequences of the genomes are attached in Table 4. The phylogeny of these cellulose synthase operons was analyzed using FastTree and was inferred using the Neighbour Joining method with a bootstrap consensus of 1000 replicates. The evolutionary distances were calculated using the Maximum Likelihood Method and are in the units of the number of base substitutions per site. The amino acid sequences of *V. parahaemolyticus* RIMD 2210633 were retrieved from the National Centre for Biotechnology Information (NCBI) Database and selected as the reference genome, followed by transferring into a local database for BLASTP searching via DIAMOND (Buchfink et al., 2021). Each sequenced *V. parahaemolyticus* genome was searched to identify the presence of genes (with > 70 % identity and > 70 % sequence coverage).

2.10. Gene co-occurrence analysis

Gene co-occurrence networks were computed using Coinfinder with a significance cut-off of 0.1 and gene presence/absence matrix from Roary. Networks were visualized using the R package igraph of Fruchterman Reingold.

2.11. Data availability

The sequence data determined in this work was deposited in the NCBI database under project accession No. PRJNA808748. The cellulose operon information for the strains used in this study are presented in Table 4.

3. Results and discussion

3.1. Genome assembly and annotation

The size of the *V. parahaemolyticus* genomes after trimming ranged from 5110607 bp to 7614640 bp, with an average value of 5532445 bp (Table 2). The GC contents of these genomes were not significantly different, ranging from 45.16 % to 45.47 %. Genome annotation was obtained from Prokka, which was to predict the coding of DNA sequences (CDS) in the assemblies. The number of CDS varied from 4591 to 4937. The variations in genome size and CDS number could be due to gene diversity during the evolution of *V. parahaemolyticus* isolates. The strain *V. parahaemolyticus* RIMD 2210633 isolated from Japan was used as the reference genome for most of our molecular analysis. This genome consisted of chromosomes of 5165770 bp, genes of 4832 bp and a GC content of 45.4 %. Overall, assembled genomes provided reasonable gene completeness and represented a reliable resource for analysis. In our study, the gene pools of the two clinical isolates were lower than the environmental (food) isolates; this was in accordance with reports by Pang et al. (2019) where there were 4718 genes on average from 19

Table 2
Summary of assemblies of *V. parahaemolyticus* genomes.

Assembly	PFR21C03	PFR24B07	PFR29A04	PFR30G02	PFR30J09	PFR34B02	PFR37C06	PFR37D08	PFR37E03	RIMD 2,210,633
Contigs (number)	93	133	129	99	329	104	95	85	110	–
Total length (bp)	5,338,398	5,417,990	5,402,653	5,172,378	7,816,420	5,367,801	5,218,301	5,149,233	5,110,607	5,165770
GC* (%)	45.27	45.16	45.19	45.27	45.47	45.19	45.25	45.24	45.27	45.4
CDS**	4830	4937	4905	4659	4913	4857	4725	4639	4591	4832
rRNA***	7	8	8	7	5	8	9	8	7	11
tRNA***	119	112	115	102	110	117	116	106	112	127
tmRNA***	1	1	1	1	1	1	1	1	1	1

* GC indicates Gunanine: Cytosine ratio.
** CDS indicates predicted coding sequences of DNA.
*** rRNA is ribosomal RNA, tRNA is transfer RNA and tmRNA is transfer-messenger RNA. - indicates not applicable, RIMD 2210633 is of scaffold assembled level.

V. parahaemolyticus environmental isolates compared with 4580 genes on average from 20 clinical isolates, suggesting higher genetic diversity and environmental adaptability in the environmental isolates.

3.2. ANI analysis and phylogenetic tree

ANI provides a commonly used metric to define intra- or inter-species relationships in prokaryotic genomes. An ANI of > 95 % has been proposed to represent the intra-species boundary (Jain et al., 2018). In this study, the dendrogram directly reflects the degree of identity between genomes; clustering across the dendrograms was based on ANI, forming one major clade of *V. parahaemolyticus* (Fig. 1a). The sequenced genomes revealed a high nucleotide identity (97.98–99.67 %), indicating they belong to the same species, *V. parahaemolyticus,* and are from a common origin. *V. parahaemolyticus* nucleotide identity compared with other *Vibrio* species was 70.98–76.15 %. The ANI across all *Vibrio spp*. ranges from 69.78 % to 76.15 %. The ANI result for *E. coli* K-12 ER3413 and *L. monocytogenes* ATCC 19117 is below 70 %, demonstrating they are of different genera.

Nucleotide acid sequences of 16 s rDNA have been used for decades to assign phylogenetic relationships (Fig. 1b). Among selected 16 s rDNA sequences of *E. coli* K-12 ER3413, *L. monocytogenes* ATCC 19117, *V. parahaemolyticus* RIMD 2210633, *V. parahaemolyticus* BB22OP, *Vibrio cholerae* ATCC 14035, *Vibrio fischeri* ES114 and *Vibrio vulnificus* ATCC 27562, our PFR species remained in one subclade with *V. parahaemolyticus* RIMD 2210633 and *V. parahaemolyticus* BB22OP, indicating that they are closely related. The most divergent species are *E. coli* K-12 ER3413 and *L. monocytogenes* ATCC 19117.

Another method to assign the phylogenetic relationship is based on the orthologous groups, using 275 conserved single copy orthogroups, seen in Fig. 1c. The result was consistent with the phylogeny analysis result based on 16 s rDNA similarity that *V. parahaemolyticus* was in the same clade with *V. parahaemolyticus* RIMD 2210633 and *V. parahaemolyticus* BB22OP, divergent from *E. coli* K-12 ER3413 and *L. monocytogenes* ATCC 19117. Among the subclade, PFR30J09 formed a different branch showing some deviation from the other isolates. These data provide an evolutionary picture of our sequenced *V. parahaemolyticus* genomes and reveal evolutionary distances from other studied seafood pathogens.

3.3. Pangenome analysis

The sizes of core and dispensable genomes were estimated using pangenome analysis via two distinct pipelines, Roary and Anvi'o. Roary analysis suggested there were 3253 core genes (33.20 %), 2266 shell genes (23.13 %), and 4278 cloud genes (43.67 %) out of 9797 genes in 9 sequenced *V. parahaemolyticus* genomes. The genome group is open, indicating that additional data input will alter the proportion of the core genome and that new orthogroups will be discovered (Supplemental Fig. 1). A circular graph was created via Anvi'o containing information on gene numbers in gene clusters, maximum number of paralogs, genomic homogeneity index, functional homogeneity index, combined homogeneity index, and single copy gene (SCG) clusters (Fig. 2a). Despite both Roary and Anvi'o approaches using the MCL algorithm to identify clusters, the Anvi'o pangenomics workflow identified fewer gene clusters (6845), which may have led to a smaller number of core genes. The differences might be due to different ways to establish orthologs of protein clusters. Roary divides groups of homologous sequences into paralogs and orthologs using conserved gene neighborhood information, while Anvi'o clusters orthologs based on the homology and synteny of genes (Maturana & Cárdenas, 2021). The core genome was a relatively low portion of the total (~33.20 %) which was revealed by Roary analysis, suggesting that the *V. parahaemolyticus* species may contain more accessory genes that are critical for adaptation to different environments and survival. A similar finding has been reported in previous *V. parahaemolyticus* studies by Pang et al. (2019) and Qin et al. (2021).

3.4. Comparative genome analysis

The strong biofilm-forming strains of *V. parahaemolyticus* (PFR30J09 and PFR34B02) both had more gene families than the weak biofilm-forming isolates (PFR21C03 and PFR37D08), but medium biofilm forming isolates covered the range of gene families (CDS, Table 2). Analysis of the amino acid sequences by OrthoVenn2 showed a total of 3692 gene families to be shared across the genomes of the four strong and weak biofilm forming isolates (Fig. 2b). OrthoVenn2 identified 136 gene families exclusively present in strong biofilm producers.

The GO annotation was performed to assign the functional category of these 136 unique gene families. The significantly enriched GO terms ($p < 0.05$) included UDP-glucose metabolism (GO: 0006011), cellulose biosynthesis (GO: 0030244), rhamnose metabolism (GO: 0019299) and O antigen biosynthesis (GO: 0009243) (Supplemental Table 1). Cellulose synthesis, rhamnose catabolic and metabolic processes are absent in the *V. parahaemolyticus* reference strain RIMD 2210633. These distinctive functions may play crucial roles in robust biofilm formation. BlastKOALA was used to assign the amino acid sequences to KEGG functions, resulting in 24.3 percent (33/136) of them being successfully annotated (Table 3), while many of them remained unidentified. The most frequently annotated pathways were metabolic pathways (path: ko01100), microbial metabolism in diverse environments (path: ko01120), pentose and glucuronate interconversions (path: ko00040), fructose and mannose metabolism (path: ko00051), biosynthesis of secondary metabolites (path: ko01110), galactose metabolism (path: ko00052) as well as starch and sucrose metabolism (path: ko00500). These pathways represent the synthesis and metabolism of simple/complex sugars, which may aid in the utilization of diverse nutrient sources and adaptation to environmental stress.

Gene clusters associated with cellulose synthesis, rhamnose catabolism, and metabolic process were identified via KEGG gene annotation (Table 3), that was consistent with the result from GO enrichment. Meanwhile, the clustered regularly interspaced short palindromic repeats (CRISPR)-Cas system was identified in the unique 136 gene families via KEGG annotation. This system has been reported to play critical

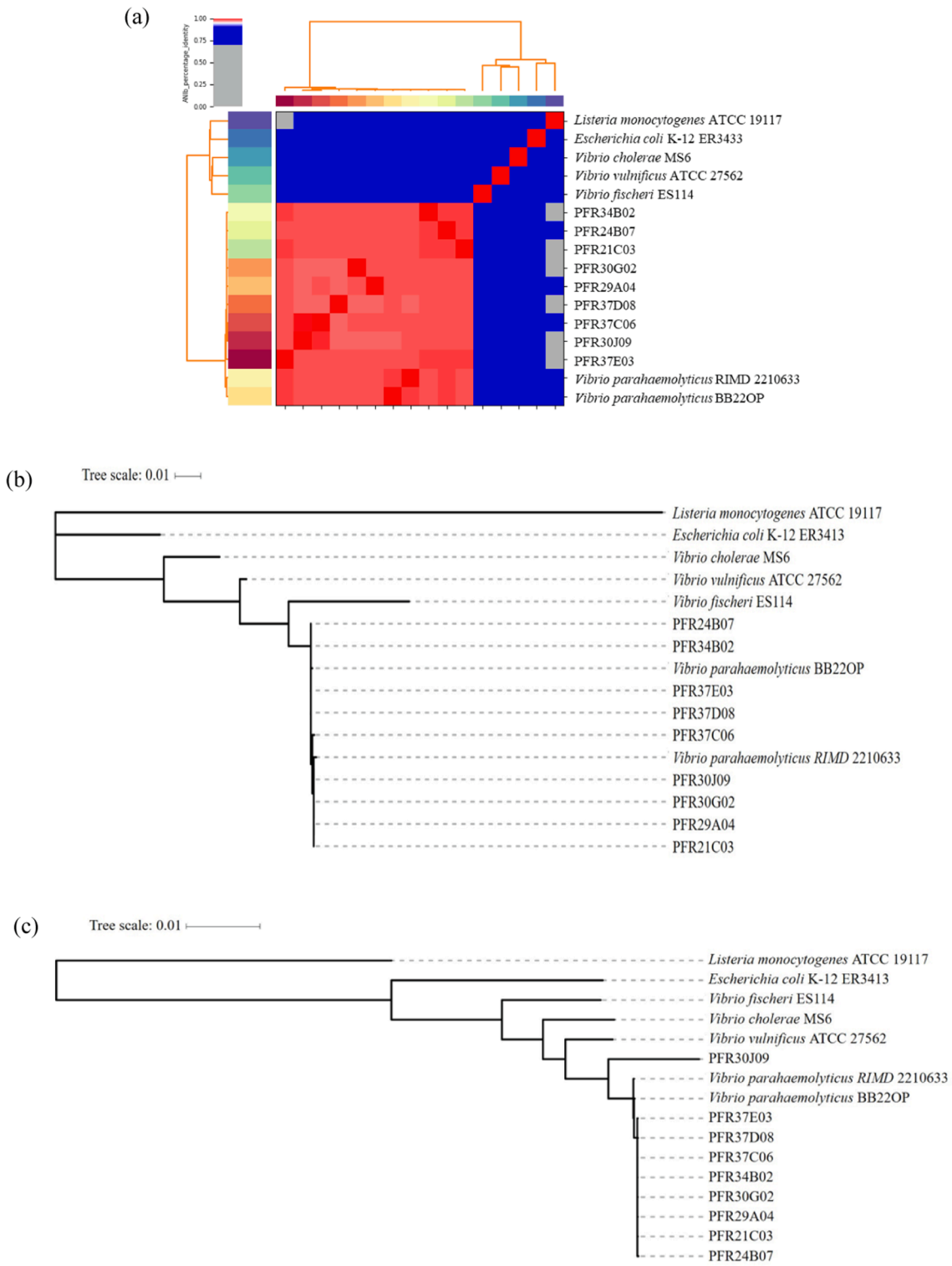

Fig. 1. Phylogenetic analysis and amino acid identities across genomes. (a) Phylogeny based on a heatmap with row and column dendrograms from the average nucleotide identity (ANI). (b) The Maximum Likelihood phylogeny was constructed using the 16 s rDNA sequences. (c) Maximum Composite Likelihood phylogeny was constructed based on single copy orthogroups.

roles in environmental stress defense, DNA repair and biofilm formation in microorganism cells. In this study, the CRISPR-Cas system in *V. parahaemolyticus* was identified as the subtype I-F, that is consistent with previous studies (Makarova et al., 2011). However, whether or how the CRISPR-Cas system promotes biofilm formation in *V. parahaemolyticus* is unknown. The genome data also indicated that strong biofilm forming strains are more likely to have *mshA*, *mshC* and *mshD* genes that are lacking in the weak biofilm forming strains. MSHA is required for early attachment of *V. cholerae* to abiotic surfaces. Cells that are deficient in *mshA* cannot perform cell aggregation although they are involved in the formation of three-dimensional structures (Moorthy and Watnick, 2004) and in *V. parahaemolyticus*, *mshA* mutants show reduced adherence to surfaces (Shime-Hattori et al., 2006).

Our study supports the hypothesis that *V. parahaemolyticus* employs

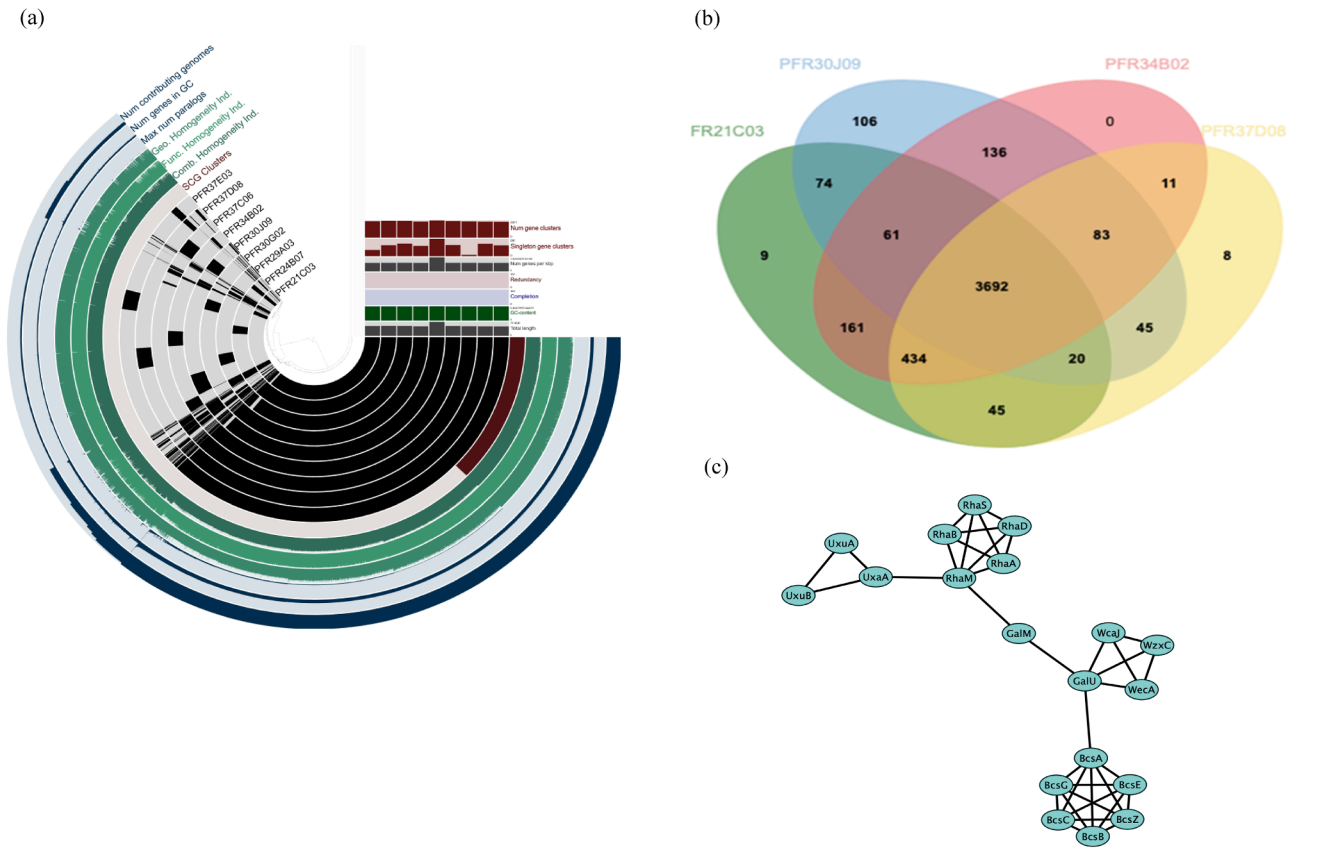

Fig. 2. Genomes of *Vibrio parahaemolyticus* exhibit genetic diversity. (a) Strain-specific gene families of nine *V. parahaemolyticus* genomes computed by the Anvi'o pangenome analysis pipeline employing the default setting. (b) A comparison of protein orthologues in weak and strong biofilm-forming *V. parahaemolyticus*, generated by OrthoVenn2 with an E-value setting of 1e-5. (c) PPI network of 136 gene families exclusively present in strong biofilm-forming *V. parahaemolyticus*.

multiple strategies to promote robust biofilm formation, and some of these strategies interact. In the PPI (protein–protein interaction) analysis of 136 gene families, an interaction between rhamnose metabolism, cellulose biosynthesis, as well as O-antigen and capsular biosynthesis were demonstrated (Fig. 2c). Regmi and Boyd (2019) reported that the *rha* gene cluster, encoding rhamnose catabolic and metabolic processes, had limited distribution among *V. parahaemolyticus* strains. Another study reported that the presence of the rhamnose operon is indispensable but not sufficient for biofilm formation in *Rhodobacteraceae* (Michael et al., 2016), but scenario varied here. In this study, the presence of the *rha* gene cluster might have played as one 'biobrick' for robust biofilm formation, but the absence did not interfere with the ability to form biofilms as presented in PFR21C03 and PFR37D08. The presence of cellulose synthase operons suggests that cellulose production is a phenotype of robust biofilm forming *V. parahaemolyticus* strains, but until now, cellulose has only been reported in environmental isolates of *V. parahaemolyticus* (Meparambu Prabhakaran et al., 2022). Given cellulose roles in virulence in other microorganisms (Arrebola et al., 2015; Heredia-Ponce et al., 2020), cellulose could be a neglected virulence factor that has previously been considered absent in *V. parahaemolyticus*. GalU is required to encode UDP-glucose pyrophosphorylase and produce a single substrate UDP-glucose for cellulose synthases. GalU is essential for cellulose secretion *in vivo* (Abidi et al., 2021; Valla et al., 1989). The subcluster encompassing WcaJ, WecA and WzxC was suggested to be responsible for the O antigen (Pang et al., 2019), as shown in Fig. 2. The PPI network revealed an indirect link between cellulose biosynthesis and O-antigen via GalU in strong biofilm forming strains of *V. parahaemolyticus*. L-Rhamnose is required for rhamnan O antigen synthesis (Broughton et al., 2006), which is to some extent in agreement with that *rha* gene cluster associated with O-antigen formation in this study. The variation in O-antigen can result in the structural diversity of O-antigen and linkages between monosaccharides, that in turn influence LPS and biofilm formation (Lerouge & Vanderleyden, 2002). Based on our study, cellulose production might relate to a specific type O-antigen in *V. parahaemolyticus* biofilms. However, whether cellulose biosynthesis is related with a specific O-antigen in other strong biofilm forming strains and how the O-antigen assist in the robust biofilm formation require further exploration in *V. parahaemolyticus*, although some evidence has been demonstrated by Lerouge and Vanderleyden (2002). To sum up, this is the first time that an interaction between rhamnose metabolic and catabolic processes, cellulose biosynthesis, and O-antigen biosynthesis pathways has been linked to robust biofilm formation. Biochemical studies are needed to determine whether this interaction is specific to the strains studied or more widespread in *V. parahaemolyticus* species.

3.5. Verification of cellulose production

Strong biofilm forming strains PFR30J09 and PFR34B02 produced red colonies in the Congo Red Assay, indicating the production of curli and cellulose (Fig. 3a). The weak biofilm forming strains, PFR21C03 and PFR37D08, only produced brown colonies, suggesting no cellulose was produced. The absence of cellulose biosynthesis in PFR21C03 and PFR37D08 is consistent with the results from the anthrone absorbance assay indicating no cellulose formation in PFR21C03 and PFR37D08 while cellulose was detected in PFR30J09 and PFR34B02 with OD_{620} of 0.234 ± 0.0356 and 0.185 ± 0.0546 (Fig. 3b), respectively. Crystalline cellulose was used as a control with 25 mg/mL producing an OD_{620} of 0.235 ± 0.0092.

Table 3
KEGG functional assignment of unique genes from strong biofilm forming strains.

KO	Gene name	EC number	Pathways/associated functions
K02851	*wecA*, UDP-GlcNAc: undecaprenyl-phosphate/ decaprenyl-phosphate GlcNAc-1-phosphate transferase	2.7.8.35	00542 O-Antigen repeat unit biosynthesis 00572 Arabinogalactan biosynthesis - Mycobacterium 01100 Metabolic pathways
K01785	*alM*, aldose 1-epimerase	5.1.3.3	00010 Glycolysis / Gluconeogenesis 00052 Galactose metabolism 01100 Metabolic pathways 01110 Biosynthesis of secondary metabolites 01120 Microbial metabolism in diverse environments
K00694	*bcsA*, cellulose synthase	2.4.1.12	00500 Starch and sucrose metabolism Cellulose biosynthesis
K01119	*cpdB*, 2′,3′-cyclic-nucleotide 2′-phosphodiesterase / 3′-nucleotidase	3.1.4.16/ 3.1.3.6	00230 Purine metabolism 00240 Pyrimidine metabolism 01100 Metabolic pathways
K00008	*gutB*, L-iditol 2-dehydrogenase	1.1.1.14	00040 Pentose and glucuronate interconversions 00051 Fructose and mannose metabolism 01100 Metabolic pathways
K02554	*mhpD*, 2-keto-4-pentenoate hydratase	4.2.1.80	00360 Phenylalanine metabolism 00362 Benzoate degradation 00621 Dioxin degradation 00622 Xylene degradation 01100 Metabolic pathways 01120 Microbial metabolism in diverse environments 01220 Degradation of aromatic compounds
K01813	*rhaA*, L-rhamnose isomerase	5.3.1.14	00051 Fructose and mannose metabolism 01100 Metabolic pathways 01120 Microbial metabolism in diverse environments
K00848	*rhaB*, rhamnulokinase	2.7.1.5	00040 Pentose and glucuronate interconversions 00051 Fructose and mannose metabolism 01100 Metabolic pathways 01120 Microbial metabolism in diverse environments
K01629	*rhaD*, rhamnulose-1-phosphate aldolase	4.1.2.19	00040 Pentose and glucuronate interconversions 00051 Fructose and mannose metabolism 01100 Metabolic pathways 01120 Microbial metabolism in diverse environments
K00963	UTP glucose-1-phosphate uridylyltransferase	2.7.7.9	00040 Pentose and glucuronate interconversions 00052 Galactose metabolism 00500 Starch and sucrose metabolism 00520 Amino sugar and nucleotide sugar metabolism 00541 O-Antigen nucleotide sugar biosynthesis 01100 Metabolic pathways 01110 Biosynthesis of secondary metabolites 01240 Biosynthesis of cofactors 01250 Biosynthesis of nucleotid8e sugars
K00040	*uxuB*, fructuronate reductase	1.1.1.57	00040 Pentose and glucuronate interconversions 01100 Metabolic pathways Translation
K16694	*tuaB*; teichuronic acid exporter	–	Transporters
K08138	*xylE*; MFS transporter, SP family, xylose:H + symportor	–	Transporters
K03765	*cadC*; transcriptional activator of cad operon	–	Transcription factors
K19130	*csy4*, CRISPR-associated endonuclease Csy4	3.1.-.-	Prokaryotic defense system
K19129	*csy3*, CRISPR-associated endonuclease Csy3	–	Prokaryotic defense system
K19128	*csy2*, CRISPR-associated endonuclease Csy2	–	Prokaryotic defense system
K19127	*csy1*, CRISPR-associated endonuclease Csy1	–	Prokaryotic defense system
K07012	*cas3*; CRISPR-associated endonuclease/helicase Cas3	3.1.-.- /5.6.2.4	Prokaryotic defense system
K15342	*cas1*; CRISP-associated protein Cas1	–	Prokaryotic defense system
K03534	*rhaM;* L-rhamnose mutarotase	5.1.3.32	Fructose and mannose metabolism Metabolic pathways Microbial metabolism in diverse environments
K02855	*rhaS*; AraC family transcriptional regulator, L-rhamnose operon regulatory protein RhaS	–	Fructose and mannose metabolism Metabolic pathways Microbial metabolism in diverse environments
K07733	*alpA*; prophage regulatory protein	–	Transcription factors
K03630	*radC*; DNA repair protein RadC	–	Replication and repair
K06877	DEAD/DEAH box helicase domain-containing protein	–	Poorly characterized
K03574	*mutT*, NUDT15, MTH2; 8-oxo-dGTP diphosphatase	3.6.1.55	DNA repair and recombination proteins
K20543	*bcsC*; cellulose synthase operon protein C	–	Cellulose biosynthesis
K20542	*bcsZ*; endoglucanase	3.2.1.4	Cellulose biosynthesis
K20541	*bcsB*; cellulose synthase operon protein B	–	Cellulose biosynthesis

3.6. Putative horizontal gene transfer

Horizontal gene transfer is the major driver of genetic diversity, providing bacteria with properties that enable adaption to various environments. An average of 126 ± 4.2 and 156 ± 45.3 potential HGT gene families providing gene acquisition from outside the *Vibrionales* were identified in weak and strong biofilm forming strains respectively. There were 123 and 129 predicted HGT events in PFR21C03 and PFR37D08 whereas 188 and 124 were predicted in PFR30J09 and PFR34B02. The results indicate that strong biofilm forming strains have

Table 4
Summary of the cellulose synthase operon in *V. parahaemolyticus*.

Genome	Strain	Nation	BioSample ID	Source	TLH	TRH	TDH	VgrG_T6SS	Hcp_T6SS
GCF_001304775.1	FORC_006	South Korea: Gyeongnam	SAMN03140318	Environment: cutting board	+	–	–	+	+
GCF_006517795.1	Vb0624	China: Shenzhen	SAMN12123413	Environment: market	+	–	–	+	+
GCF_002504185.1	HA2	China:Tianjin	SAMN07680340	Environment: aquaculture	+	–	–	+	+
GCF_004194515.1	D3112	China	SAMN10591529	Environment: seawater	+	–	–	+	+
GCF_001758605.1	FORC_023	South Korea: Pusan	SAMN03701448	Human	+	–	–	+	+
GCF_002209725.2	MAVP-Q	USA	SAMN05579852	Human	+	+	+	+	+
GCF_016403045.1	81TDH2	India: Mangaluru	SAMN16844329	Seafood	+	–	–	+	+
GCF_013393845.1	LVP1	China	SAMN11579495	Seafood: crayfish	+	–	–	+	+
GCF_000430405.1	FDA_R31	USA:LA	SAMN02179882	Seafood: oyster	+	–	+	+	+
GCF_016834555.1	HP1	India: Kumta	SAMN16844529	Seafood: shrimp	+	–	–	+	+
GCF_021730085.1	VP157	China:Tianjin	SAMN17188296	Seafood: shrimp	+	–	–	+	+
GCF_001700835.1	CHN25	China	SAMN03325855	Seafood: shrimp	+	–	–	+	+
GCF_001636035.1	FORC_014	South Korea: Pusan	SAMN03457164	Seafood: toothfish	+	–	–	+	+
GCF_012274985.1	2012 V-1165	USA	SAMN12648280	Stool	+	+	–	+	+
GCF_001682175.1	MAVP-Q	USA	SAMN03766034	Stool	+	+	+	+	+
GCF_002220985.3	MAVP-R	USA	SAMN06042545	Stool	+	+	–	+	+
GCF_009763565.1	2013 V-1181	USA	SAMN12648285	Stool	+	+	+	+	+
GCF_000568495.1	UCM-V493	Unknown	SAMN03081521	Unknown	+	–	–	+	+
GCF_003612715.1	FORC_071	South Korea: Seoul	SAMN07629009	Unknown	+	–	–	+	+
GCF_012274865.1	AM51557	USA	SAMN12648303	Unknown	+	–	+	+	+
GCF_003612695.1	FORC_072	South Korea: Seoul	SAMN07629020	Unknown	+	–	–	+	+
GCF_009764055.1	2012AW-0224	USA	SAMN12648278	Unknown	+	–	–	+	+
PFR30G02	PFR21C03	New Zealand	–	Seafood: oyster	+	–	–	+	+
PFR30J09	PFR30J09	New Zealand	–	Seafood: oyster	+	–	–	+	+
PFR34B02	PFR34B02	New Zealand	–	Seafood: oyster	+	–	–	+	+
PFR37C06	PFR37D08	New Zealand	–	Human	+	–	–	+	+

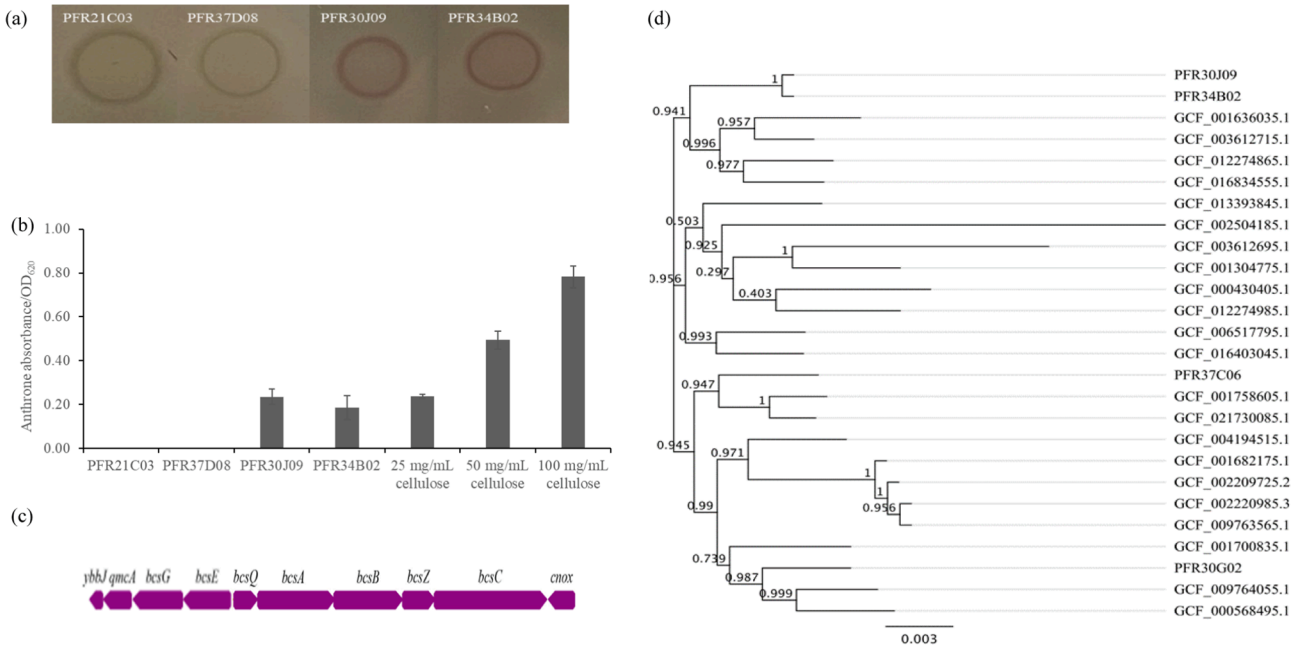

Fig. 3. Cellulose production at the phenotypic and genomic level in *V. parahaemolyticus*. (a) Colony morphotypes on CRI agar plates indicating cellulose/curli production. (b) Anthrone absorbance assay to determine cellulose production in *V. parahaemolyticus*, crystalline cellulose (CC) chemical was assayed as the control. (c) Diagram of cellulose synthase operon in *V. parahaemolyticus*, consisting of *bcsG*, *bcsE*, *bcsQ*, *bcsA*, *bcsB*, *bcsZ*, and *bcsC*. (d) Evolutionary relationships between cellulose synthase operons in *V. parahaemolyticus*. The phylogenetic tree was inferred using the Neighbor-joining method and FastTree.

more putative horizontal gene transfer events than the weak biofilm forming strains. This may relate to better adaption to diverse environments through biofilm formation. The phylum of *Proteobacteria* (taxid: 1224) and *Gammaproteobacteria* (taxid: 1236) were suggested to be the main donor taxa, providing 39.5 ± 3.54 and 51 ± 21.21, 52.5 ± 0.71 and 60 ± 9.90 HGT genes among weak and strong biofilm forming strains, respectively. This observation is not surprising as the aquatic environments are preferred by *Proteobacteria* and *Gammaproteobacteria*, providing many opportunities for recombination (Koblížek, 2015). The unique gene clusters of rhamnose metabolism and the CRISPR-Cas system present in strong biofilm forming strains appear to be obtained by HGT events resulting in the increased gene diversity in these strains.

Although these strategies were absent in weak biofilm-forming strains of *V. parahaemolyticus*, they are identified not only in strong biofilm-forming strains but also in part of the intermediate biofilm former (Data not shown). Interestingly, the cellulose synthase operon was not included in the putative HGT event results, indicating that the cellulose synthase function present in *V. parahaemolyticus* came from within the *Vibrionales* order. In our study, the *bcs* gene cluster was only present in environmental isolates, and whether cellulose only exists in environmental strains globally requires further study.

3.7. Prevalence of the cellulose synthase operon in V. Parahaemolyticus genomes

The cellulose synthase operon (*bsc*) was present in 22 out of 138 *V. parahaemolyticus* reference genomes from the NCBI dataset. The cellulose synthase operons of 22 *V. parahaemolyticus* genomes were aligned with one another, revealing a cellulose synthase operon conservatism. Fig. 3c is a diagram of the *bcs* locus and the transcription direction of cellulose synthase genes. The cellulose biosynthetic cluster was inserted on chromosome 1 between the *ybbJ* and *cnoX* genes, except for strain *V. parahaemolyticus* HA2 accommodating another two genes of the helix-turn-helix (HTH) domain containing gene and the IS630 family gene (RefGenome: GCF_002504185.1).

The cellulose biosynthesis locus was initially assumed to be absent from other *Vibrio* species (Yildiz & Visick, 2009). Although the genomic analysis indicated that the cellulose synthase genes were highly conserved in *V. parahaemolyticus*, the evolutionary analysis demonstrated that these genes were not 100 % identical, suggesting that the *bcs* operon might be useful in distinguishing *V. parahaemolyticus* isolates. Interestingly here, the strong biofilm forming strains of PFR30J09 and PFR34B02 were clustered into a subgroup based on the cellulose operon (Fig. 3d). We also examined whether the *bcs* gene cluster was restricted to environmental isolates in the reference genomes. The results indicated that cellulose synthase operons were present in both environmental (73.0 %, 19/26) and pathogenic (26.92 %, 7/26) strains.

The *bcs* operon responsible for cellulose production has been proposed as useful in the classification and standardized nomenclature in *Proteobacteria* (Römling & Galperin, 2015). The distinctions of the *bcs* gene profiles from *Proteobacteria* are primarily from three areas: 1) the presence of the *bcsD* gene, 2) the presence of the *bcsE* and *bcsG* genes (and the absence of *bcsD*) and 3) the absence of *bcsD*, *bcsE*, or *bcsG* (Abidi et al., 2021; Römling & Galperin, 2015). The putative operon identified in *V. parahaemolyticus* has an organization similar to *E. coli*-like -*bcs* operons, that have been found in *E. coli*, *Salmonella*, *Yersinia enterocolitica*, *V. fischeri* and *Pseudomonas*. Cellulose is considered as one of the major causes of robust biofilm and strong resistance in *Salmonella*, *Cronobacter* and *E. coli* in the food industry (Hu et al., 2015; Kim, Jyung, & Kang, 2022; Macarisin, Patel, Bauchan, Giron, & Sharma, 2012). Cellulose plays roles as an architectural element in biofilm matrices, providing advantages in water retention, porosity, mechanical resistance, low antigenicity, and interaction with saccharidic and proteinaceous components of both bacteria and hosts. This enables higher sanitizer resistance in biofilms. Solano et al. (2002) compared the influence of 30 ppm of sodium hypochlorite on the survival of biofilms of wild-type *Salmonella* Enteritidis and cellulose mutants formed on glass. Survival of 75 % of wild-type cells following 20 min exposure to the disinfectant contrasts with only 0.3 % survival of the cellulose-deficient mutant cells surviving, which clearly indicates the protective function of cellulose.

In *E. coli*-like *bcs* operon, BcsA and BcsB have been characterized as the enzymatic core of cellulose synthase, which is crucial for *in vitro* cellulose synthesis (Omadjela et al., 2013). BcsA provides glucosyltransferase activity that is allosterically triggered by the second messenger c-di-GMP and that, in conjunction with the BcsB subunit, produces a transmembrane channel for co-synthetic secretion of cellulose (Serra & Hengge, 2019). BcsE and BcsG subunits, produce a cytosolic c-di-GMP-binding protein and a membrane-anchored periplasmic pEtN transferase, respectively, and are the distinctive components of this type of *bcs* operon. The proteins mentioned above and c-di-GMP have been considered as molecular targets to interfere cellulose biosynthesis (Abidi et al., 2021), however, more exploration is required in *V. parahaemolyticus* species. Alternative carbon sources could be considered as an approach to inhibit cellulose production in food environments. Zhong et al. (2013) reported different carbon sources resulted in different bacterial cellulose production in *Gluconacetobacter xylinus*. Highest yields of cellulose were obtained when *G. xylinus* was cultured in glucose, fructose and glycerol, whereas lowest was obtained in those using inositol, sucrose and lactose as the medium. Carbon sources influence mechanical and microstructural characteristics of cellulose, as glycerol contributed to the highest cellulose yield in *G. xylinus*, as well as highest tensile strength of cellulose with thinner fibres and lower porosity.

A previous study identified epistatic interaction in the flexible genome between gene clusters for T6SS1 and cellulose biosynthesis, while clinical strains tend to depend on the antibacterial activity of T6SS1 proteins for competitive survival in the aquatic milieu and the strains that lack T6SS depend on cellulose production (Cui et al., 2020). However, in our genomic analysis, cellulose production appeared to be associated with the presence of T6SS that was characterized by both VgrG and Hcp (Table 4), regardless of whether the strains were clinical or environmental. The reason for this discrepancy could be that different clonal complexes of *V. parahaemolyticus* genomes used in the present study differ from those used in the earlier study. In our study, gene co-occurrence analysis identified that there were 468 co-occurring genes along with *bcsA*, notably, the CRISPR-Cas defense system (*cas*, *cys* gene families), the type II secretion system (*eps* gene family), the rhamnose metabolic process (*rha* gene family) and capsular polysaccharide synthesis (*wec* gene family). This was consistent with the results from PPI analysis for the two strong biofilm-forming *V. parahaemolyticus* (PFR30J09 and PFR34B02). Symbiosis polysaccharide protein SypF has been identified in association with cellulose biosynthesis and production in *V. fischeri* via the polysaccharide biosynthesis protein VpsR, indicating that cellulose contributes to symbiotic initiation by promoting biofilm formation on the symbiotic organ's surface (Darnell et al., 2008). Here, the interaction between the VpsR homologue CpsR and cellulose production was also identified in *V. parahaemolyticus*, reported for the first time in this study (Fig. 4). Whether cellulose is important in symbiosis in *V. parahaemolyticus* requires further investigation.

4. Conclusion and perspectives

The expansion of gene families is a common strategy whereby bacteria cope with diverse environments. In this study, strong biofilm forming strains appeared to have higher genetic diversity compared with the weak biofilm forming strains. In *V. parahaemolyticus*, the low portion of core genome size (~33.20 %) allows for more accessory genes that are critical for survival in multiple environments. We identified the phyla of *Proteobacteria* (taxid: 1224) and *Gammaproteobacteria* (taxid: 1236) act as the main donor taxa for the genetic diversity of strong biofilm forming strains. Genes that are more abundant in strong biofilm forming genomes were associated with cellulose biosynthesis, rhamnose and other sugar metabolism, O-antigen biosynthesis, MSHA pili led attachment as well as the CRISPR-Cas defence system. The genes of rhamnose metabolism and catabolismand the CRISPR-Cas defence system was most likely from HGT with cellulose biosynthesis being acquired within the order of *Vibrionales*. The PPI analysis demonstrated an interaction between rhamnose metabolism, cellulose production, and O-antigen biosynthesis. This study is limited to the contrast between high and low biofilm producers to highlight the key factors involved. This is a limitation of the study as it does not include the intermediate biofilm forming strains. Future studies will look at gene expression to provide some quantitative information that will cover the intermediate biofilm

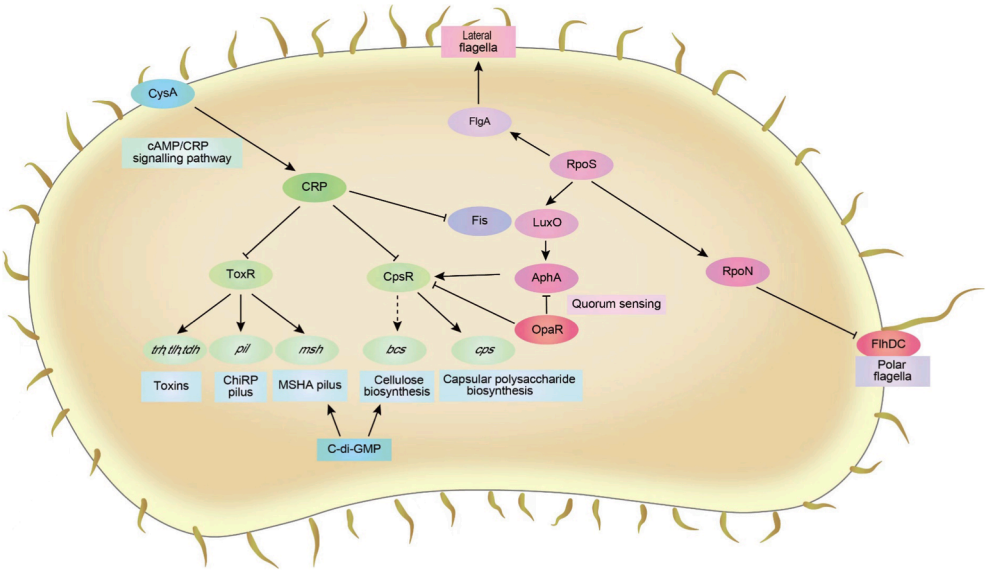

Fig. 4. Molecular mechanisms of biofilm formation in *V. parahaemolyticus*. The known molecular mechanism was based on Abidi et al., (2021) and Guo et al., (2020), presented using the solid line; the dash line represents the finding in this study.

forming strains.

We believe this study is the first to reveal the relationship between cellulose secretion and robust biofilm formation in *V. parahaemolyticus*. The cellulose synthase operons in *V. parahaemolyticus* were examined for their prevalence (22/138, 15.94 %) and were found to consist of the genes *bcsG*, *bcsE*, *bcsQ*, *bcsA*, *bcsB*, *bcsZ*, *bcsC*. The precise biochemical functions and chemical structures of the cellulose component in biofilm matrices require further identification. The chemical elements of cellulose vary slightly between species with distinct *bcs* operon structures. For example, in *E. coli*, the cellulose in the biofilm matrix has been identified as phosphoethanolamined, whereas in *Pseudomonas*, the cellulose is amorphous. The chemical structure of cellulose in *V. parahaemolyticus* and how it interacts with other biofilm matrix components is unknown. Robust biofilm formation is likely to protect the organism from sanitisers, creating higher risks of cross contamination and subsequent foodborne illness. Understanding cellulose in the biofilm matrix will help develop novel and effective biofilm control measures.

CRediT authorship contribution statement

Dan Wang: Conceptualization, Methodology, Investigation, Formal analysis, Writing – original draft. **Graham C. Fletcher:** Resources, Conceptualization, Methodology, Supervision, Writing – review & editing. **Dragana Gagic:** Conceptualization, Methodology, Supervision, Writing – review & editing. **Stephen L.W. On:** Conceptualization, Methodology, Supervision, Writing – review & editing. **Jon S. Palmer:** Conceptualization, Methodology, Supervision, Writing – review & editing. **Steve H. Flint:** Project administration, Resources, Supervision, Conceptualization, Methodology, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

Dr. Tetsuya Iida (Research Institute for Microbial Diseases, Osaka, Japan) kindly supplied the culture of genome reference strain RIMD 2210633. The authors are grateful to Jaime Martinez-Urtaza and Andy Powell of Centre for Environment, Fisheries, and Aquaculture Science (Cefas) who provided sequence data.

Funding

This study is funded by Massey University, the grant number is P963141222WangD.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodres.2023.112605.

References

Abidi, W., Torres-Sánchez, L., Siroy, A., & Krasteva, P. V. (2021). Weaving of bacterial cellulose by the Bcs secretion systems. *FEMS Microbiology Reviews, 46*(2). https://doi.org/10.1093/femsre/fuab051

Anriany, Y., Sahu, S. N., Wessels, K. R., McCann, L. M., & Joseph, S. W. (2006). Alteration of the rugose phenotype in *waaG* and *ddhC* mutants of *Salmonella enterica* serovar Typhimurium DT104 is associated with inverse production of curli and cellulose. *Applied and Environmental Microbiology, 72*(7), 5002–5012.

Arrebola, E., Carrión, V. J., Gutiérrez-Barranquero, J. A., Pérez-García, A., Rodríguez-Palenzuela, P., Cazorla, F. M., & de Vicente, A. (2015). Cellulose production in *Pseudomonas syringae* pv. *syringae*: A compromise between epiphytic and pathogenic lifestyles. *FEMS Microbiology Ecology, 91*(7). https://doi.org/10.1093/femsec/fiv071

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., … Prjibelski, A. D. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology, 19*(5), 455–477.

Broughton, W., Hanin, M., Relic, B., Kopcińska, J., Golinowski, W., Simsek, S., … Kobayashi, H. (2006). Flavonoid-inducible modifications to rhamnan O antigens are necessary for *Rhizobium sp.* strain NGR234-legume symbioses. *Journal of Bacteriology, 188*(10), 3654–3663.

Buchfink, B., Reuter, K., & Drost, H. G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods, 18*(4), 366–368. https://doi.org/10.1038/s41592-021-01101-x

Chen, T.-Y., Kuo, S.-H., Chen, S.-T., & Hwang, D.-F. (2016). Differential proteomics to explore the inhibitory effects of acidic, slightly acidic electrolysed water and sodium hypochlorite solution on *Vibrio parahaemolyticus*. *Food Chemistry, 194*, 529–537. https://doi.org/10.1016/j.foodchem.2015.08.019

Chen, Y., Dai, J., Morris, J. G., & Johnson, J. A. (2010). Genetic analysis of the capsule polysaccharide (K antigen) and exopolysaccharide genes in pandemic *Vibrio*

parahaemolyticus O3:K6. *BMC Microbiology, 10*(1), 274. https://doi.org/10.1186/1471-2180-10-274
Cui, Y., Yang, C., Qiu, H., Wang, H., Yang, R., & Falush, D. (2020). The landscape of coadaptation in *Vibrio parahaemolyticus*. *Elife, 9*, e54136.
Darnell, C. L., Hussa, E. A., & Visick, K. L. (2008). The putative hybrid sensor kinase SypF coordinates biofilm formation in *Vibrio fischeri* by acting upstream of two response regulators, SypG and VpsR. *Journal of Bacteriology, 190*(14), 4941–4950.
Eren, A. M., Esen, Ö. C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., & Delmont, T. O. (2015). Anvi'o: An advanced analysis and visualization platform for 'omics data. *PeerJ, 3*, e1319.
Fang, Y., Visvalingam, J., Zhang, P., & Yang, X. (2022). Biofilm formation by Non-O157 Shiga toxin-producing *Escherichia coli* in monocultures and co-cultures with meat processing surface bacteria. *Food Microbiology, 102*, Article 103902. https://doi.org/10.1016/j.fm.2021.103902
FAO. (2021). Advances in science and risk assessment tools for *Vibrio parahaemolyticus* and *V. vulnificus* associated with seafood: meeting report (9240024875). Retrieved from https://www.who.int/publications/i/item/9789240024878. Accessed October 6, 2022.
Güvener, Z. T., & McCarter, L. L. (2003). Multiple regulators control capsular polysaccharide production in *Vibrio parahaemolyticus*. *Journal of Bacteriology, 185*(18), 5431–5441.
Guo, L., Wang, J., Gou, Y., Tan, L., Liu, H., Pan, Y., & Zhao, Y. (2020). Comparative proteomics reveals stress responses of *Vibrio parahaemolyticus* biofilm on different surfaces: Internal adaptation and external adjustment. *Science of The Total Environment, 731*, Article 138386. https://doi.org/10.1016/j.scitotenv.2020.138386
Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics, 29*(8), 1072–1075.
Heredia-Ponce, Z., Gutiérrez-Barranquero, J. A., Purtschert-Montenegro, G., Eberl, L., Cazorla, F. M., & de Vicente, A. (2020). Biological role of EPS from *Pseudomonas syringae* pv. *syringae* UMAF0158 extracellular matrix, focusing on a Psl-like polysaccharide. *npj Biofilms and Microbiomes, 6*(1), 1–13.
Hu, L., Grim, C. J., Franco, A. A., Jarvis, K. G., Sathyamoorthy, V., Kothary, M. H., … Tall, B. D. (2015). Analysis of the cellulose synthase operon genes, *bcsA*, *bcsB*, and *bcsC* in *Cronobacter* species: Prevalence among species and their roles in biofilm formation and cell–cell aggregation. *Food Microbiology, 52*, 97–105. https://doi.org/10.1016/j.fm.2015.07.007
Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., & Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications, 9*(1), 1–8.
Kim, S. H., Jyung, S., & Kang, D. H. (2022). Comparative study of *Salmonella* Typhimurium biofilms and their resistance depending on cellulose secretion and maturation temperatures. *LWT, 154*, Article 112700. https://doi.org/10.1016/j.lwt.2021.112700
Kim, Y. K., & McCarter, L. L. (2000). Analysis of the polar flagellar gene system of *Vibrio parahaemolyticus*. *Journal of Bacteriology, 182*(13), 3693–3704.
Koblížek, M. (2015). Ecology of aerobic anoxygenic phototrophs in aquatic environments. *FEMS Microbiology Reviews, 39*(6), 854–870.
Lei, T., Jiang, F., He, M., Zhang, J., Zeng, H., Chen, M., … Wu, Q. (2020). Prevalence, virulence, antimicrobial resistance, and molecular characterization of fluoroquinolone resistance of *Vibrio parahaemolyticus* from different types of food samples in China. *International Journal of Food Microbiology, 317*, Article 108461. https://doi.org/10.1016/j.ijfoodmicro.2019.108461
Lerouge, I., & Vanderleyden, J. (2002). O-antigen structural variation: Mechanisms and possible roles in animal/plant–microbe interactions. *FEMS microbiology reviews, 26*(1), 17–47.
Lianou, A., Nychas, G. J. E., & Koutsoumanis, K. P. (2020). Strain variability in biofilm formation: A food safety and quality perspective. *Food Research International, 137*, Article 109424. https://doi.org/10.1016/j.foodres.2020.109424
Macarisin, D., Patel, J., Bauchan, G., Giron, J. A., & Sharma, V. K. (2012). Role of curli and cellulose expression in adherence of *Escherichia coli* O157: H7 to spinach leaves. *Foodborne pathogens and disease, 9*(2), 160–167. https://doi.org/10.1089/fpd.2011.1020
Makarova, K. S., Haft, D. H., Barrangou, R., Brouns, S. J. J., Charpentier, E., Horvath, P., … Koonin, E. V. (2011). Evolution and classification of the CRISPR–Cas systems. *Nature Reviews Microbiology, 9*(6), 467–477. https://doi.org/10.1038/nrmicro2577
Martinez-Urtaza, J., Bowers, J. C., Trinanes, J., & DePaola, A. (2010). Climate anomalies and the increasing risk of *Vibrio parahaemolyticus* and *Vibrio vulnificus* illnesses. *Food Research International, 43*(7), 1780–1790. https://doi.org/10.1016/j.foodres.2010.04.001
Maturana, J. L., & Cárdenas, J. P. (2021). Insights on the evolutionary genomics of the *Blautia* Genus: Potential new species and genetic content among lineages. *Frontiers in Microbiology, 12*. https://doi.org/10.3389/fmicb.2021.660920
McCarter, L. L. (2001). Polar flagellar motility of the *Vibrionaceae*. *Microbiology and Molecular Biology Reviews, 65*(3), 445–462.
Meparambu Prabhakaran, D., Patel, H. R., Chandrika, S. K., & S., & Thomas, S.. (2022). Genomic attributes differ between *Vibrio parahaemolyticus* environmental and clinical isolates including pathotypes. *Environmental Microbiology Reports, 14*(3), 365–375. https://doi.org/10.1111/1758-2229.13000
Michael, V., Frank, O., Bartling, P., Scheuner, C., Göker, M., Brinkmann, H., & Petersen, J. (2016). Biofilm plasmids with a rhamnose operon are widely distributed determinants of the 'swim-or-stick' lifestyle in roseobacters. *The ISME Journal, 10*(10), 2498–2513. https://doi.org/10.1038/ismej.2016.30
Omadjela, O., Narahari, A., Strumillo, J., Mélida, H., Mazur, O., Bulone, V., & Zimmer, J. (2013). BcsA and BcsB form the catalytically active core of bacterial cellulose synthase sufficient for *in vitro* cellulose synthesis. *Proceedings of the National Academy of Sciences, 110*(44), 17856–17861.
Moorthy, S., & Watnick, P. I. (2004). Genetic evidence that the *Vibrio cholerae* monolayer is a distinct stage in biofilm development. *Molecular Microbiology, 52*, 573–587.
Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T., … Parkhill, J. (2015). Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics, 31*(22), 3691–3693.
Pang, Y., Guo, X., Tian, X., Liu, F., Wang, L., Wu, J., … Liu, B. (2019). Developing a novel molecular serotyping system based on capsular polysaccharide synthesis gene clusters of *Vibrio parahaemolyticus*. *International Journal of Food Microbiology, 309*, Article 108332.
Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research, 25*(7), 1043–1055.
Pritchard, L., Glover, R. H., Humphris, S., Elphinstone, J. G., & Toth, I. K. (2016). Genomics and taxonomy in diagnostics for food security: Soft-rotting enterobacterial plant pathogens. *Analytical Methods, 8*(1), 12–24. https://doi.org/10.1039/C5AY02550H
Qin, X., Wang, H., Miao, C., Yang, X., Zhang, Y., Feng, J., … Jiang, Y. (2021). Comparative genomics reveals environmental adaptation differences between *Cronobacter* species. *Food Research International, 147*, Article 110541. https://doi.org/10.1016/j.foodres.2021.110541
Regmi, A., & Boyd, E. F. (2019). Carbohydrate metabolic systems present on genomic islands are lost and gained in *Vibrio parahaemolyticus*. *BMC Microbiology, 19*(1), 112. https://doi.org/10.1186/s12866-019-1487-6
Römling, U., & Galperin, M. Y. (2015). Bacterial cellulose biosynthesis: Diversity of operons, subunits, products, and functions. *Trends in Microbiology, 23*(9), 545–557.
Roy, P. K., Mizan, M. F. R., Hossain, M. I., Han, N., Nahar, S., Ashrafudoulla, M., … Ha, S. D. (2021). Elimination of *Vibrio parahaemolyticus* biofilms on crab and shrimp surfaces using ultraviolet C irradiation coupled with sodium hypochlorite and slightly acidic electrolyzed water. *Food Control, 128*, Article 108179. https://doi.org/10.1016/j.foodcont.2021.108179
Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics, 30*(14), 2068–2069.
Serra, D. O., & Hengge, R. (2019). Cellulose in Bacterial Biofilms. In E. Cohen, & H. Merzendorfer (Eds.), *Extracellular Sugar-Based Biopolymers Matrices* (pp. 355–392). Springer International Publishing. https://doi.org/10.1007/978-3-030-12919-4_8.
Shime-Hattori, A., Iida, T., Arita, M., Park, K. S., Kodama, T., & Honda, T. (2006). Two type IV pili of *Vibrio parahaemolyticus* play different roles in biofilm formation. *FEMS microbiology letters, 264*(1), 89–97.
Simões, M., Simões, L. C., & Vieira, M. J. (2010). A review of current and emergent biofilm control strategies. *LWT, 43*(4), 573–583. https://doi.org/10.1016/j.lwt.2009.12.008
Solano, C., García, B., Valle, J., Berasain, C., Ghigo, J. M., Gamazo, C., & Lasa, I. (2002). Genetic analysis of *Salmonella enteritidis* biofilm formation: Critical role of cellulose. *Molecular Microbiology, 43*(3), 793–808.
Stalder, T., & Top, E. (2016). Plasmid transfer in biofilms: A perspective on limitations and opportunities. *npj Biofilms and Microbiomes, 2*(1), 16022. https://doi.org/10.1038/npjbiofilms.2016.22
Valla, S., Coucheron, D. H., Fjærvik, E., Kjosbakken, J., Weinhouse, H., Ross, P., … Benziman, M. (1989). Cloning of a gene involved in cellulose biosynthesis in *Acetobacter xylinum*: Complementation of cellulose-negative mutants by the UDPG pyrophosphorylase structural gene. *Molecular and General Genetics MGG, 217*, 26–30.
Verstraeten, N., Braeken, K., Debkumari, B., Fauvart, M., Fransaer, J., Vermant, J., & Michiels, J. (2008). Living on a surface: Swarming and biofilm formation. *Trends in Microbiology, 16*(10), 496–506. https://doi.org/10.1016/j.tim.2008.07.004
Wang, D., Fletcher, G. C., On, S. L. W., Palmer, J. S., Gagic, D., & Flint, S. H. (2023). Biofilm formation, sodium hypochlorite susceptibility and genetic diversity of *Vibrio parahaemolyticus*. *International Journal of Food Microbiology, 385*, Article 110011. https://doi.org/10.1016/j.ijfoodmicro.2022.110011
Wang, D., Flint, S. H., Gagic, D., Palmer, J. S., Fletcher, G. C., & On, S. L. W. (2021). *In silico* analysis revealing CsrA roles in motility-sessility switching and tuning VBNC cells in *Vibrio parahaemolyticus*. *Biofouling, 37*(6), 680–688. https://doi.org/10.1080/08927014.2021.1955357
Wang, D., Flint, S. H., Palmer, J. S., Gagic, D., Fletcher, G. C., & On, S. L. W. (2022). Global expansion of *Vibrio parahaemolyticus* threatens the seafood industry: Perspective on controlling its biofilm formation. *LWT, 158*, Article 113182. https://doi.org/10.1016/j.lwt.2022.113182
Yang, Z., Wafula, E. K., Kim, G., Shahid, S., McNeal, J. R., Ralph, P. E., … dePamphilis, C. W. (2019). Convergent horizontal gene transfer and cross-talk of mobile nucleic acids in parasitic plants. *Nature Plants, 5*(9), 991–1001. https://doi.org/10.1038/s41477-019-0458-0
Yildiz, F. H., & Visick, K. L. (2009). *Vibrio* biofilms: So much the same yet so different. *Trends in Microbiology, 17*(3), 109–118. https://doi.org/10.1016/j.tim.2008.12.004
Zerbino, D. R., & Birney, E. (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research, 18*(5), 821–829.
Zhong, C., Zhang, G.-C., Liu, M., Zheng, X.-T., Han, P.-P., & Jia, S.-R. (2013). Metabolic flux analysis of *Gluconacetobacter xylinus* for bacterial cellulose production. *Applied Microbiology and Biotechnology, 97*(14), 6189–6199. https://doi.org/10.1007/s00253-013-4908-8
Zhu, Q., Kosoy, M., & Dittmar, K. (2014). HGTector: An automated method facilitating genome-wide discovery of putative horizontal gene transfers. *BMC Genomics, 15*(1), 1–18.