




Article

Comparative Genomics of Clinical and Environmental Isolates of *Vibrio* spp. of Colombia: Implications of Traits Associated with Virulence and Resistance

Alejandra Pérez-Duque^{1,2}, Andrea Gonzalez-Muñoz¹, Jorge Arboleda-Valencia^{1,3,*} ,
Lizbeth Janet Vivas-Aguas⁴, Tania Córdoba-Meza⁴, Ghennie Tatiana Rodriguez-Rey² , Paula Díaz-Guevara⁵,
Jaime Martínez-Urtaza⁶  and Magdalena Wiesner-Reyes^{5,*}

- ¹ Centro de Bioinformática y Biología Computacional BIOS, Manizales 170002, Colombia; alejandra.mapd@gmail.com (A.P.-D.); andrea.gonzalez@bios.co (A.G.-M.)
² BIONAT: Grupo de Investigaciones en Biodiversidad y Recursos Naturales, Universidad de Caldas, Manizales 170002, Colombia; ghennie.rodriguez@ucaldas.edu.co
³ Grupo FITOBIOL, Instituto de Biología, Facultad de Ciencias Exactas y Naturales, Universidad de Antioquia, Medellín 050010, Colombia
⁴ Programa Calidad Ambiental Marina, Instituto de Investigaciones Marinas y Costeras José Benito Vives De Andreis INVEMAR, Santa Marta 470006, Colombia; janet.vivas@invemar.org.co (L.J.V.-A.); tania.cordoba@invemar.org.co (T.C.-M.)
⁵ Grupo de Microbiología, Instituto Nacional de Salud, Bogotá 111321, Colombia; pdiaz@ins.gov.co
⁶ Department of Genetics and Microbiology, Faculty of Biosciences, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain; jaime.martinez.urtaza@uab.cat
* Correspondence: jorge.arboleda@bios.co (J.A.-V.); mwiesner@ins.gov.co (M.W.-R.)



Citation: Pérez-Duque, A.; Gonzalez-Muñoz, A.; Arboleda-Valencia, J.; Vivas-Aguas, L.J.; Córdoba-Meza, T.; Rodriguez-Rey, G.T.; Díaz-Guevara, P.; Martínez-Urtaza, J.; Wiesner-Reyes, M. Comparative Genomics of Clinical and Environmental Isolates of *Vibrio* spp. of Colombia: Implications of Traits Associated with Virulence and Resistance. *Pathogens* **2021**, *10*, 1605. <https://doi.org/10.3390/pathogens10121605>

Academic Editor:
Daniele Provenzano

Received: 7 October 2021
Accepted: 9 November 2021
Published: 10 December 2021

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



Copyright: © 2021 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/>).

Abstract: There is widespread concern about the increase in cases of human and animal infections caused by pathogenic *Vibrio* species due to the emergence of epidemic lineages. In Colombia, active surveillance by the National Institute of Health (INS) has confirmed the presence of *Vibrio*; however, in routine surveillance, these isolates are not genomically characterized. This study focused on the pangenome analysis of six *Vibrio* species: *V. parahaemolyticus*, *V. vulnificus*, *V. alginolyticus*, *V. fluvialis*, *V. diabolicus* and *V. furnissii* to determine the genetic architectures of potentially virulent and antimicrobial resistance traits. Isolates from environmental and clinical samples were genome sequenced, assembled and annotated. The most important species in public health were further characterized by multilocus sequence typing and phylogenomics. For *V. parahaemolyticus*, we found the virulent ST3 and ST120 genotypes. For *V. vulnificus*, we identified isolates belonging to lineages 1 and 2. Virulence gene homologues between species were found even in non-pathogenic species such as *V. diabolicus*. Annotations related to the mobilome, integrative mobile and conjugative elements and resistance genes were obtained from environmental and clinical isolates. This study contributes genomic information to the intensified surveillance program implemented by the INS to establish potential sources of vibriosis in Colombia.

Keywords: vibriosis; virulence; antibiotic resistance; pangenome; whole genome sequencing

1. Introduction

The genus *Vibrio* comprises bacterial species present as free-living organisms in aquatic and marine habitats, or associated with fish and marine invertebrates. Several species of this bacterial genus cause food-borne diseases and wound infections resulting from contact with water or raw seafood. The number of cases of human infection by *Vibrio* species have increased worldwide. The diseases caused by *Vibrio* spp. are divided into two large groups: cholera, caused by enteropathogenic *Vibrio cholerae* and vibriosis, caused by other pathogenic *Vibrio* species [1]. Cholerae is a rapidly disseminating diarrheal disease that causes quick dehydration and can be fatal if not treated on-time [2]. Patients with vibriosis present a variety of symptoms depending on the infecting *Vibrio* species, infection route and

host susceptibility. Symptoms range from gastrointestinal illness, such as gastroenteritis, to extraintestinal pathologies, including skin infections and septicemia [1].

Vibrio is commonly found in the United States of America, where the number of cases of infection has increased since 2000, reaching approximately 80,000 cases and 100 deaths annually [3]. Similarly, countries in Asia reported impacts of *Vibrio* infections on humans and massive death of marine animals, affecting the seafood industry [4]. In Europe, the presence of *Vibrio vulnificus*, *Vibrio parahaemolyticus* and other *Vibrio* species has raised concerns regarding their potential to generate significant clinical problems and outbreaks associated with climate change [5]. In America, alerts have been raised throughout the continent due to the cholera outbreaks in Haiti and the Dominican Republic in 2010 and 2011, as well as case reports in Cuba, Venezuela and Mexico [6].

Colombia is not exempt from possible cholera or vibriosis outbreaks due to its geographic location and the presence of the Atlantic and Pacific coasts [7]. Although *V. cholerae*, *V. parahaemolyticus* and *V. vulnificus* are the main reported pathogens, there is also concern regarding an increase in cases of infection by other *Vibrio* species and the emergence of epidemic lineages [8]. A vibriosis outbreak would affect the economic sector of the Atlantic and Pacific coasts of Colombia, especially, seafood production and tourism.

In response to this contingency, the Instituto Nacional de Salud (National Institute of Health, INS) of Colombia reinforced the Intensified Cholera Surveillance program in 2010 [9]. Although Colombia is currently in an inter-epidemic period, the intensified surveillance between 2010 and 2015 recovered 502 probable *V. cholerae* isolates (e.g., toxin negative and positive), as well as other *Vibrio* species. The presence of these isolates suggests that the country is at risk for a possible reemergence of cholera and vibriosis. Furthermore, the recovered isolates belonging to non-cholera vibrios have not been genotyped or characterized at the genome level, so it is unknown if these isolates carry toxin or virulence genes and antimicrobial resistance traits [10]. Therefore, genetic characterization studies are important to understand the public health risks associated with the presence of these bacteria in the environment [11–15]. Additionally, continuous monitoring of the genotypes is essential to determine their pathogenicity and to establish potential sources of epidemic outbreaks. This information is relevant for adopting timely measures in response to a contingency, allowing to reduce the impact of this pathogen on the country.

Species of *Vibrio* have highly plastic genomes since there is a high probability of horizontal transfer of virulence and antibiotic resistance genes from virulent to non-virulent strains due to their free-living nature. Whole genome sequencing of several *Vibrio* isolates and comparative genomics analyses have revealed a wide array of mutations, chromosomal rearrangements and gene gain and loss events resulting from duplication or horizontal gene transfer [16]. Virulence genes are normally identified within mobile genetic elements; therefore, these elements can favor the appearance of novel virulent strains [17]. The *Vibrio* genome comprises a core genome composed of conserved genes necessary for essential functions and a flexible gene pool, including mobile elements, virulence factors and antibiotic resistance determinants, among others, that are involved in adaptation processes [18]. Pangenome analyses represent an adequate comparative genomics approach to study the intraspecific diversity of *Vibrio* by analyzing the core and accessory genomes of different isolates of a species.

This research involved species-level pangenome analyses to genetically characterize isolates of six non-cholera *Vibrio* species that were obtained from environmental and clinical samples in Colombia to identify potential sources of vibriosis. Furthermore, this study relied on whole genome sequencing (WGS) as a high-throughput and high-resolution approach to characterize isolates for genomic surveillance in epidemiology. The pangenome allowed us to contrast the diversity within the *Vibrio* species analyzed and the virulence and antibiotic-resistance determination allow us to identify the potential pathogenesis of the isolates. For the most important species in public health, the multilocus sequence typing (MLST) approach was used to determine pandemic clones of *V. parahaemolyticus* and phylogenetic analyses allowed effectively identifying lineages of *V. vulnificus*. Cou-

pled with isolate source information (e.g., environmental or clinical sample, geographic location), this study contributes to identifying potential environmental sources of vibriosis. This is relevant for the intensified cholera surveillance program implemented by the National Institute of Health of Colombia that aims to predict future vibriosis outbreaks in the country.

2. Results

2.1. Characteristics of the Assembled Genomes of *Vibrio* spp.

A total of 60 isolates were analyzed in this study, belonging to *V. parahaemolyticus* (n = 17), *V. vulnificus* (n = 9), *V. alginolyticus* (n = 8), *V. fluvialis* (n = 16), *V. diabolicus* (n = 4) and *V. furnissii* (n = 6). General characteristics of the obtained assemblies, such as assembled length, number of scaffolds, %GC, N50 length, number of coding sequences (CDS) and reference genome fraction are reported in Supplementary Table S1. The assembled genomes showed a similar size and GC content to the corresponding reference genome of each species. Within each species, the strains shared more than 95% average nucleotide identity (ANI). Moreover, among species, the most genetically similar were *V. alginolyticus* and *V. diabolicus* with ~90% ANI and *V. fluvialis* and *V. furnissii* with ~85% ANI (Figure 1).

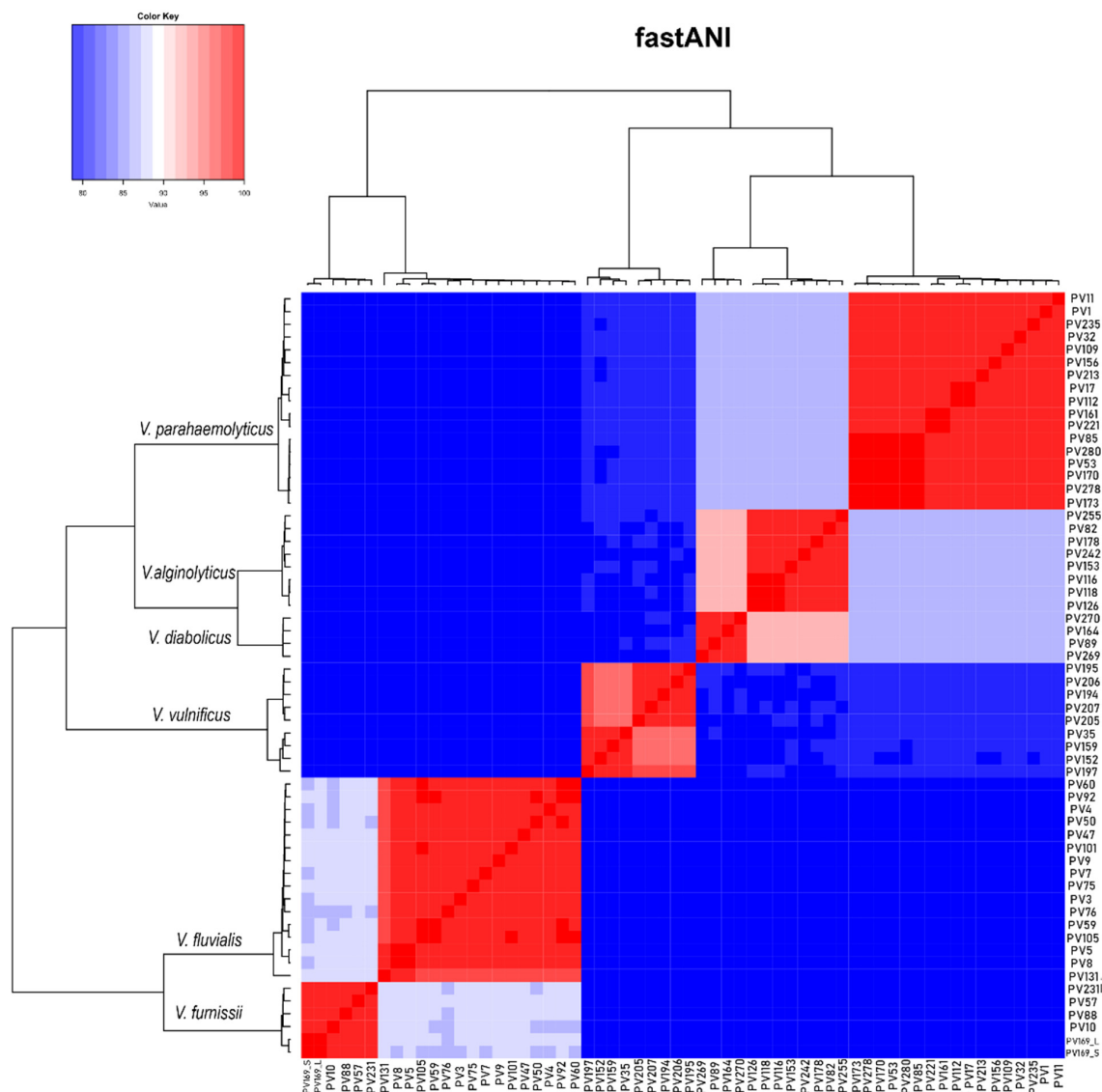


Figure 1. Heatmap based on hierarchical clustering of the Whole-genome Average Nucleotide Identity (ANI) calculated for the 60 *Vibrio* spp. isolates of the study.

2.2. *Vibrio parahaemolyticus*

Of the 17 *V. parahaemolyticus* isolates (seven clinical, ten environmental), we found seven environmental isolates with new sequence types, six clinical isolates assigned to ST3, one clinical isolate assigned to ST120 and two untypeable isolates (Supplementary Table S4).

The pangenome of *V. parahaemolyticus* comprised a total of 11313 gene clusters, including 7405 that were assigned to COG (cluster of orthologous of proteins). The core genome contained 3859 gene clusters, the accessory genome 1747 and the singletons 5707. In the pangenome scheme, we observed a defined block of gene clusters belonging to ST3 and a block of gene clusters shared by clinical isolates (ST120 and ST3). The isolates belonging to ST3 were grouped together in the maximum likelihood (ML) tree (Figure 2).

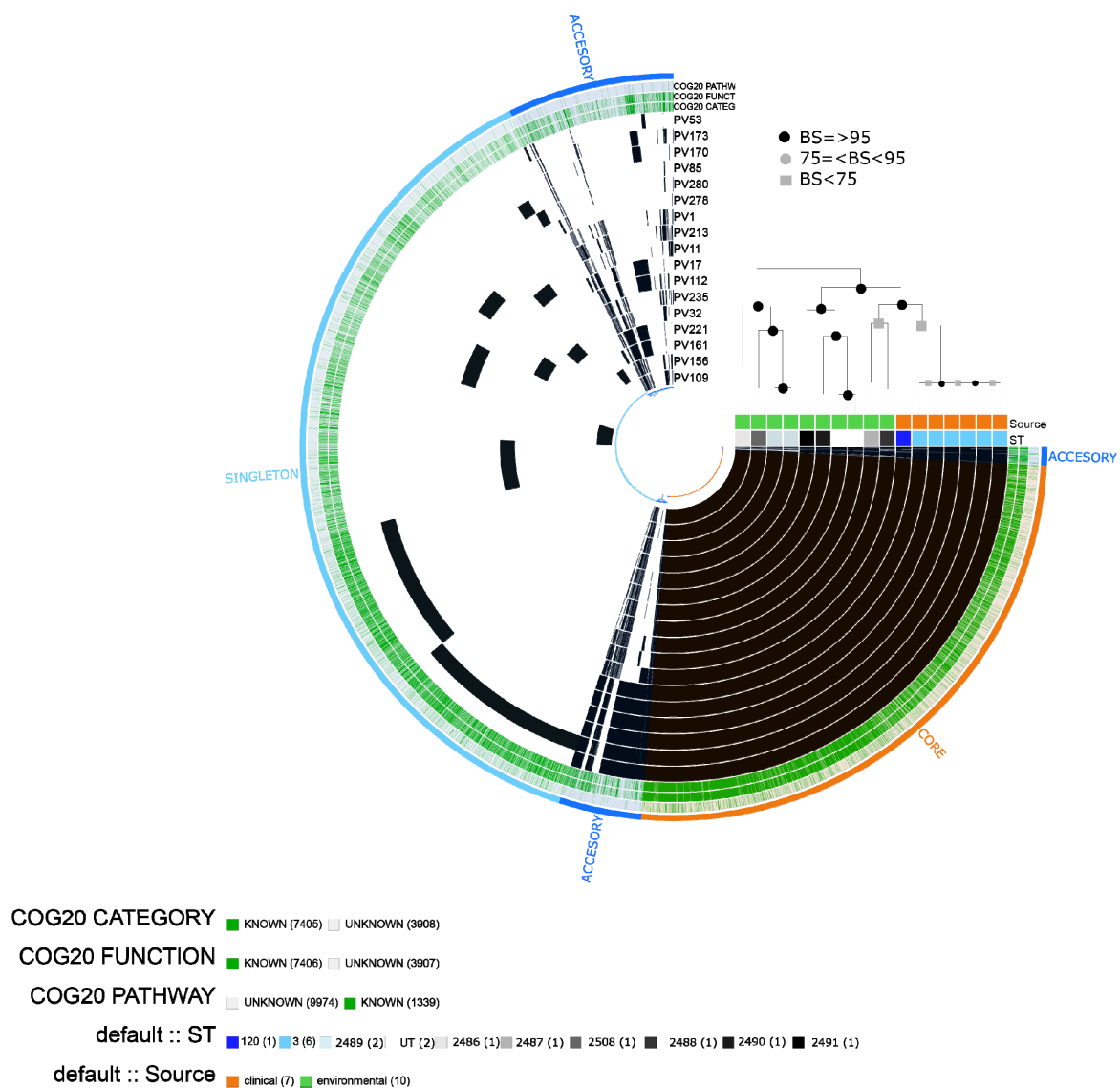


Figure 2. *Vibrio parahaemolyticus* pangenome based on 17 isolates analyzed. Presence/absence of gene clusters are represented by black and gray bars, respectively. Each bin within the pangenome is signaling core, accessory and singleton fractions. The labels represent in order: Number of gene clusters assigned to COG20_CATEGORY, COG20_FUNCTION and COG20_PATHWAY; each isolate (PV); information on ST; and source of the isolate (clinical/environmental). The pangenome is ordered according to a maximum likelihood (ML) tree based on 2149 single-copy concatenated core genes. Bootstrap supports (BS) are based on 100 replicates.

In the clinical group, we found a significant enrichment of categories (adjusted q-value $\leq 1.03E-02$) associated with: [N] cell motility, [U] intracellular trafficking, secretion

and vesicular transport related to type III secretory pathway and type VI secretion system, including proteins related to flagella and other component proteins. In contrast, the most enriched categories in the environmental isolates were associated with cell cycle control, cell division, carbohydrate, amino acid and nucleotide transport and metabolism (Table 1).

Table 1. Functional enrichment analyses of the accessory genome of clinical and environmental isolates of *Vibrio parahaemolyticus*.

COG20_Category	Enrichment_Score	Unadjusted_p_Value	Adjusted_q_Value	Associated_Isolate_Sources
Cell motility Intracellular trafficking, secretion and vesicular transport	17	3.74×10^{-5}	3.03×10^{-3}	clinical
Cell cycle control, cell division, chromosome partitioning Cell motility	13.388	2.53×10^{-4}	3.16×10^{-3}	environmental
Amino acid transport and metabolism Nucleotide transport and metabolism	13.246	2.73×10^{-4}	3.16×10^{-3}	clinical
Cell cycle control, cell division, chromosome partitioning Coenzyme transport and metabolism	13.246	2.73×10^{-4}	3.16×10^{-3}	clinical
Nucleotide transport and metabolism Signal transduction mechanisms	13.246	2.73×10^{-4}	3.16×10^{-3}	clinical
Signal transduction mechanisms Signal transduction mechanisms	13.246	2.73×10^{-4}	3.16×10^{-3}	clinical
Carbohydrate transport and metabolism Transcription	13.246	2.73×10^{-4}	3.16×10^{-3}	environmental
Cell cycle control, cell division, chromosome partitioning	10.578	1.14×10^{-3}	1.03×10^{-2}	clinical
Signal transduction mechanisms Intracellular trafficking, secretion and vesicular transport	10.578	1.14×10^{-3}	1.03×10^{-2}	clinical
Intracellular trafficking, secretion and vesicular transport General function prediction only Transcription	9.745	1.80×10^{-3}	1.03×10^{-2}	clinical
Nucleotide transport and metabolism	9.745	1.80×10^{-3}	1.03×10^{-2}	environmental

The *tth* (thermolabile hemolysin) gene, MAM7 (multivalent adhesion molecule) and genes related to T3SS1 were found in 100% of the *V. parahaemolyticus* isolates (n = 17). In support of the pangenome results, a cluster of virulence genes was observed in all ST3 (n = 6) and ST120 (n = 1) clinical isolates, whereas it was absent in the environmental isolates. This cluster included *tdh* (thermostable direct hemolysin A) and different proteins belonging to the type III secretion system (T3SS1 and T3SS2), including effector proteins (*vopC*, *vopT*, *vopZ*, VPA1331, *vopA/vopP*, *vopL*), translocator proteins (*vopD2*, *vopB2*), transcriptional regulator proteins (*vtrA*), C-ring proteins (*vscT2*, *vscR2*), chaperones (VPA1363) and other proteins and putative proteins (*VscS2*, *VscC2*, *VscQ2*, *vscU2*, *vcrD2*, *vscJ2*, VPA1337, *vscN2*).

Regarding antibiotic resistance, the most predicted genes were related to the tetracycline efflux pump *tet(35)* and resistance to beta-lactamases (CARB-21, CARB-23), which were found in the 17 isolates of *V. parahaemolyticus*. Four isolates (including three envi-

ronmental) contained the cAMP receptor protein (CRP), which is a global regulator that represses MdtEF multidrug efflux pump expression associated with resistance to fluoroquinolone, macrolide and penam [19]. One clinical isolate (e.g., PV173) presented an additional gene (SHV-187) related to resistance to carbapenem, cephalosporin and penam, according to CARD (Figure 3).

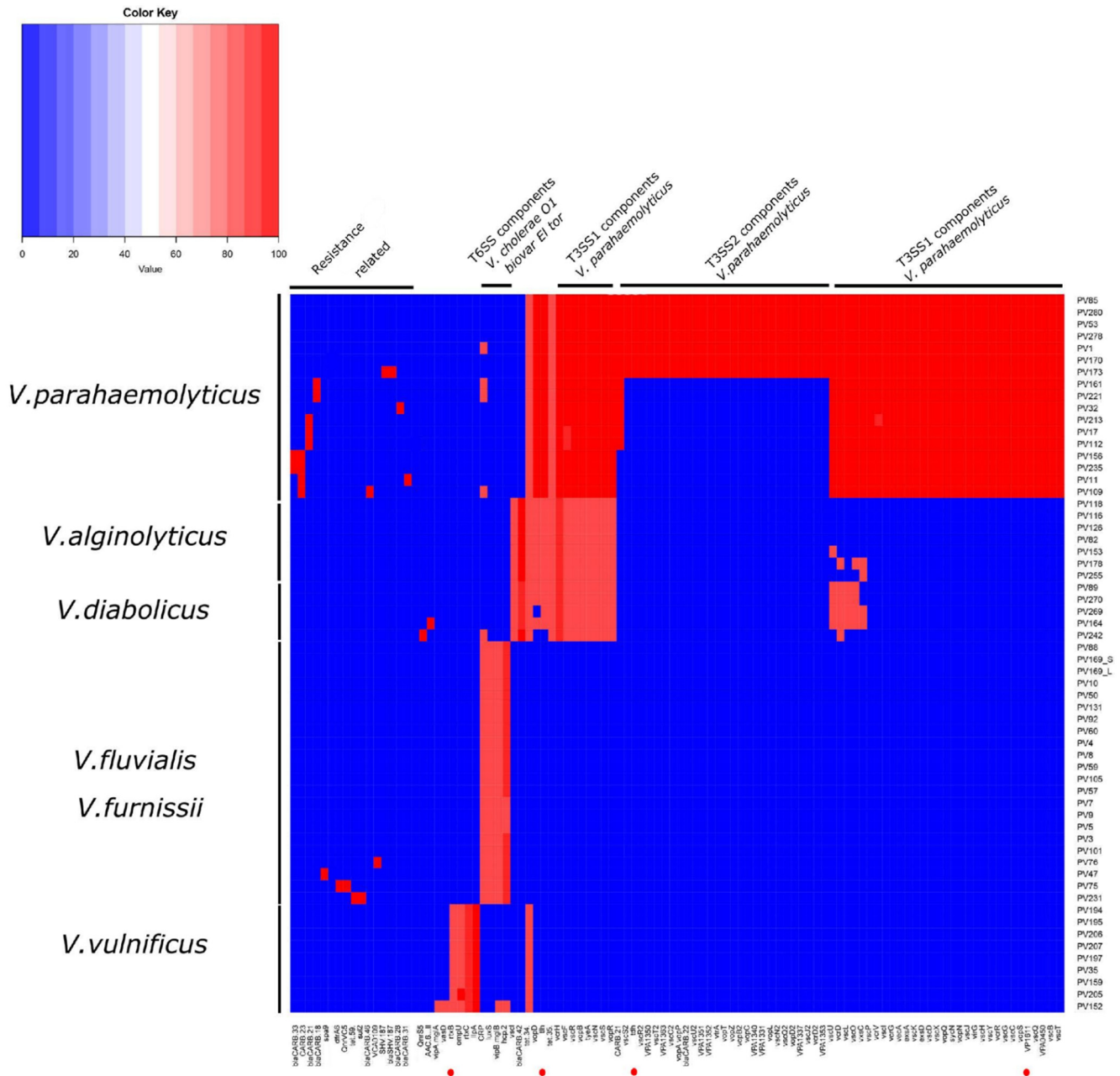


Figure 3. Heatmap of the in silico predicted virulence and resistance genes of the 60 *Vibrio* spp. isolates grouped according to sequence identity. This analysis allowed the separation between species and between clinical and environmental *V. parahaemolyticus* isolates due the different virulence profile predicted for each, except for *V. fluvialis* and *V. furnissii*.

In isolate PV173 (ST3), we found a putative integrative and mobilizable element (IME) with a size of ~80kb containing virulence genes related to T3SS2 and other genes such as *acfD*, *tdh*, *flhA*. In PV1 (ST120), we found a putative integrative and conjugative element (ICE) with T4SS containing virulence genes related to MSHA type IV pilus (Table 2).

Table 2. Integrative elements (ICE/IME) predicted for 60 *Vibrio* spp isolates. The type “putative IME” contain predicted Integrase and Relaxase and “putative ICE with T4SS” corresponds to elements with predicted Relaxase, T4CP, Integrase and T4SS components.

Species	Isolate	Type	GC (%)	Size (kbp)	Predicted Virulence Genes (Min e-Value 0.01)
<i>V. parahaemolyticus</i>	PV1	Putative ICE	44.45	195	VP1611(MAM7), <i>mshE</i> , <i>mshH</i> , <i>mshA</i> , <i>mshL</i> , <i>mshG</i> , <i>mshJ</i> , <i>tcpI</i>
<i>V. parahaemolyticus</i>	PV173	Putative IME	39.38	82	<i>vcrD2</i> , <i>vscC2</i> , <i>vopL</i> , <i>vscN2</i> , <i>vopB2</i> , <i>vscU2</i> , <i>VPA1351</i> , <i>vopA/vopP</i> , <i>VPA1353</i> , <i>vopZ</i> , <i>vopD2</i> , <i>vopT</i> , <i>vscR2</i> , <i>vscT2</i> , <i>VPA1350</i> , <i>vscJ2</i> , <i>VPA1352</i> , <i>VPA1331</i> , <i>VPA1331</i> , <i>VPA1363</i> , <i>VPA1337</i> , <i>VPA1340</i> , <i>vscS2</i> , <i>vscQ2</i> , <i>acfD</i> , <i>tdh</i> , <i>flhA</i> , <i>yscV/lcrD</i> , <i>flhA</i> , <i>vcrD</i> , <i>yscN</i> , <i>mxIA</i> , <i>flhA</i>
<i>V. parahaemolyticus</i>	PV278	Putative IME	40.6	35	
<i>V. fluvialis</i>	PV3	Putative ICE with T4SS	52	166	<i>llpA</i> , <i>motB</i> , <i>motA</i>
<i>V. fluvialis</i>	PV4	Putative IME	46.86	74	
<i>V. fluvialis</i>	PV9	Putative IME	50.94	40	
<i>V. fluvialis</i>	PV50	Putative ICE with T4SS	48.48	233	<i>tcpI</i>
<i>V. fluvialis</i>	PV50	Putative IME	45.8	77	
<i>V. fluvialis</i>	PV59	Putative ICE with T4SS	47.8	215	<i>cqsA</i>
<i>V. fluvialis</i>	PV59	Putative ICE with T4SS	48.85	231	
<i>V. fluvialis</i>	PV60	Putative ICE with T4SS	48.4	296	<i>flaC</i> , <i>flaB</i> , <i>flgA</i> , <i>flgJ</i> , <i>flaD</i> , <i>flgT</i> , <i>flgP</i> , <i>flgI</i> , <i>flgL</i> , <i>flgC</i> , <i>flgO</i> , <i>flgF</i> , <i>flgB</i> , <i>cheR</i> , <i>cheV</i> , <i>flgK</i> , <i>flgD</i> , <i>flaA</i> , <i>flgH</i> , <i>flgG</i> , <i>tcpI</i> , <i>flgN</i>
<i>V. fluvialis</i>	PV76	Putative ICE with T4SS	47.74	226	<i>cqsA</i>
<i>V. fluvialis</i>	PV92	Putative IME with T4SS	47.99	183	<i>cqsA</i>
<i>V. fluvialis</i>	PV92	Putative IME	50.62	48	
<i>V. fluvialis</i>	PV105	Putative ICE with T4SS	49.15	165	<i>flgD</i>
<i>V. fluvialis</i>	PV131	Putative IME	47.34	52	
<i>V. alginolyticus</i>	PV126	Putative IME	41.8	75	
<i>V. alginolyticus</i>	PV242	Putative IME	45.81	15	
<i>V. vulnificus</i>	PV35	Putative IME	43.19	79	
<i>V. vulnificus</i>	PV159	Putative IME	40.53	15	
<i>V. vulnificus</i>	PV194	Putative IME	49.13	62	
<i>V. vulnificus</i>	PV207	Putative IME	40.18	57	
<i>V. furnissii</i>	PV10	Putative ICE with T4SS	53.14	173	
<i>V. furnissii</i>	PV88	Putative IME	45.96	80	
<i>V. furnissii</i>	PV169_L	Putative ICE with T4SS	51.94	388	

2.3. *Vibrio vulnificus*

Of the nine *V. vulnificus* isolates (eight environmental and one clinical), one isolate (PV207) could be assigned to a pre-existing MLST profile (ST335), while the other eight isolates were assigned to new sequence types (Supplementary Table S5).

We obtained a total of 6806 gene clusters, including 4504 assigned to COG functions. The pangenome contained 3510 core genes, 1069 accessory genes and 2224 singletons. Due to the low number of clinical isolates ($n = 1$), the differential analysis of enrichment categories was not done. Instead, we observed that all isolates contained categories related to important pathogenicity traits, such as [P] Inorganic ion transport and metabolism related to iron uptake, [U] Vesicular intracellular traffic, secretion and transport, [O] Post-translational modification, protein turnover, chaperones related to metalloproteases; [Q] Secondary metabolites biosynthesis, transport and catabolism, [N] Cell motility related to flagella, [T] Signal transduction mechanisms related to chemotaxis, among others.

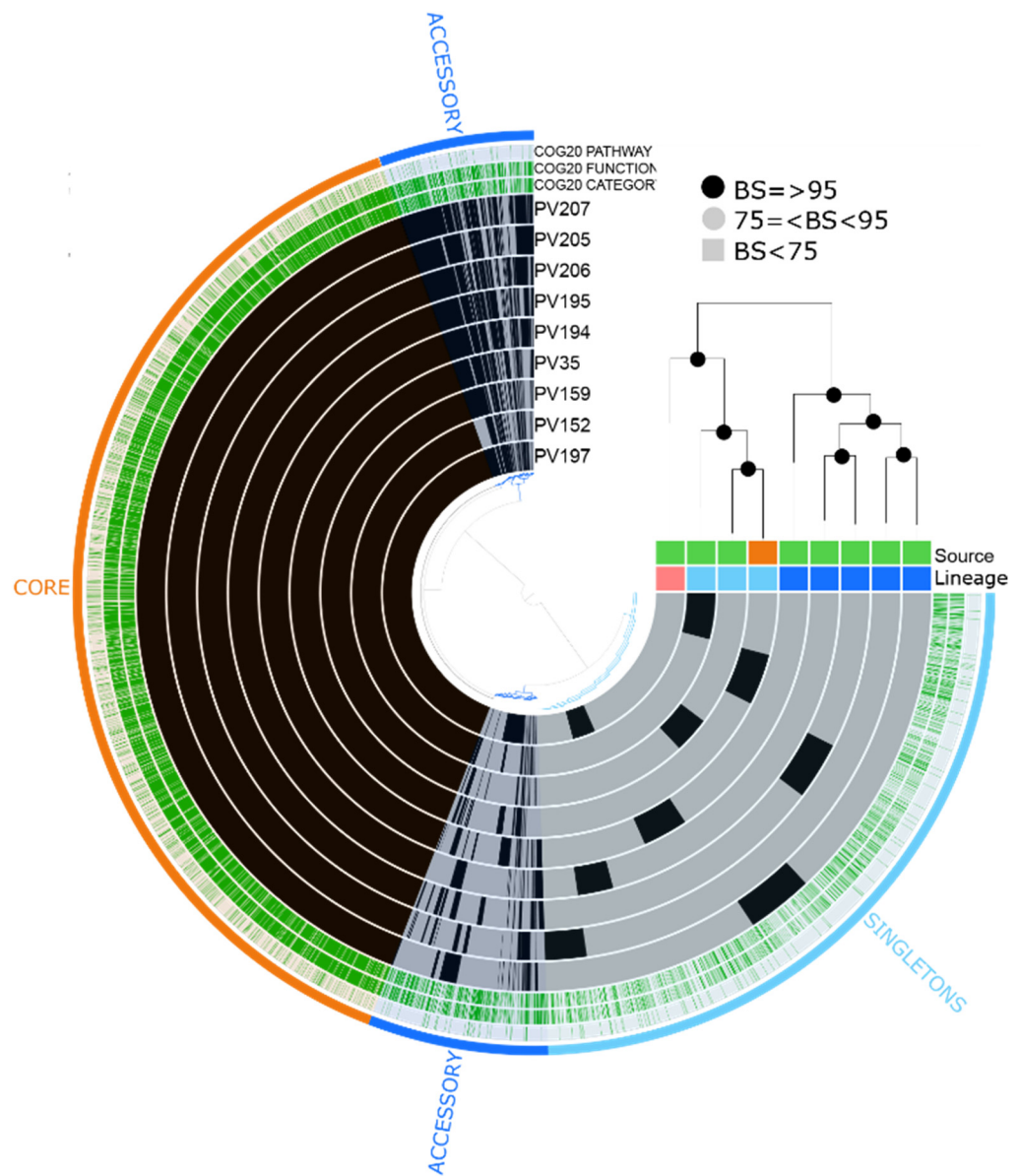
Based on the core genome phylogeny, the nine isolates were grouped into three clusters (Figure 4). Furthermore, the phylogenetic analysis (Supplementary Figure S1) and the whole-genome average nucleotide identity (Supplementary Figure S2) based on alignments to 80 reference *V. vulnificus* isolates obtained from [20], showed that five environmental isolates belonged to Lineage 1 and two environmental isolates and one clinical to Lineage 2. In the whole genome average nucleotide identity analyses, isolate PV197 showed high similarity to the reference isolate V252, assigned to lineage 5. However, despite this result, these two isolates did not form a monophyletic group in the phylogenetic analyses.

In 100% of the *V. vulnificus* isolates, we found virulence genes homologous to *V. cholerae* O1 biovar El Tor, including genes related to RTX toxin activation (*rtxC*) and transportation (*rtxB*). Additionally, *ompU* and *IlpA* were found in all isolates. Regarding resistance, we predicted a gene related to tetracycline resistance *tet(34)* in all the isolates (Figure 3).

2.4. *Vibrio alginolyticus*

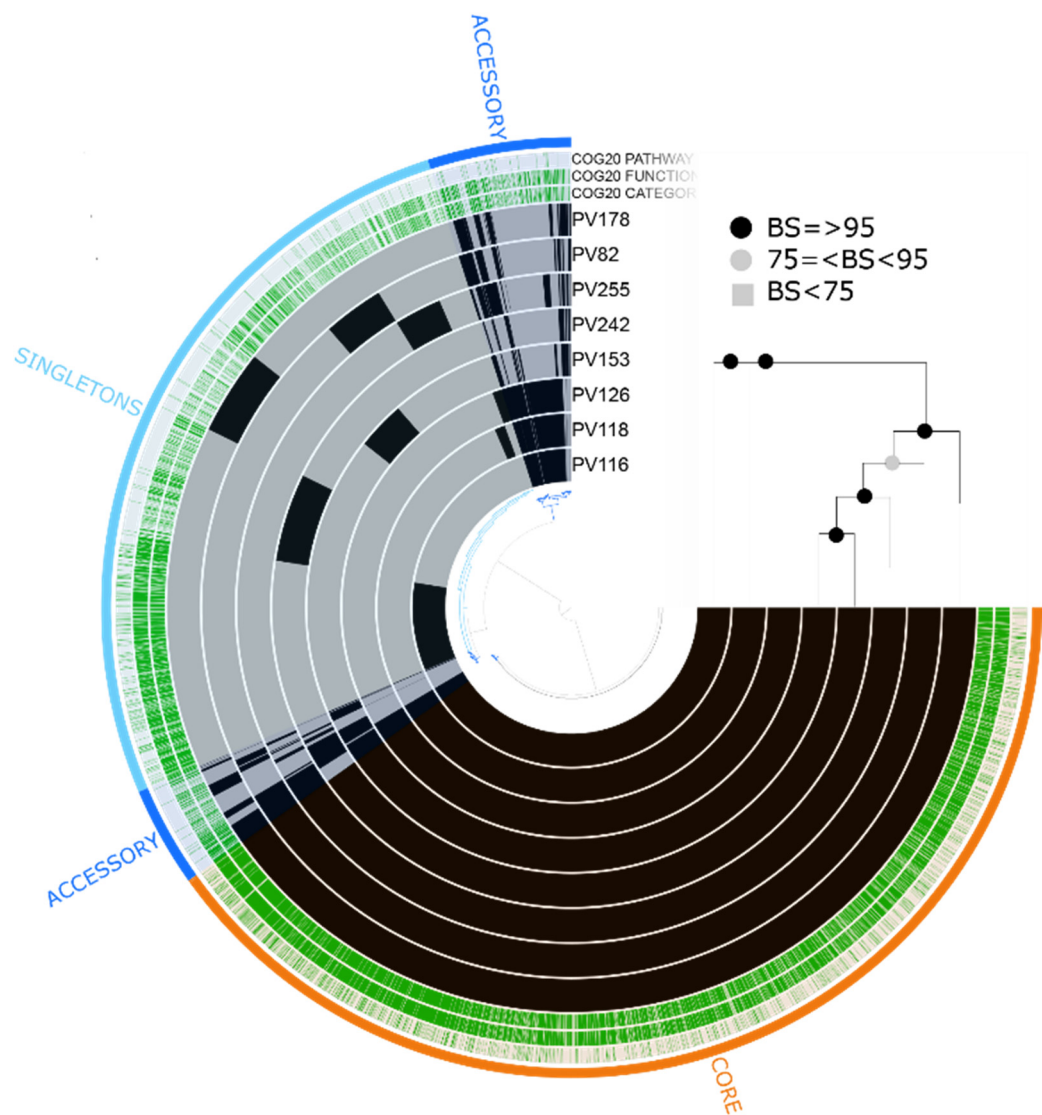
For the eight environmental isolates of *V. alginolyticus*, we obtained a total of 7226 gene clusters, including 5081 assigned to COG functions. The pangenome contained 3866 core genes, 808 accessory genes and 2550 singletons. Categories related to virulence were found in the eight isolates, namely [N | W] Cell motility | Extracellular structures related to type IV pilus and type II secretory pathway, [O] Posttranslational modification, protein turnover, chaperones related to proteases, [T] Signal transduction mechanisms related to chemotaxis, [V] Defense mechanisms, [R] General function prediction only related to putative hemolysin, [P] Inorganic ion transport and metabolism related to iron utilization and transport, among others. Based on the phylogenetic analyses, we observed a group composed of PV116, PV118 and PV126, which shared a specific cluster of genes mainly related to [Q] Secondary metabolites biosynthesis, transport and catabolism [T] Signal transduction mechanisms [P] Inorganic ion transport and metabolism, [P] coenzyme transport and metabolism, [C] Energy production and conversion, among others, that could be involved in adaptation to specific hosts or environments (Figure 5).

The eight isolates contained the *tlh* gene and genes of the T3SS1 secretion system, such as *VopD*, *VopB*, *VcrH*, *tyeA*, *vscN*, *vscR*, *vscS*, *vopR*, *vscI* and *vscF*. Similar to *V. parahaemolyticus*, the eight isolates of *V. alginolyticus* shared the predicted gene *tet(35)* related to the tetracycline efflux pump. One isolate (PV242) presented CRP and a plasmid-mediated quinolone resistance protein (QnrS5) (Figure 3).



COG20 CATEGORY ■ KNOWN (4504) ■ UNKNOWN (2302)
 COG20 FUNCTION ■ KNOWN (4504) ■ UNKNOWN (2302)
 COG20 PATHWAY ■ UNKNOWN (5943) ■ KNOWN (863)
 default :: Lineage ■ L1 (5) ■ L2 (3) ■ ND (1)
 default :: Source ■ environmental (8) ■ clinical (1)

Figure 4. *Vibrio vulnificus* pangenome. The labels represent in order: number of gene clusters assigned to COG20_CATEGORY, COG20_FUNCTION and COG20_PATHWAY, lineage (L1 or L2) and source (clinical or environmental). Maximum likelihood tree based on 2429 single-copy concatenated core genes. Bootstrap supports (BS) are based on 100 replicates.



COG20 CATEGORY ■ KNOWN (5081) ■ UNKNOWN (2145)
COG20 FUNCTION ■ KNOWN (5081) ■ UNKNOWN (2145)
COG20 PATHWAY ■ UNKNOWN (6245) ■ KNOWN (981)

Figure 5. *Vibrio alginolyticus* pangenome. The labels represent in order: number of gene clusters assigned to COG20_CATEGORY, COG_20FUNCTION and COG_20 PATHWAY. Maximum likelihood tree based on 2853 single-copy concatenated core genes. Bootstrap supports (BS) are based on 100 replicates.

2.5. *Vibrio diabolicus*

We predicted 6327 total genes, including 4476 assigned to COG functions. There were 3922 core genes, 599 accessory genes and 1788 singleton genes in the pangenome based on the four *V. diabolicus* environmental isolates. Similar to *V. alginolyticus*, we found the following functional categories related to virulence: [N | W] Cell motility | Extracellular structures related to type IV pilus and flagella-related proteins, [U] Intracellular trafficking, secretion and vesicular transport related to type II secretion pathway, [Q] Secondary metabolites biosynthesis, [T] Signal transduction mechanisms related to chemotaxis, [V] Defense mechanisms, [P] Inorganic ion transport and metabolism related to iron utilization, transport and siderophore, among others. The isolates were grouped into two clusters comprising isolates PV270/PV269 and PV89/PV164. Furthermore, PV270 and PV269 shared a gene block with functional categories mainly related to [G] Carbohydrate transport and

metabolism, [D] Cell cycle, cell division, chromosome partitioning, [C] Energy production and conversion, among others (Figure 6).

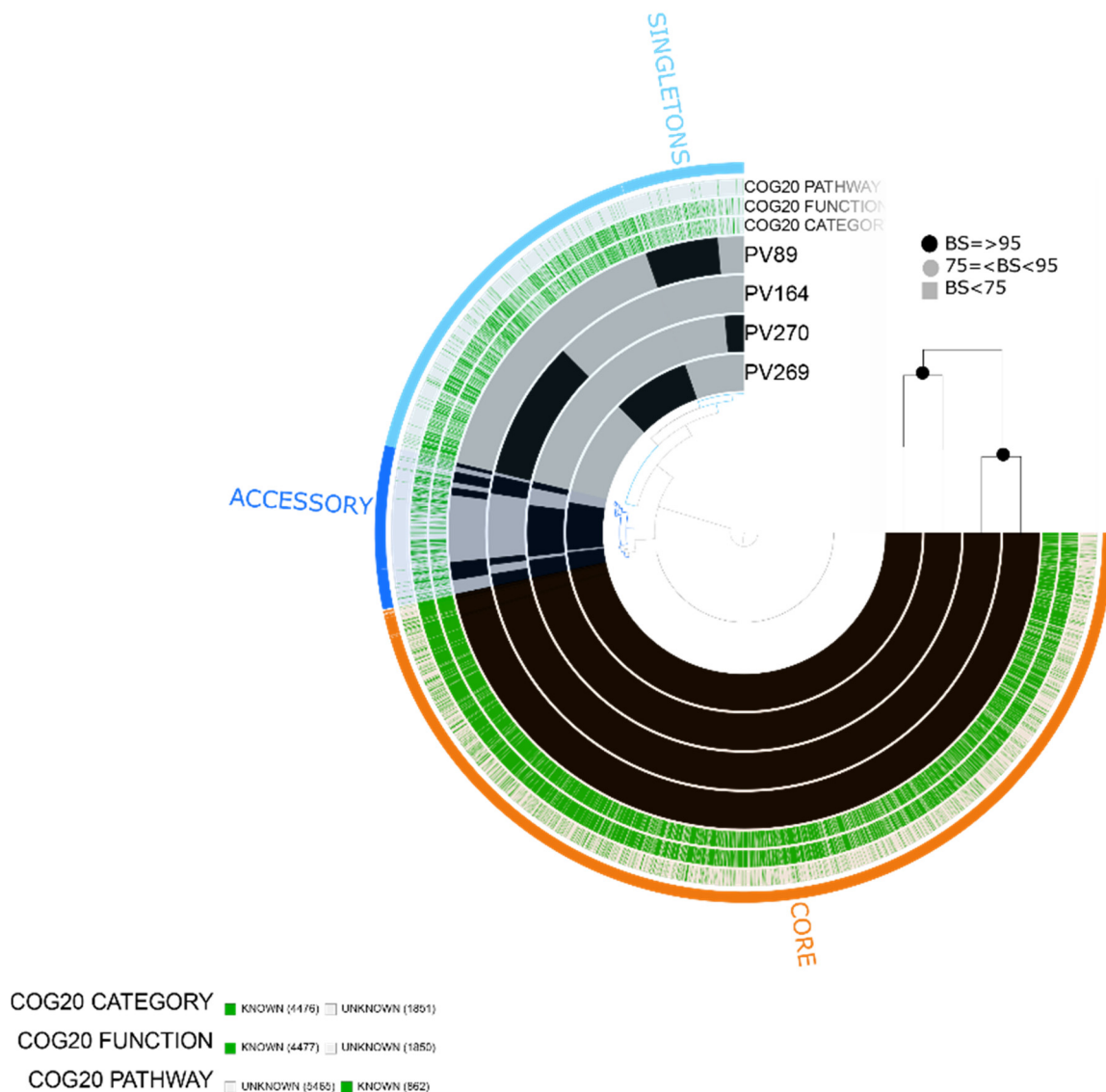


Figure 6. *Vibrio diabolicus* pangenome. The labels represent in order: number of clusters of genes assigned to COG20_CATEGORY, COG20_FUNCTION and COG20_PATHWAY. Maximum likelihood tree based on 2780 single-copy concatenated core genes. Bootstrap supports (BS) are based on 100 replicates.

We found that the virulence profile of *V. diabolicus* was similar to that of *V. alginolyticus* since both species shared genes related to T3SS1. A cluster of virulence genes was found with additional genes related to T3SS1 (*vsclU*, *vsclD*, *vsclL*, *vsclO*, *YsclO* and *vsclC*). Similar to *V. parahaemolyticus* and *V. alginolyticus*, the four *V. diabolicus* isolates contained the *tet(35)* antibiotic resistance determinant gene (Figure 3).

2.6. *Vibrio fluvialis*

We predicted a total of 8139 gene clusters (5222 assigned to COG functions) distributed into 3666 core genes, 1574 accessory genes and 2898 singletons, which comprised the pangenome of the 16 isolates of *V. fluvialis* (five clinical isolates and 11 environmental isolates) (Figure 7).

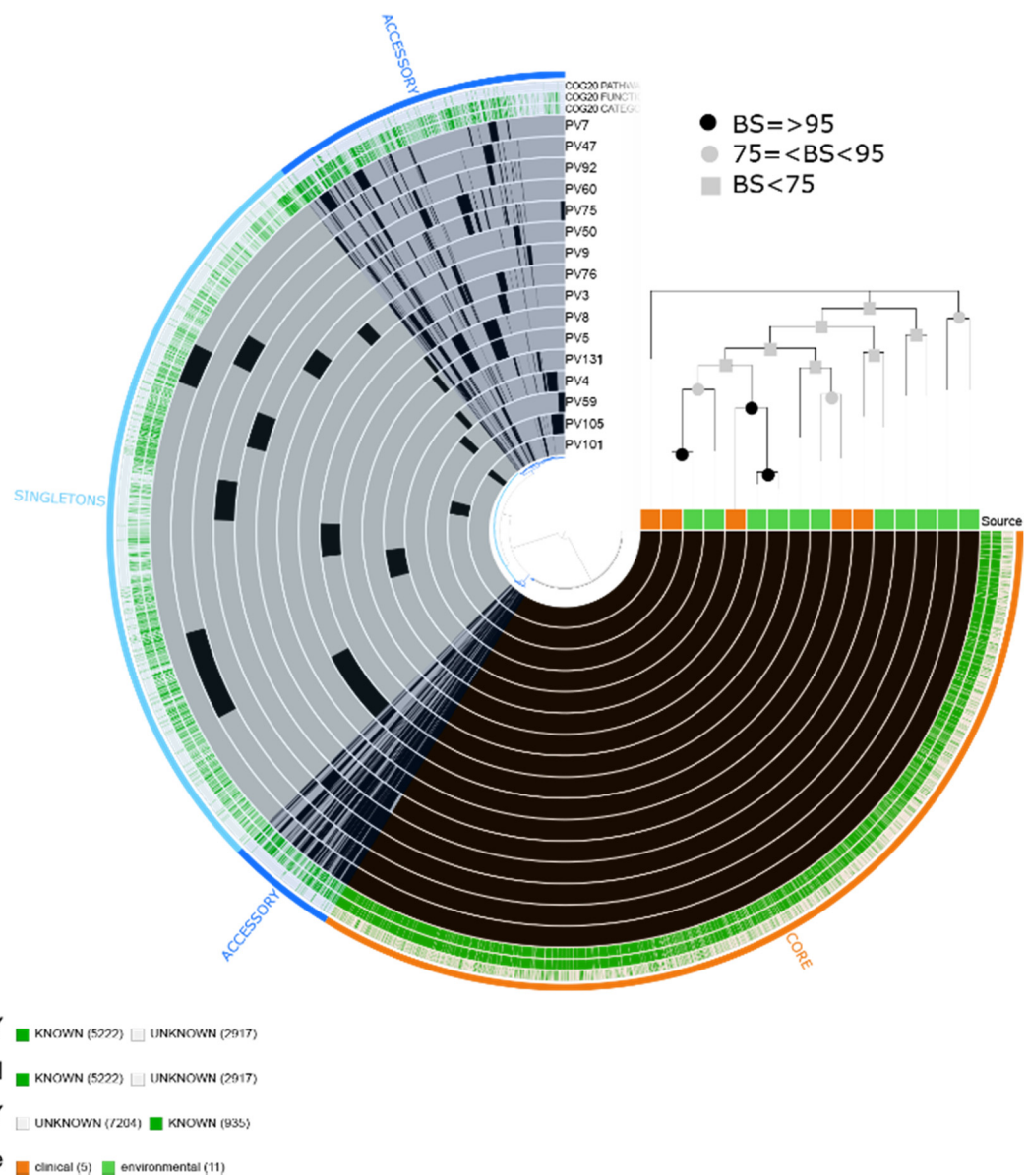


Figure 7. *Vibrio fluvialis* pangenome. The labels represent in order: number of clusters of genes assigned to COG20 CATEGORY, COG20_FUNCTION and COG_20 PATHWAY and isolate source (clinical or environmental). Maximum likelihood tree based on 2382 single-copy concatenated core genes. Bootstrap supports (BS) are based on 100 replicates.

The enrichment test did not show significantly enriched functional categories among the clinical and environmental isolates. Moreover, the phylogenetic analyses did not indicate a differential aggrupation among the isolates and we did not observe specific blocks of gene clusters related to groups within the pangenome. Therefore, we report the categories related to virulence mechanisms in the core genome. We found categories such as [N|U|W] Cell motility| Intracellular trafficking, secretion and vesicular transport| Extracellular structures related to secretory pathway type VI, [R] General prediction only related to putative hemolysins, [Q] Inorganic ion transport and metabolism related to mechanisms implicated in iron uptake, [T] Signal transduction mechanisms related to chemotaxis, among others.

The sequence homology-based search showed that the virulence factors predicted in *V. fluvialis* were mostly associated with *V. cholerae* O1 biovar El Tor. In both clinical and environmental isolates, we found genes associated with secretion system type VI (T6SS) including type VI secretion system substrate (*hcp-2*), type VI secretion system tubule-forming protein VipB (*vipB/mglB*) and the *luxS* gene. The 16 isolates contained CRP and

one isolate (PV75) presented a predicted gene (*QnrVCS*) related to resistance to quinolone and *dfrA6* related to resistance to trimethoprim (Figure 3).

Out of the 16 isolates, we found predicted IMEs and ICEs in 10 isolates. These elements contained predicted virulence genes such as *cqsA*, *tcpl*, *IlpA*, *motB*, *motA* and *flgD*. Furthermore, in PV60, a predicted ICE contained genes related to flagella (*flaC*, *flaB*, *flgA*, *flgJ*) and chemotaxis (*cheR*, *cheV*) (Table 2).

2.7. *Vibrio furnissii*

We obtained 5950 total genes, including 4300 assigned to COG functions. The pangenome contained 3822 core genes, 787 accessory genes and 1339 singletons based on six *V. furnissii* isolates. The isolates (four clinical and two environmental) did not show significantly enriched categories or differential aggrupation among clinical and environmental isolates. Similar to the results for *V. fluvialis*, the isolates contained gene clusters assigned to virulence-related categories such as [U] Intracellular trafficking, secretion and vesicular transport implicated in diverse components of type VI secretion system; [N] Cell motility related to flagellar biosynthesis proteins [P] Inorganic ion transport and metabolism including mechanisms related to iron uptake such as transport systems, enterobactin and siderophore related [G] Carbohydrate transport and metabolism related to chitinase, among others (Figure 8).

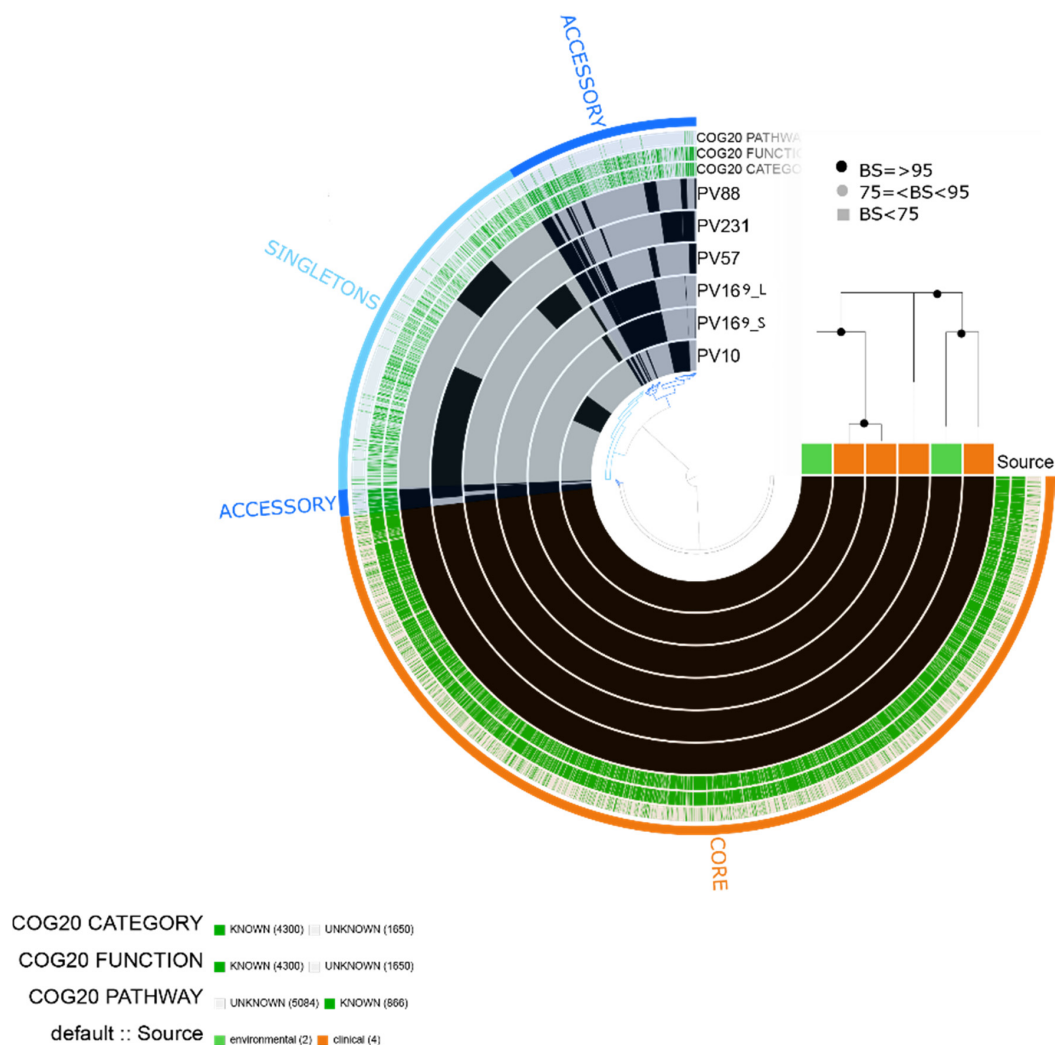


Figure 8. *Vibrio furnissii* pangenome. The labels represent in order: number of clusters of genes assigned to COG20 CATEGORY, COG20_FUNCTION and COG_20 PATHWAY and isolate source (clinical or environmental). Maximum likelihood tree based on 3411 single-copy concatenated core genes. Bootstrap supports (BS) are based on 100 replicates.

Due to the high similarity between *V. fluvialis* and *V. furnissii*, the isolates also presented a similar virulence profile. Both clinical and environmental isolates of *V. furnissii* contained genes related to the secretion system type VI (T6SS) (*hcp-2*, *vipB/mglB*) and the *luxS* gene. All the isolates contained a predicted CRP. Finally, isolate PV231 contained a predicted *sul2*, which is a resistance determinant associated with plasmids (Figure 3).

3. Discussion

This research sets a precedent in genomic surveillance of environmental and clinical isolates of *Vibrio* spp. in Colombia. We characterized the genetic potential for virulence, resistance and the presence of mobile elements in non-cholera vibrio isolates based on WGS and pangenome analyses.

In *Vibrio*, due to the high genomic plasticity, the differentiation of sister species is difficult using traditional phenotypic and biochemical methods for identification. Therefore, closely related species may be misclassified due to the similar phenotypic characteristics [21]. The calculation of average nucleotide identity (ANI) allows classifying prokaryotic species based on a similarity threshold of 95–96%. Our results showed that the isolates from each species shared ~95% ANI. Furthermore, we were able to differentiate between closely related species, such as *V. alginolyticus* / *V. diabolicus* and *V. fluvialis* / *V. furnissii*, based on the genome sequence.

We found the ST3 genotype in six isolates of *V. parahaemolyticus*. This genotype has been associated with serotype O3:K6. This pandemic clone was detected for the first time in Asia and spread through countries in America and Europe. Since 1996, it has been detected in Peru causing outbreaks mainly associated with climate change [22]. In [23], there is a report of a Colombian isolate sampled in 2016 belonging to genotype ST3. The virulence profile of this isolate was *tdh* + and *trh* -, congruent with the laboratory determination by PCR of *trh* and *tdh* of the *V. parahaemolyticus* isolates of the present study, where 100% (n = 17) were negative to *trh* and seven isolates were positive to *tdh* corresponding to ST3 (n = 6) and ST120 (n = 1) (Supplementary Table S2). The other sequence type found in this study corresponded to one isolate belonging to ST120. This ST was reported in 2009, causing an outbreak in Peru due to an introduction of Asian populations of pathogenic *V. parahaemolyticus* to the Pacific coasts of South America [24]. Here, we report the first record of ST120 in Colombia.

Regarding the *in silico* predicted virulence factors, according to [25], *tdh* has been reported in more than 90% of clinical isolates but is rarely present in environmental isolates. Therefore, it is considered a reliable virulence marker. Indeed, virulent strains exhibit certain biological characteristics that differentiate them from non-pathogenic environmental strains. This is congruent with our results since we found significant differences in the gene content and function of the accessory genome in clinical vs. environmental strains. In addition, in addition to *tdh*, the clinical strains (n = 7) presented secretion systems T3SS1 and T3SS2; the former causes cytotoxicity and the latter is essential for enterotoxicity associated with pathogenicity in humans [22]. Boyd et al., [26] suggested that T3SS2 may be an integrative element due to the presence of transposases and low content of GC (40%). This pathogenic island was named *Vibrio parahaemolyticus* island-7 (VpaI-7). We found a putative IME of ~80 kb and 39.38% (GC) in one of the isolated genotyped as ST3. According to [26], this IME could correspond to VPAI-7 and should be present in the potentially pathogenic strains of *V. parahaemolyticus* determined here due to the presence of the associated virulence genes (T3SS2 and *tdh*).

The isolates of *V. vulnificus* were classified according to the lineages proposed by [20]. Here, we found five environmental isolates belonging to Lineage 1. According to [20], this lineage comprises both clinical and environmental strains implicated in human infections due to the consumption of seafood and represent the most dangerous strains related to public health. We found one clinical and two environmental isolates associated with lineage 2. The isolates from lineage 2 have been involved in both human and fish infections. The laboratory determination of *vcgC* and *vcgE* of the *V. vulnificus* isolates of the present study

(Supplementary Table S2) showed that five of the nine isolates were positive to *vcgC* and, according to our genomic analyses, these isolates belonged to Lineage 1. Regarding Lineage 2, two of the three isolates were positive to *vcgE* and one to *vcgC*. In terms of the virulence of *V. vulnificus*, the *VvhA* and RTX cytotoxins are the most important factors; specifically, the former shares homology with cytotoxins of *V. cholerae*. Here, we found components of RTX in all the isolates of *V. vulnificus*. This cytotoxin forms pores on the host cell membrane and causes cell lysis [27]. Other virulence genes found in all the isolates were *OmpU* and *IlpA*. These genes have been classified as outer membrane proteins important in adherence, immune response and cytotoxicity [28].

Gene homology between the virulence mechanisms of *V. alginolyticus* and *V. parahaemolyticus* has been shown [29]. In fact, the *tdh* and *trh* genes in *V. alginolyticus* have been proposed as important virulence markers [30]. Nevertheless, in this study, these genes were not predicted in any of the isolates. Another homologous system to *V. parahaemolyticus* is T3SS1. We found several components of T3SS1 and *tlh* in the environmental isolates of *V. alginolyticus* and *V. diabolis*. Regarding this system, cytotoxic activity in fish has been described [31]. To date, there are no reports of *V. diabolis* causing infection; however, Song et al. [32] reported that one virulent isolate in mice was misidentified as *V. alginolyticus*. In fact, the shared virulence profile with *V. alginolyticus* may imply that *V. diabolis* could be potentially dangerous to aquatic organisms.

Chibani et al. [33] demonstrated in *V. alginolyticus* isolates that the formation of ecotypes is mainly driven by horizontal gene transfer (HGT) since the authors detected a specific gene-block cluster shared by habitat-specific strains and a closed pangenome among them. We found specific clusters of genes shared among some isolates of *V. alginolyticus*, as well as of *V. parahaemolyticus*. As future work, it would be valuable to further analyze these genomic signatures implicated in a specific lifestyle, host or habitat for epidemiological studies. Furthermore, the use of complete genomes is highly recommended in pangenomic studies.

In *V. fluvialis*, both clinical and environmental strains can express putative virulence factors; for instance, one of the most important is the hemolysine *vfh* [34]. In this study, we found genes homologous to T6SS of *V. cholera* in *V. fluvialis* and *V. furnissii*. This secretion system has been implicated in biofilm formation and cytotoxicity and contributes to bacterial pathogenicity by exerting toxic effects on host cells. In *V. fluvialis*, this secretion system and other extracellular proteases and hemolysins are regulated by quorum sensing [35]. Here, we found the *LuxS* gene in all the isolates.

ICEs occur between bacteria by conjugation, resulting in the transfer of several functions, including acquired antimicrobial resistance and virulence factors [36]. Here, the ICEs and IMEs found in the *V. fluvialis* isolates contributed genes related to flagella and TCP. Nevertheless, the contribution of this kind of elements has been related mainly to AMR resistance. For example, [34] mention that plasmid-borne *qnrVC*-like genes have been reported for quinolone resistance in *V. fluvialis* strains. This study found *QnrVCS* and *QnrSS* integron and plasmid-mediated quinolone resistance proteins in the environmental *V. fluvialis* isolate PV75 and the *V. alginolyticus* isolate PV242. Cattoir et al. [37] found this type of resistance determinant in *Vibrio splendidus* and compared it with other species of the family Vibrionaceae, concluding that Gram-negative species of aquatic environments may be reservoirs of plasmid-mediated *Qnr*-like determinants. Other resistance determinants related to plasmids and integrons, respectively, were *sul2* in *V. furnissii* (i.e., predicted in the environmental isolate PV231) and *dfrA6* in *V. fluvialis* PV75. Related to these elements, [38] found *V. cholerae* isolates with resistance genes related to trimethoprim and sulfamethoxazole (*dfrA1* and *sul2*) detected in ICE's.

Loo et al. [39] reviewed the incidence of antibiotic resistance in *Vibrio* spp., relating that recently both environmental and clinical isolates have developed resistance to ampicillin, chloramphenicol, tetracycline, streptomycin, kanamycin, trimethoprim and carbapenem. This is consistent with our findings, which showed *tet(35)* and CARB genes encoding for a tetracycline inactivation enzyme and betalactamases in all the *V. parahaemolyticus* isolates.

Furthermore, all the isolates of *V. alginolyticus*, *V. diabolicus* and *V. vulnificus* contained a gene related to tetracycline resistance. Another gene widely found in our isolates was CRP, which is a global regulator of multidrug resistance mainly described in *Escherichia coli* [19] and, according to CARD database, shows sequence variants in the resistomes of several Gram-negative bacterial species.

Due to the recent implementation of WGS in surveillance systems, there are some discrepancies between traditional in vitro and in silico methods related to taxonomical classification and gene determination. Genome-level analyses are sensitive for species determination [40]; however, given that discrepancies between in vitro and in silico determinations of antimicrobial resistance have been reported [41], in vitro validations of the presence of predicted antimicrobial genes are highly recommended.

The acquisition of mobile DNA can mediate the appearance of virulent or more virulent strains, even in formerly non-pathogenic species [42]. Despite the limitation of analyzing incomplete scaffold-level genomes, we attempted to determine the presence of mobile elements in the *Vibrio* isolates. We found functional categories assigned to [X] Mobilome: prophages, transposons including phage-related proteins, transposase, transcriptase and plasmid-related components in the accessory and singleton fractions of the six pangenomes. In addition, we found ICE's and IME's and resistance genes related to mobile elements, which support mobile elements as important sources of variation in the genus and that environmental isolates could be reservoirs of virulence and resistance genes.

Knowledge of the risk of vibriosis outbreaks due to the presence of potentially virulent and/or resistant isolates is essential for establishing management and control measures to prepare for a contingency and reduce socio-economic impacts of outbreaks. The implementation of WGS provides high resolution to characterize pathogen isolates for genomic surveillance. The results of this study are important to reinforce the intensified laboratory surveillance that is currently conducted by the Microbiology Group of the National Institute of Health of Colombia

4. Materials and Methods

4.1. Study Area and Sampling

A total of 60 isolates were analyzed in this study, including *V. parahaemolyticus* (n = 17), *V. vulnificus* (n = 9), *V. alginolyticus* (n = 8), *V. fluvialis* (n = 16), *V. diabolicus* (n = 4) and *V. furnissii* (n = 6). These isolates were recovered from water samples with three different salinity conditions (marine, estuarine and continental/fresh water) and clinical samples (stool and wound samples) (Supplementary Table S2). The isolates were obtained from two different programs in Colombia. One was the surveillance network for the conservation and protection of marine water and coasts of Colombia (*Red de Vigilancia para la conservación y Protección de las Aguas Marinas y Costeras de Colombia-REDCAM*- <http://www.invemar.org.co/redcam> accessed on 2 August 2021), which evaluated 38 monitoring stations distributed in the departments of the Archipelago of San Andres, Providence and Santa Catalina (n = 3), Magdalena (n = 16), Atlantico (n = 5), Choco (n = 3), Cauca (n = 3), Valle del Cauca (n = 4) and Nariño (n = 4). These stations were selected because they were located in sites of economic and environmental importance, such as bays, coastal lakes, swamps, low river basins and tourist beaches.

Between 0.1 and 50 were processed through membrane filtration using 0.45 µm sterile nitrocellulose filters. The filters were then transferred on a Petri dish with thiosulfate citrate bile salts sucrose (TCBS) (Oxoid™) to select for all *Vibrio* strains, as well as CHROMagar™ *Vibrio* to select for *V. cholerae*/*V. vulnificus*, *V. parahaemolyticus* and *V. alginolyticus*. The cultures were incubated at 37 ± 0.5 °C for 18 ± 4 h, then, putative *Vibrio* spp. colony-forming units (CFU) were counted. The morphotypes were replicated in new TCBS and CHROMagar™ *Vibrio* media. The isolated colonies were then transferred on a nutritive agar with 1% sodium chloride and incubated at 37 ± 0.5 °C for 18 ± 4 h. The isolates were identified by phenotypic and PCR methods and 10% of these were selected for whole genome sequencing [43].

The second program was the Intensified Cholera Surveillance program coordinated by the National Institute of Health of Colombia between 2010 and 2019, from which 14 environmental and 18 clinical samples were obtained. Sample processing was done by the Microbiology Group of National Institute of Health, following the standard protocols of the National Institute of Health [44]. The isolates were recovered by inoculation in alkaline peptone water at 8.4 pH and 37 °C for 6–8 h according to Bergey’s manual of systematic bacteriology [45]. Next, the cultures were transferred to a selective TCBS agar and incubated at 37 °C between 18 and 24 h, according to the *Vibrio cholerae* surveillance guide of the Microbiology group of the National Institute of Health [46]. The isolates were identified by phenotypic and PCR methods and 10% of these were selected for whole genome sequencing [43].

4.2. Whole Genome Sequencing, De Novo Assembly and Annotation

All isolates were grown in brain heart infusion (BHI) Agar overnight at 37 °C. DNA extraction was performed by boiling at 100 °C for 15 minutes from a suspension of 2–3 bacterial colonies in 100 µL of Tris HCl 0.1N to lyse the bacteria. The suspension was then centrifuged for 2 min at 12,000 rpm to precipitate the cell debris, while the supernatant was collected in a new tube for use as template and stored at 20 °C until use. DNA was extracted from each isolate using the QIAamp DNA Mini Kit (Qiagen®, Hilden, Germany) following the protocol and recommendations of the manufacturer. DNA integrity and quantity were assessed on a Nanodrop UV-vis spectrophotometer (Thermo Fisher Scientific™, Waltham, MA, USA) and 1% agarose gel electrophoresis visualized with ethidium bromide in a ChemiDoc XRS photodocumenter (BioRad®, Hercules, CA, USA). Paired-end libraries with 500 bp insert size and 250 bp read length were sequenced on an Illumina HiSeq platform, generating a raw data output of ~1 Gb per sample (~190X depth).

Prior to the genome assembly, the taxonomic assignment of each isolate was confirmed by comparing the sequenced reads against the RefSeq Bacteria database using KAIJU v1.7.3 [47]. The results were visualized with KRONA v2.7.1 [48]. We found that four isolates were contaminated with reads from *Enterobacter* spp. or *Morganella* spp.; therefore, read decontamination was done using BLAST against local databases containing all *Enterobacter* (txid: 547) or *Morganella* (txid:581) sequences available in the Nucleotide database of the National Center for Biotechnology Information (NCBI) (version August 2020). Contaminated reads were searched for using ncbi-blast v2.10.1 and non-contaminated reads were filtered and extracted using Seqtk v1.2 [49] and custom bash scripts.

Genome assembly was done according to the de novo reference-guided assembly pipeline developed and described by [50], which was modified to use SPADES v3.6.2 de novo assembler [51]. This pipeline was developed to reduce the complexity of the genomic assembly using a conspecific reference genome, without introducing bias. The pipeline spans read quality control and trimming, contig assembly and scaffolding. Briefly, raw data quality control was assessed using FastQC v0.11.4 [52] and visualized with MultiQC v1.9 [53]. Read trimming and quality filtering were done using Trimmomatic v0.36 [54] with the following parameters: HEADCROP:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN: <<100, 120 or 150>>. The minimum length was adjusted for each sample to discard the lowest number of reads possible based on overall read quality. Next, homologue regions between the reference genome and target genome were identified by read mapping using Bowtie2 v2.3.4 [55], which allowed constructing superblocks or conserved genomic regions using SAMtools v1.5 [56], BEDtools v2.25 [57] and BCFtools v1.3.1 [58]. Mapped reads were split according to the superblocks using Seqtk v1.2 and assembled separately with SPADES v3.6.2; furthermore, unmapped reads were also assembled independently. The contigs were merged into non-redundant supercontigs against the reference genome using AMOScmp v3.1.0 [59]. Finally, error correction was done with the Genome Analysis Toolkit (GATK) version [60,61] and the original reads and scaffolding was performed using MUMMER v3.23 [62] and SOAPdenovo2 v240 [63]. Each isolate was

assembled using the genome reference of its species (Supplementary Table S3). Assembly quality and completeness were assessed using QUAST v4.5 [64] and BUSCO v4.0.6 [65].

Whole genome average nucleotide identity was calculated for the 60 isolates using FastANI [66]. The output was converted to a symmetrical matrix and plotted in a heatmap with hierarchical clustering in R v3.6.1 ggplot2 [67].

4.3. Characterization of the Most Important Species in Public Health

To determine the genetic variants (sequence types) of the most important species in public health (*V. parahaemolyticus* and *V. vulnificus*), MLST v2.10 [68] was used. For this, the allelic profile of housekeeping genes was obtained and a ST was signed according to the defined typing scheme for each species in PubMLST database [69]. The isolates were submitted to the database (Supplementary Tables S4 and S5).

The lineages of *V. vulnificus* were determined by a phylogenetic analysis using a database of 80 closed genomes with a defined poblational structure obtained from Roig et al. [20]. For this, core genes were aligned with Parsnp [70] and the phylogenetic tree was inferred using IQTREE v1.6.12 [71], with 1000 bootstrap replicates and the GTR + I + G model. The tree was visualized with FigTree [72]. FastANI was used to calculate the whole genome average nucleotide identity of the 80 *V. vulnificus* isolates.

4.4. Pangenome and Phylogenomic Analyses

The Anvi'o workflow v2.1.0 [73] was used to build a pangenome for each species. For this, the genomes were annotated using the Clusters of Orthologues Genes (COG) database, according to the categories COG20_FUNCTION, COG20_PATHWAY and COG20_CATEGORY [74]. Three bins related to the core genes (genes present in 99–100% of the isolates), accessory genes (15–95%) and singletons (0–15%) were created and a summary of the annotation for each bin was obtained. To investigate the differences between the genomic content of accessory genomes between groups of clinical and environmental isolates within each pangenome, we created layers of additional data (e.g., source) and conducted analyses to determine significantly enriched categories (p value < 0.05) between these groups according to [75]. The core genes for each species were obtained from Anvi'o and used to construct a maximum likelihood (ML) phylogenetic tree using IQTREE v1.6.12 with the LG amino acid evolutionary model, I+G sites distribution and an ultrafast bootstrap of 100 replicates. The pangenome was ordered according to the phylogenetic tree obtained.

4.5. Identification of Virulence Factors and Antimicrobial Resistance Determinants

To identify the virulence and antibiotic resistance genes, ABRicate [76] was used with specific databases, such as NCBI AMRFinderPlus [77] and the Comprehensive Antibiotic Resistance Database (CARD) [78] for resistance determinants and the Virulence Factor Database (VFDB) [79] for virulence factors. Hits were filtered at thresholds of 80% identity and 80% coverage. As input, we used the CDS file of each genome. The predicted virulence and resistance genes were grouped according to the identity percentage and visualized with heatmap (R ggplot). This allowed us to observe the main virulence genes shared between species.

ICEfinder [36] was used to determine the presence of putative integrative and conjugative elements (ICE) and integrative and mobile elements (IME) in the assembled genomes. The association of these elements with virulence and resistance determinants was analyzed by extracting the predicted ICE or IME sequence and searching for virulence and resistance genes using VFDB (e-value = 0.01) and CARD.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/pathogens10121605/s1>, Supplementary Table S1: Characteristics of the obtained assemblies of the six *Vibrio* species analyzed in this study. The characteristics represent statistics and metrics resulted from the quality control, preprocessing, assembly and annotation processes, Supplementary Table S2: Sample information of the *Vibrio* spp. isolates analyzed in this study, Supplementary Table S3: Reference *Vibrio* spp. genomes used for reference-guided de novo genome assemblies, Supplementary Table S4:

MLST results of *Vibrio parahaemolyticus*, Supplementary Table S5: MLST results of *Vibrio vulnificus*, Figure S1: Maximum likelihood tree of 80 *V. vulnificus* isolates. BS are based on 1000 replicates. The nine isolates of the present study are marked with orange, Figure S2: Whole-genome Average Nucleotide Identity of *V. vulnificus* isolates.

Author Contributions: Conceptualization, A.P.-D., A.G.-M., M.W.-R., L.J.V.-A., J.A.-V. and J.M.-U.; methodology, A.P.-D., A.G.-M., M.W.-R., G.T.R.-R. and J.M.-U.; formal analysis, A.G.-M., A.P.-D.; investigation, A.P.-D., A.G.-M., M.W.-R.; writing—original draft preparation, A.P.-D., A.G.-M.; writing—review and editing, A.P.-D., A.G.-M., J.A.-V., L.J.V.-A., T.C.-M., G.T.R.-R., P.D.-G., M.W.-R., J.M.-U.; visualization, A.P.-D.; supervision, A.G.-M., J.A.-V., M.W.-R., G.T.R.-R. and J.M.-U.; project administration, M.W.-R. All authors have read and agreed to the published version of the manuscript.

Funding: This research and the APC were funded by Colciencias project number 210471250745 (Title: *Vibrio* spp. in water reserves in Colombia as potential agents of cholera and vibriosis), contract N°-RC-862-2017.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committee of the Instituto Nacional de Salud (Comité de ética y metodologías de investigación CEMIN) according to code 10-2017 and act No.8 of 2 June 2017.

Informed Consent Statement: Patient consent was waived because surveillance programs for public health events do not require informed consents and cholera surveillance is an event of mandatory notification according to the international sanitary code; therefore, all suspicious samples must be collected and processed immediately without requiring informed consent.

Data Availability Statement: The genomic data of the *Vibrio* isolates were deposited under BioProject ID PRJNA754786 with BioSample accessions SAMN20802061, SAMN20931810, SAMN20804964 - SAMN20804979, SAMN20805029 - SAMN20805070.

Acknowledgments: We acknowledge the departmental public health laboratories participating in the intensified cholera surveillance program and all research of the *Instituto Nacional de Salud*, who indirectly contributed to this study by conducting field work to collect the samples and laboratory work to process and recover the isolates. We acknowledge the support of the monitoring program of the Red de Vigilancia para la Conservación y Protección de las Aguas Marinas y Costeras de Colombia- REDCAM (Surveillance Network for the Conservation and Protection of Marine and Coastal Waters of Colombia) and the personnel who carried out sampling. We specially thank Lucy Angeline Montaña for analyzing and processing the *Vibrio* spp. isolates recovered in the cholera surveillance program. A special acknowledgment to Silvia Yasbleidy Narvaez Florez (Universidad ECCI. Grupo de Gestión Ambiental y Desarrollo Sostenible. Bogotá, Colombia) for her constant support throughout the entire project and Carmen Amaro (Full Professor, University of Valencia, Spain) for her contribution to the *Vibrio vulnificus* analyses.

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

References

1. Baker-Austin, C.; Oliver, J.D.; Alam, M.; Ali, A.; Waldor, M.K.; Qadri, F.; Martinez-Urtaza, J. *Vibrio* spp. infections. *Nat. Rev. Dis. Primers* **2018**, *4*, 1–19. [CrossRef]
2. OMS. Cólera. 2019. Available online: <https://www.who.int/es/news-room/fact-sheets/detail/cholera> (accessed on 25 January 2021).
3. CDC. *Vibrio* Species Causing Vibriosis. 2020. Available online: <https://www.cdc.gov/vibrio/index.html> (accessed on 2 February 2021).
4. Mok, J.S.; Ryu, A.; Kwon, J.Y.; Kim, B.; Park, K. Distribution of *Vibrio* species isolated from bivalves and bivalve culture environments along the Gyeongnam coast in Korea: Virulence and antimicrobial resistance of *Vibrio parahaemolyticus* isolates. *Food Control* **2019**, *106*, 106697. [CrossRef]
5. Baker-Austin, C.; Trinanés, J.; González-Escalona, N.; Martínez-Urtaza, J. Non-Cholera Vibrios: The Microbial Barometer of Climate Change. *Trends Microbiol.* **2017**, *25*, 76–84. [CrossRef]
6. Hernández-Flórez, C.E.; Cáceres-Manrique, F.D.M. Cólera, ¿se aproxima una nueva pandemia? *Med. UIS* **2014**, *27*, 67–83. Available online: http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0121-03192014000200008 (accessed on 12 January 2021).

7. Escobar, L.E.; Ryan, S.J.; Stewart-Ibarra, A.M.; Finkelstein, J.L.; King, C.A.; Qiao, H.; Polhemus, M.E. A global map of suitability for coastal *Vibrio cholerae* under current and future climate conditions. *Acta Trop.* **2015**, *149*, 202–211. [CrossRef] [PubMed]
8. Raszl, S.M.; Froelich, B.A.; Vieira, C.R.W.; Blackwood, A.D.; Noble, R.T. *Vibrio parahaemolyticus* and *Vibrio vulnificus* in South America: Water, seafood and human infections. *J. Appl. Microbiol.* **2016**, *121*, 1201–1222. [CrossRef]
9. López, M. Acciones de Vigilancia Intensificada de Cólera Ante Posible Reintroducción en Los Componentes de Vigilancia Epidemiológica y Laboratorio, Colombia, 2011–2012. 2012. Available online: <https://www.ins.gov.co/buscador-eventos/IQEN/IQEN%20vol%2018%202013%20num%2024.pdf> (accessed on 11 November 2021).
10. Instituto Nacional de Salud. Vigilancia Fenotípica y Genotípica de *Vibrio Cholerae* 2010–2013. 1–11. 2013. Available online: <https://www.ins.gov.co/buscador/Informacin%20de%20laboratorio/Vigilancia%20C%3%B3lera%20Colombia%202013.pdf> (accessed on 10 October 2021).
11. Mohamad, N.; Amal MN, A.; Saad, M.Z.; Yasin, I.S.M.; Zulkipli, N.A.; Mustafa, M.; Nasruddin, N.S. Virulence-associated genes and antibiotic resistance patterns of *Vibrio* spp. isolated from cultured marine fishes in Malaysia. *BMC Vet. Res.* **2019**, *15*, 1–13. [CrossRef]
12. Bruto, M.; Labreuche, Y.; James, A.; Piel, D.; Chenivesse, S.; Petton, B.; Polz, M.F.; Le Roux, F. Ancestral gene acquisition as the key to virulence potential in environmental *Vibrio* populations. *ISME J.* **2018**, *12*, 2954–2966. [CrossRef] [PubMed]
13. Klein, S.; Pipes, S.; Lovell, C.R. Occurrence and significance of pathogenicity and fitness islands in environmental vibrios. *AMB Express* **2018**, *8*, 177. [CrossRef]
14. Nathamuni, S.; Jangam, A.K.; Katneni, V.K.; Selvaraj, A.; Krishnan, K.; Kumar, S.; Avunje, S.; Balasubramaniam, S.; Grover, M.; Alavandi, S.V.; et al. Insights on genomic diversity of *Vibrio* spp. through Pan-genome analysis. *Ann. Microbiol.* **2019**, *69*, 1547–1555. [CrossRef]
15. Mok, J.S.; Ryu, A.; Kwon, J.Y.; Park, K.; Shim, K.B. Abundance, antimicrobial resistance, and virulence of pathogenic *Vibrio* strains from molluscan shellfish farms along the Korean coast. *Mar. Pollut. Bull.* **2019**, *149*, 110559. [CrossRef]
16. Thompson, F.L.; Iida, T.; Swings, J. Biodiversity of Vibrios. *J. Clin. Microbiol.* **2004**, *68*, 403–431. [CrossRef]
17. Castillo, D.; Kauffman, K.; Hussain, F.; Kalatzis, P.; Rørbo, N.; Polz, M.F.; Middelboe, M. Widespread distribution of prophage-encoded virulence factors in marine *Vibrio* communities. *Sci. Rep.* **2018**, *8*, 2–10. [CrossRef]
18. Dobrindt, U.; Hochhut, B.; Hentschel, U.; Hacker, J. Genomic islands in pathogenic and environmental microorganisms. *Nat. Rev. Microbiol.* **2004**, *2*, 414–424. [CrossRef]
19. Nishino, K.; Senda, Y.; Yamaguchi, A. CRP regulator modulates multidrug resistance of *Escherichia coli* by repressing the *mdtEF* multidrug efflux genes. *J. Antibiot.* **2008**, *61*, 120–127. [CrossRef]
20. Roig, F.J.; González-Candelas, F.; Sanjuán, E.; Fouz, B.; Feil, E.J.; Llorens, C.; Baker-Austin, C.; Oliver, J.D.; Danin-Poleg, Y.; Gibas, C.J.; et al. Phylogeny of *Vibrio vulnificus* from the analysis of the core-genome: Implications for intra-species taxonomy. *Front. Microbiol.* **2018**, *8*, 1–13. [CrossRef] [PubMed]
21. Thompson, C.C.; Vicente, A.C.P.; Souza, R.C.; Vasconcelos, A.T.R.; Vesth, T.; Alves, N.; Ussery, D.W.; Iida, T.; Thompson, F.L. Genomic taxonomy of vibrios. *BMC Evol. Biol.* **2009**, *9*, 1–16. [CrossRef] [PubMed]
22. Gavilan, R.G.; Zamudio, M.L.; Martínez-Urtaza, J. Molecular Epidemiology and Genetic Variation of Pathogenic *Vibrio parahaemolyticus* in Peru. *PLoS Negl. Trop. Dis.* **2013**, *7*, e2210. [CrossRef]
23. Baker-Austin, C.; Jenkins, C.; Dadzie, J.; Mestanza, O.; Delgado, E.; Powell, A.; Bean, T.; Martínez-Urtaza, J. Genomic epidemiology of domestic and travel-associated *Vibrio parahaemolyticus* infections in the UK, 2008–2018. *Food Control* **2020**, *115*, 107244. [CrossRef]
24. Gonzalez-Escalona, N.; Gavilan, R.G.; Toro, M.; Zamudio, M.L.; Martínez-Urtaza, J. Outbreak of *Vibrio parahaemolyticus* sequence type 120, Peru, 2009. *Emerg. Infect. Dis.* **2016**, *22*, 1235–1237. [CrossRef] [PubMed]
25. Velazquez-Roman, J.; León-Sicairos, N.; Hernández-Díaz, L.D.J.; Canizalez-Roman, A. Pandemic *Vibrio parahaemolyticus* O3: K6 on the American continent. *Front. Cell. Infect. Microbiol.* **2014**, *3*, 1–14. [CrossRef]
26. Boyd, E.F.; Cohen, A.L.V.; Naughton, L.M.; Ussery, D.W.; Binnewies, T.T.; Stine, O.C.; Parent, M.A. Molecular analysis of the emergence of pandemic *Vibrio parahaemolyticus*. *BMC Microbiol.* **2008**, *8*, 1–14. [CrossRef] [PubMed]
27. Kim, Y.R.; Lee, S.E.; Kook, H.; Yeom, J.A.; Na, H.S.; Kim, S.Y.; Chung, S.S.; Choy, H.E.; Rhee, J.H. *Vibrio vulnificus* RTX toxin kills host cells only after contact of the bacteria with host cells. *Cell. Microbiol.* **2008**, *10*, 848–862. [CrossRef] [PubMed]
28. Jones, M.K.; Oliver, J.D. *Vibrio vulnificus*: Disease and pathogenesis. *Infect. Immun.* **2009**, *77*, 1723–1733. [CrossRef] [PubMed]
29. Xie, Z.Y.; Hu, C.Q.; Chen, C.; Zhang, L.P.; Ren, C.H. Investigation of seven *Vibrio* virulence genes among *Vibrio alginolyticus* and *Vibrio parahaemolyticus* strains from the coastal mariculture systems in Guangdong, China. *Lett. Appl. Microbiol.* **2005**, *41*, 202–207. [CrossRef]
30. Hernández-Robles, M.F.; Álvarez-Contreras, A.K.; Juárez-García, P.; Natividad-Bonifacio, I.; Curiel-Quesada, E.; Vázquez-Salinas, C.; Quiñones-Ramírez, E.I. Virulence factors and antimicrobial resistance in environmental strains of *Vibrio alginolyticus*. *Int. Microbiol.* **2016**, *19*, 191–198. [CrossRef]
31. Osorio, C.R. T3SS effectors in Vibrios: Homology in sequence, diversity in biological functions? *Virulence* **2018**, *9*, 721–723. [CrossRef]
32. Song, J.; Liu, X.; Wu, C.; Zhang, Y.; Fan, K.; Zhang, X.; Wei, Y. Isolation, identification and pathogenesis study of *Vibrio diabolicus*. *Aquaculture* **2021**, *533*, 736043. [CrossRef]
33. Chibani, C.M.; Roth, O.; Liesegang, H.; Wendling, C.C. Genomic variation among closely related *Vibrio alginolyticus* strains is located on mobile genetic elements. *BMC Genom.* **2020**, *21*, 1–14. [CrossRef]

34. Ramamurthy, T.; Chowdhury, G.; Pazhani, G.P.; Shinoda, S. *Vibrio fluvialis*: An emerging human pathogen. *Front. Microbiol.* **2014**, *5*, 1–8. [CrossRef]
35. Liu, X.; Pan, J.; Gao, H.; Han, Y.; Zhang, A.; Huang, Y.; Liu, P.; Kan, B.; Liang, W. CqsA/LuxS-HapR Quorum sensing circuit modulates type VI secretion system VftT6SS2 in *Vibrio fluvialis*. *Emerg. Microbes Infect.* **2021**, *10*, 589–601. [CrossRef] [PubMed]
36. Liu, M.; Li, X.; Xie, Y.; Bi, D.; Sun, J.; Li, J.; Tai, C.; Deng, Z.; Ou, H.Y. ICEberg 2.0: An updated database of bacterial integrative and conjugative elements. *Nucleic Acids Res.* **2019**, *47*, D660–D665. [CrossRef] [PubMed]
37. Cattoir, V.; Poirel, L.; Mazel, D.; Soussy, C.J.; Nordmann, P. *Vibrio splendidus* as the source of plasmid-mediated QnrS-like quinolone resistance determinants. *Antimicrob. Agents Chemother.* **2007**, *51*, 2650–2651. [CrossRef]
38. Sarkar, A.; Morita, D.; Ghosh, A.; Chowdhury, G.; Mukhopadhyay, A.K.; Okamoto, K.; Ramamurthy, T. Altered Integrative and Conjugative Elements (ICEs) in Recent *Vibrio cholerae* O1 Isolated From Cholera Cases, Kolkata, India. *Front. Microbiol.* **2019**, *10*, 1–13. [CrossRef]
39. Loo, K.Y.; Letchumanan, V.; Law, J.W.F.; Pusparajah, P.; Goh, B.H.; Ab Mutalib, N.S.; He, Y.W.; Lee, L.H. Incidence of antibiotic resistance in *Vibrio* spp. *Rev. Aquac.* **2020**, *12*, 2590–2608. [CrossRef]
40. Paul, B.; Dixit, G.; Murali, T.S.; Satyamoorthy, K. Genome Based Taxonomic Classification. *Genome* **2019**, *62*, 1–17. [CrossRef]
41. Lepuschitz, S.; Baron, S.; Larvor, E.; Granier, S.A.; Pretzer, C.; Mach, R.L.; Farnleitner, A.H.; Ruppitsch, W.; Pleininger, S.; Indra, A.; et al. Phenotypic and Genotypic Antimicrobial Resistance Traits of *Vibrio cholerae* Non-O1/Non-O139 Isolated From a Large Austrian Lake Frequently Associated With Cases of Human Infection. *Front. Microbiol.* **2019**, *10*, 1–9. [CrossRef] [PubMed]
42. Gennari, M.; Ghidini, V.; Caburlootto, G.; Lleo, M.M. Virulence genes and pathogenicity islands in environmental *Vibrio* strains nonpathogenic to humans. *FEMS Microbiol. Ecol.* **2012**, *82*, 563–573. [CrossRef]
43. Sánchez, L.P.; Martínez, M.; León, T.; Córdoba, T.; Díaz, P.; Calvo, M.; Montaña, A.; Escandón, P.; Narváez, S.; Vivas, J.; et al. Desarrollo e Implementación de una PCR Multiplex Para la Detección de Cuatro Especies de *Vibrio* spp. *Biomédica* **2019**, *39*. Available online: <https://revistabiomedica.org/index.php/biomedica/issue/download/171/58> (accessed on 8 September 2021).
44. Instituto Nacional de Salud. Manual de Procedimientos Para la Toma, Conservación y Envío de Muestras al Laboratorio Nacional de Referencia. Dirección Redes en Salud Pública. Available online: https://www.ins.gov.co/Direcciones/RedesSaludPublica/DocumentosdeInteresSRNL/Manual_toma_envio_muestras_INS-2019.pdf (accessed on 20 September 2021).
45. Williams; Wilkins. Bergey’s Manual of Systematic Bacteriology. In *The Proteobacteria*, 2nd ed.; Garrity, G., Ed.; Springer: Berlin/Heidelberg, Germany, 1984; Volume 2.
46. Instituto Nacional de Salud Dirección Redes en Salud Pública. Guía Para la Vigilancia por Laboratorio de *Vibrio Cholerae* (p. 17). Dirección Redes en Salud Pública; 2017. Available online: <https://www.ins.gov.co/buscador-eventos/Informacin%20de%20laboratorio/Gu%C3%ADa%20para%20la%20vigilancia%20por%20laboratorio%20de%20Vibrio%20Cholerae.pdf> (accessed on 11 November 2021).
47. Menzel, P.; Ng, K.L.; Krogh, A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat. Commun.* **2016**, *7*, 11257. [CrossRef]
48. Ondov, B.D.; Bergman, N.H.; Phillippy, A.M. Interactive metagenomic visualization in a Web browser. *BMC Bioinform.* **2011**, *12*, 385. [CrossRef]
49. Li, H. Seqtk. 2013. Available online: <https://github.com/lh3/seqtk> (accessed on 17 August 2021).
50. Lischer, H.E.L.; Shimizu, K.K. Reference-guided *de novo* assembly approach improves genome reconstruction for related species. *BMC Bioinform.* **2017**, *18*, 1–12. [CrossRef]
51. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Pribelski, A.D.; et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol. A J. Comput. Mol. Cell Biol.* **2012**, *19*, 455–477. [CrossRef] [PubMed]
52. Andrews; Simon. FastQC: A Quality Control Tool for High Throughput Sequence Data. 2017. Available online: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc> (accessed on 10 January 2021).
53. Ewels, P.; Magnusson, M.; Lundin, S.; Käller, M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **2016**, *32*, 3047–3048. [CrossRef] [PubMed]
54. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef] [PubMed]
55. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [CrossRef] [PubMed]
56. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef]
57. Quinlan, A.R.; Hall, I.M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26*, 841–842. [CrossRef]
58. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **2011**, *27*, 2987–2993. [CrossRef]
59. Pop, M.; Phillippy, A.; Delcher, A.L.; Salzberg, S.L. Comparative genome assembly. *Brief. Bioinform.* **2004**, *5*, 237–248. [CrossRef]
60. Van der Auwera, G.A.; Carneiro, M.O.; Hartl, C.; Poplin, R.; Del Angel, G.; Levy-Moonshine, A.; Jordan, T.; Shakir, K.; Roazen, D.; Thibault, J.; et al. From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinform.* **2013**, *43*, 11.10.1–11.10.33. [CrossRef]

61. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernysky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **2010**, *20*, 1297–1303. [[CrossRef](#)]
62. Delcher, A.L.; Salzberg, S.L.; Phillippy, A.M. Using MUMmer to identify similar regions in large sequence sets. *Curr. Protoc. Bioinform.* **2003**, 10.3.1–10.3.18. [[CrossRef](#)] [[PubMed](#)]
63. Luo, R.; Liu, B.; Xie, Y.; Li, Z.; Huang, W.; Yuan, J.; He, G.; Chen, Y.; Pan, Q.; Liu, Y.; et al. SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *GigaScience* **2012**, *1*, 18. [[CrossRef](#)] [[PubMed](#)]
64. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* **2013**, *29*, 1072–1075. [[CrossRef](#)]
65. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [[CrossRef](#)]
66. Jain, C.; Rodriguez-R, L.M.; Phillippy, A.M.; Konstantinidis, K.T.; Aluru, S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* **2018**, *9*, 1–8. [[CrossRef](#)]
67. Wickham, H. Ggplot2. *Wiley Interdiscip. Rev. Comput. Stat.* **2011**, *3*, 180–185. [[CrossRef](#)]
68. GitHub—Tseemann/Mlst: Scan Contig Files against PubMLST Typing Schemes. Available online: <https://github.com/tseemann/mlst> (accessed on 18 August 2021).
69. Jolley, K.A.; Bray, J.E.; Maiden, M.C.J. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res.* **2018**, *3*, 1–20. [[CrossRef](#)] [[PubMed](#)]
70. Treangen, T.J.; Ondov, B.D.; Koren, S.; Phillippy, A.M. The harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol.* **2014**, *15*, 1–15. [[CrossRef](#)] [[PubMed](#)]
71. Nguyen, L.T.; Schmidt, H.A.; Von Haeseler, A.; Minh, B.Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [[CrossRef](#)]
72. Release FigTree v1.4.4 · Rambaut/Figtree · GitHub. Available online: <https://github.com/rambaut/figtree/releases/tag/v1.4.4> (accessed on 20 August 2021).
73. Eren, A.M.; Esen, O.C.; Quince, C.; Vineis, J.H.; Morrison, H.G.; Sogin, M.L.; Delmont, T.O. Anvi'o: An advanced analysis and visualization platform for omics data. *PeerJ* **2015**, *2015*, 1–29. [[CrossRef](#)] [[PubMed](#)]
74. Tatusov, R.L.; Galperin, M.Y.; Natale, D.A.; Koonin, E.V. The COG database: A tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **2000**, *28*, 33–36. [[CrossRef](#)]
75. Shaiber, A.; Willis, A.D.; Delmont, T.O.; Roux, S.; Chen, L.X.; Schmid, A.C.; Yousef, M.; Watson, A.R.; Lolans, K.; Esen, Ö.C.; et al. Functional and genetic markers of niche partitioning among enigmatic members of the human oral microbiome. *BioRxiv* **2020**, 1–35. [[CrossRef](#)] [[PubMed](#)]
76. GitHub—Tseemann/Abricate: Mass Screening of Contigs for Antimicrobial and Virulence Genes. Available online: <https://github.com/tseemann/abricate> (accessed on 18 August 2021).
77. Feldgarden, M.; Brover, V.; Haft, D.H.; Prasad, A.B.; Slotta, D.J.; Tolstoy, I.; Tyson, G.H.; Zhao, S.; Hsu, C.H.; McDermott, P.F.; et al. Validating the AMRFINDER tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. *Antimicrob. Agents Chemother.* **2019**, *63*. [[CrossRef](#)]
78. Jia, B.; Raphenya, A.R.; Alcock, B.; Waglechner, N.; Guo, P.; Tsang, K.K.; Lago, B.A.; Dave, B.M.; Pereira, S.; Sharma, A.N.; et al. CARD 2017: Expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **2017**, *45*, D566–D573. [[CrossRef](#)] [[PubMed](#)]
79. Liu, B.; Zheng, D.; Jin, Q.; Chen, L.; Yang, J. VFDB 2019: A comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res.* **2019**, *47*, D687–D692. [[CrossRef](#)]