



# Insights on genomic diversity of *Vibrio* spp. through Pan-genome analysis

Suganya Nathamuni<sup>1</sup> · Ashok Kumar Jangam<sup>1</sup> · Vinaya Kumar Katneni<sup>1</sup> · Ashok Selvaraj<sup>1</sup> · Karthic Krishnan<sup>1</sup> · Sujeet Kumar<sup>2</sup> · Satheesha Avunje<sup>2</sup> · Sivamani Balasubramaniam<sup>1</sup> · Monendra Grover<sup>3</sup> · Shankar V. Alavandi<sup>2</sup> · Vijayan K. Koyadan<sup>1</sup>

Received: 22 July 2019 / Accepted: 20 November 2019 / Published online: 7 December 2019

© The Author(s) 2019

## Abstract

**Purpose** The aquaculture sector is a major contributor to the economic and nutritional security for a number of countries. India's total seafood exports for the year 2017–2018 accounted for US\$ Million 7082. One of the major setbacks in this sector is the frequent outbreaks of diseases often due to bacterial pathogens. Vibriosis is one of the major diseases caused by bacteria of *Vibrio* spp., causing significant economic loss to the aquaculture sector. The objective of this study was to understand the genetic composition of *Vibrio* spp.

**Methods** Thirty-five complete genomes were downloaded from GenBank comprising seven vibrio species, namely, *Vibrio alginolyticus*, *V. anguillarum*, *V. campbellii*, *V. harveyi*, *V. furnissii*, *V. parahaemolyticus*, and *V. vulnificus*. Pan-genome analysis was carried out with coding sequences (CDS) generated from all the *Vibrio* genomes. In addition, genomes were mined for genes coding for toxin-antitoxin systems, antibiotic resistance, genomic islands, and virulence factors.

**Results** Results revealed an open pan-genome comprising of 2004 core, 8249 accessory, and 6780 unique genes. Downstream analysis of genomes and the identified unique genes resulted in 312 antibiotic resistance genes, 430 genes coding for toxin and antitoxin systems along with 4802, and 4825 putative virulent genes from genomic island regions and unique gene sets, respectively.

**Conclusion** Pan-genome and other downstream analytical procedures followed in this study have the potential to predict strain-specific genes and their association with habitat and pathogenicity.

**Keywords** Genomic island · Pan-genome · Phylogeny · *Vibrio* · Virulent genes

## Introduction

Aquaculture, the farming of aquatic animals such as finfishes, crustaceans, mollusks, and seaweed, has the potential to contribute significantly to the society and economy. The annual

production of the year 2016, accounting for 80 million tons worth USD 231.6 billion, with 19.3 million people employed worldwide, signifies the production potentials and importance of the aquaculture sector (FAO 2018). One of the major impediments to the aquaculture sector is the diseases caused by various factors, especially diseases caused by infectious agents such as bacteria and viruses. Vibriosis is one such disease caused by *Vibrio* species, significantly affecting the productivity of aquaculture. The incidences of human infections caused by *Vibrio* spp. are also significant (Kim and Lee 2017; Ina-Salwany et al. 2019).

*Vibrios* are naturally occurring gram-negative, facultative anaerobes found in brackish and marine water. These are opportunistic pathogens and cause disease in finfish and shellfish when the host species are under stress (Ina-Salwany et al. 2019). They are also found to cause human illness such as septicemia, gastroenteritis, and necrotizing wounds (Ina-Salwany et al. 2019). More than 100 *Vibrio* species have been

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s13213-019-01539-7>) contains supplementary material, which is available to authorized users.

✉ Ashok Kumar Jangam  
ashok@ciba.res.in

<sup>1</sup> Nutrition Genetics and Biotechnology Division, ICAR-Central Institute of Brackishwater Aquaculture, Chennai, Tamil Nadu, India

<sup>2</sup> Aquatic Animal Health and Environment Division, ICAR-Central Institute of Brackishwater Aquaculture, Chennai, Tamil Nadu, India

<sup>3</sup> Division of Computer Applications, ICAR-Indian Agricultural Statistics Research Institute, New Delhi, India

identified comprising both virulent and non-virulent strains. At times the non-virulent strains have also found to cause chronic infections due to acquisition of virulence factors (Le Roux et al. 2015). Some of the species known to cause Vibriosis in finfish and shellfish are *V. parahaemolyticus*, *V. vulnificus*, *V. campbellii*, and *V. alginolyticus*. Recent studies reported presence of virulent genes, viz., *chiA*, *vhpA*, *luxR*, *flaC*, *tdh*, and *tlh* (Mohamad et al. 2019; Mok et al. 2019; Soto-Rodriguez et al. 2019) antibiotic resistance to beta-lactam, tetracycline, sulfamethizole, etc., and associated genes, viz., *aadA16*, *blaTEM1B*, *ARR3*, *sull1*, *tet(B)*, *tet(E)*, *tet(R)*, *dfrA27*, *Zeo*, *qacEA*, and RND efflux pump (Silvester et al. 2015; Mohamad et al. 2019; Mok et al. 2019; Soto-Rodriguez et al. 2019) and toxin-antitoxin systems especially the *Phd/doc* and *ccd<sub>vfi</sub>* (Guérout et al. 2013; Lloyd et al. 2019), *HigBA*, *HipBA*, *ParDE*, etc. (Szekeres et al. 2007; Guérout et al. 2013), present within and outside the superintegrons of *Vibrio* species.

Recent advances in sequencing technologies and public domain repositories enabled researchers to access large volumes of genomic data, which includes whole genome sequences. Bioinformatics analysis would provide the gene composition and their functions of the genomes. This structural and functional information would pave the way for inter and intra species comparison of a large number of genomes. In GenBank 181 complete genome sequences of *Vibrio* spp. are available as on 21 Feb 2019. *Vibrios* have two chromosomes with a significant genetic diversity shaped by recombination and horizontal gene transfer (Baker-Austin et al. 2018). Comparative analyses of different species from the same genus helps us in understanding the evolution, gene repertoire, and the minimal genome size which brings about the concept of pan-genome into the picture. In this study, pan-genome analysis was performed to understand the gene profiles and evolutionary patterns in the genomes of *Vibrio* spp. Further, attempts were made to elucidate information on toxin-antitoxin systems (TA systems), virulence, and antibiotic resistance of *Vibrio* spp.

## Materials and methods

### Genome retrieval

Thirty-five genome sequences of seven *Vibrio* spp., comprising three strains of *V. alginolyticus* (strains ATCC 17749; ATCC 33787; ZJ-T), four strains of *V. anguillarum* (strains 90-11-286; 775; M3; NB10), three strains of *V. campbellii* (strains ATCC-BB120; LB102; LMB29), one strain of *V. furnissii* NCTC 11218, three strains of *V. harveyi* (strains ATCC 33843; ATCC 43516; QT520), 13 strains of *V. parahaemolyticus* (strains ATCC 17802; BB22OP; CDC K4557; CHN25; FDA R31; FORC\_004; FORC\_006;

FORC\_008; FORC\_014; FORC\_018; FORC\_023; RIMD 2210633; UCM-V493), eight strains of *V. vulnificus* (strains 93U204; ATL 6-1306; CMCP6; FORC\_009; FORC\_016; FORC\_017; MO6-24/O; YJ016) were downloaded from GenBank. Amongst the genomes included in this study, the strain LB102 of *V. campbellii* sequenced in-house is the only entry at the scaffold level wherein others are at closed/complete level. *V. cholerae* N16961 was included as an outgroup for phylogenetic comparisons. Table 1 provides a glimpse of the genomes included in this study. The genome similarity was visualized using a BLAST atlas generated with BLAST Ring Image Generator (BRIG) (Alikhan et al. 2011).

### Genome annotation

Annotation of the sequences was carried out using Prokka, a command line rapid annotation tool written using Perl (Seemann 2014). Along with the prediction of coding sequences (CDS), Prokka carries out identification of RNA sequences using RNAmmer (Lagesen et al. 2007) and clustered regularly interspaced short palindromic repeats (CRISPR) sequences using MINCED (Bland et al. 2007). Coding sequences generated from Prokka were subjected to further analysis.

### Pan-genome analysis

Coding sequences of all the genomes were concatenated and converted into a single FASTA file. Clustering of orthologous genes was done using the Usearch (v10.0.240\_win32) (Edgar 2010). The orthologous cluster thus produced is used to generate the binary presence/absence matrix. Pan-genome analysis was carried out using bacterial pan-genome analysis pipeline (BPGA) (Chaudhari et al. 2016). Shared genes were calculated following the stepwise addition of each genome. The trend thus obtained was plotted as core and pan profile curves as shown in Fig. 2. Along with pan-genome profiles, BPGA also provides information on genome-wise exclusively absent gene counts based on the orthologous clusters. This is followed by the generation of pan-core phylogeny. The pan phylogeny was generated using the pan-matrix data and the core phylogeny using the concatenated core genes, the trees were constructed using the neighbor-joining method with a default combination value of 20 iterations and *V. cholerae* as the outgroup. The functional and pathway analysis was carried out by searching through the clusters of Orthologous Groups (COGs) and Kyoto Encyclopedia of Genes and Genomes (KEGGs) databases and by assigning the COG and KEGG categories to the sequences from the best hits obtained.

**Table 1** Table showing general statistics of 35 genomes retrieved from GenBank

S. no.	Organism name	Assembly name	GC%	Genes	Proteins	Release date
1	<i>Vibrio alginolyticus</i> ATCC 17749	GCA_000354175.2	44.7	4753	4521	10-Sep-13
2	<i>Vibrio alginolyticus</i> ATCC 33787	GCA_001469735.1	44.48	5393	5094	9-Dec-15
3	<i>Vibrio alginolyticus</i> ZJ-T	GCA_001679745.1	44.67	5062	4777	5-Jul-16
4	<i>Vibrio anguillarum</i> 90-11-286	GCA_001660505.1	44.43	3991	3741	9-Jun-16
5	<i>Vibrio anguillarum</i> 775	GCA_000217675.1	44.48	3725	3426	9-Jul-11
6	<i>Vibrio anguillarum</i> M3	GCA_000462975.1	44.45	3811	3520	30-Aug-13
7	<i>Vibrio anguillarum</i> NB10	GCA_000786425.1	44.37	4051	3794	2-Jun-16
8	<i>Vibrio campbellii</i> ATCC-BB120	GCA_000017705.1	45.4	5802	5210	28-Aug-07
9	<i>Vibrio campbellii</i> LB102	GCA_002027615.1	45.5	5145	4920	15-Mar-17
10	<i>Vibrio campbellii</i> LMB29	GCA_001969325.1	45.02	5886	5574	23-Jan-17
11	<i>Vibrio furnissii</i> NCTC 11218	GCA_000184325.1	50.63	4649	4412	17-Dec-10
12	<i>Vibrio harveyi</i> ATCC 33843	GCA_000770115.2	44.96	5521	5234	31-Oct-14
13	<i>Vibrio harveyi</i> ATCC 43516	GCA_001558435.1	44.9	5723	5440	11-Feb-16
14	<i>Vibrio harveyi</i> QT520	GCA_001908435.2	45.05	5714	5415	19-Dec-16
15	<i>Vibrio parahaemolyticus</i> ATCC 17802	GCA_001558495.1	45.33	5097	4805	11-Feb-16
16	<i>Vibrio parahaemolyticus</i> BB22OP	GCA_000328405.1	45.33	4991	4831	21-Dec-12
17	<i>Vibrio parahaemolyticus</i> CDC_K4557	GCA_000430425.1	45.34	4924	4665	16-Jul-13
18	<i>Vibrio parahaemolyticus</i> CHN25	GCA_001700835.1	45.19	5341	5072	9-Aug-16
19	<i>Vibrio parahaemolyticus</i> FDA_R31	GCA_000430405.1	45.33	5026	4716	16-Jul-13
20	<i>Vibrio parahaemolyticus</i> FORC_004	GCA_001433415.1	45.9	4992	4713	5-Nov-15
21	<i>Vibrio parahaemolyticus</i> FORC_006	GCA_001304775.1	45.33	5017	4707	1-Oct-15
22	<i>Vibrio parahaemolyticus</i> FORC_008	GCA_001244315.1	45.44	4880	4624	5-Aug-15
23	<i>Vibrio parahaemolyticus</i> FORC_014	GCA_001636035.1	45.35	5136	4845	2-May-16
24	<i>Vibrio parahaemolyticus</i> FORC_018	GCA_001887055.1	45.44	4887	4621	28-Nov-16
25	<i>Vibrio parahaemolyticus</i> FORC_023	GCA_001758605.1	45.44	4804	4552	11-Jun-16
26	<i>Vibrio parahaemolyticus</i> RIMD 2210633	GCA_000196095.1	45.4	4991	4831	5-Mar-03
27	<i>Vibrio parahaemolyticus</i> UCM-V493	GCA_000568495.1	45.32	5080	4830	12-Feb-14
28	<i>Vibrio vulnificus</i> 93U204	GCA_000746665.1	46.7	4682	4427	29-Aug-14
29	<i>Vibrio vulnificus</i> ATL 6-1306	GCA_001558515.1	46.84	4759	4464	11-Feb-16
30	<i>Vibrio vulnificus</i> CMCP6	GCA_000039765.1	46.72	4826	4547	22-Dec-02
31	<i>Vibrio vulnificus</i> FORC_009	GCA_001433435.1	46.75	4760	4466	5-Nov-15
32	<i>Vibrio vulnificus</i> FORC_016	GCA_001653775.1	46.74	4779	4460	1-Jun-16
33	<i>Vibrio vulnificus</i> FORC_017	GCA_001675245.1	46.61	4915	4684	22-Jun-16
34	<i>Vibrio vulnificus</i> MO6-24-O	GCA_000186585.1	46.95	4707	4473	24-Jan-11
35	<i>Vibrio vulnificus</i> YJ016	GCA_000009745.1	46.67	4916	4662	1-Oct-03

## Genomic islands

The genomes were subjected to IslandViewer 4 (Bertelli et al. 2017) for the prediction of genomic islands. IslandViewer uses multiple algorithms, namely, IslandPath-DIMOB, SIGI-HMMer, IslandPick, and Islander, for prediction. It also identifies resistant and virulent gene regions present in the genomic island regions; these predictions are based on the information available on the databases, namely, CARD (Comprehensive Antibiotic Resistance Database) (Jia et al. 2017), VFDB

(Virulence Factor Database) (Chen et al. 2004), and PATRIC (Pathosystems Resource Integration Center) (Wattam et al. 2017).

## Virulence prediction

VirulentPred (Garg and Gupta 2008), a web-based tool built based on support vector machines (SVM) cascade algorithm, was used for prediction of virulent genes. CDS present in the genomic islands and predicted unique genes from the pan-



genome analysis were subjected to VirulentPred tool with default settings.

## Antibiotic resistance

The antibiotic-resistant genes were identified using two methods, (1) subjecting the genome sequences to ResFinder (Zankari et al. 2012), a web-based tool with the default parameters, and (2) collecting the antibiotic gene data available in the PATRIC database, in order to determine almost all the resistant genes present in the genomes.

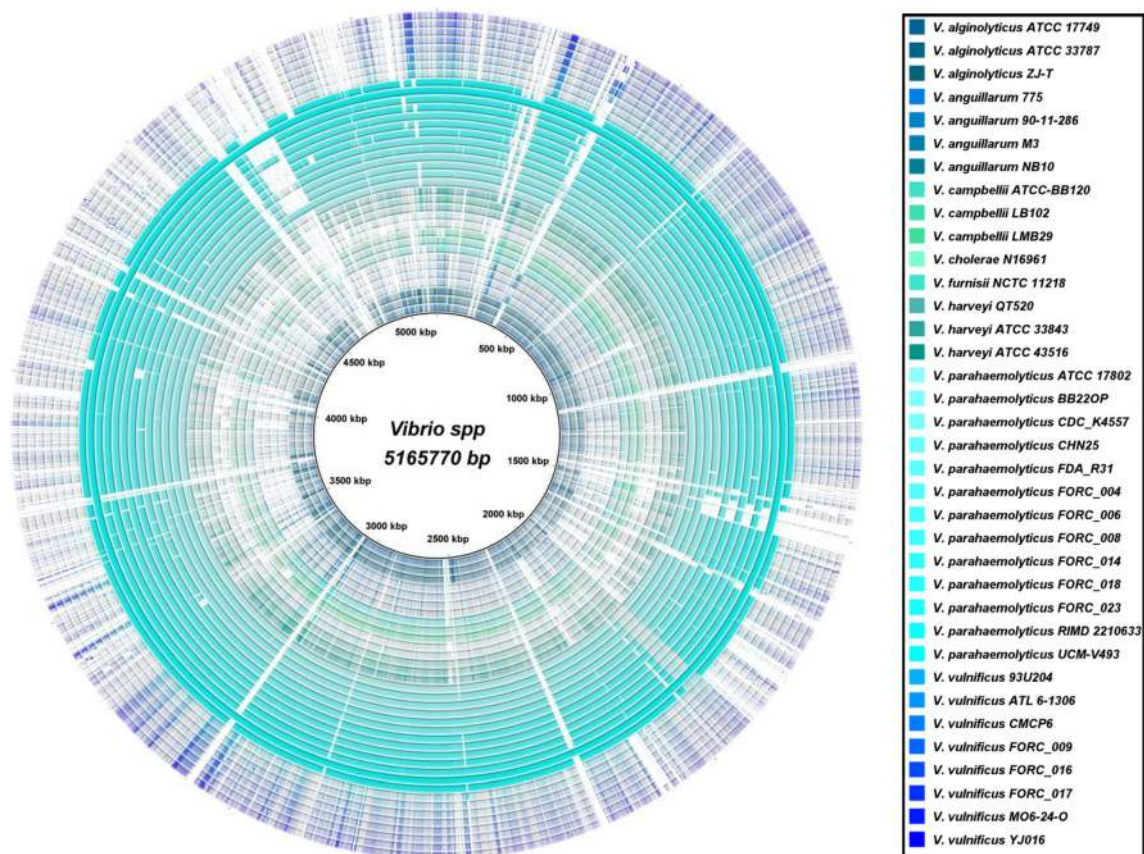
## Toxin-antitoxins prediction

To identify the TA systems present in the genomes, TAfinder (Xie et al. 2018), a web-based tool, was used to predict the type II TA systems present in the genomes. This tool is capable of identifying and predicting the type II TA loci present in the bacterial genomes. The predictions were carried out based on the homologous search module and the operon detection module with the whole genome sequences, and the GenBank files as the input.

## Results and discussion

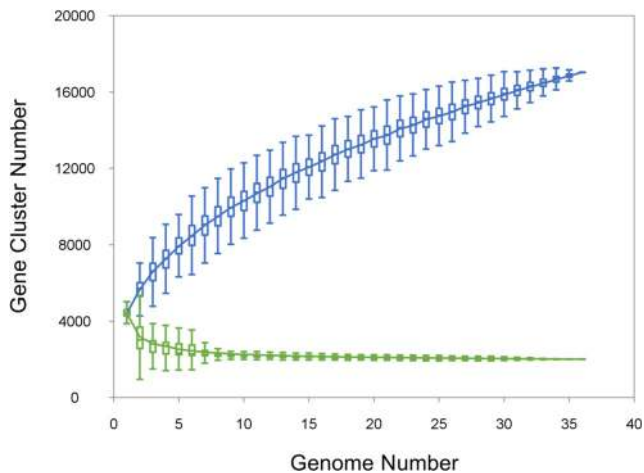
### Genome features

The genome sequences of 7 *Vibrio* species, namely, *V. alginolyticus*, *V. anguillarum*, *V. campbellii*, *V. furnissii*, *V. harveyi*, *V. parahaemolyticus*, and *V. vulnificus*, were retrieved from GenBank. These species were selected for the study due to their common occurrence in aquaculture. Their genomes ranged from 4.03–6.17 megabase pairs (Mbp) in size and the number of genes and proteins were in the ranges of 3693–5886 and 3426–5574, respectively. Amongst the 35 genomes, *V. campbellii* had the highest number of genes and proteins compared with all other species in this study. *V. harveyi* reported to be very close orthologue to *V. campbellii* stands second in terms of gene and protein count, wherein *V. anguillarum* stands last. GC content of the selected genomes ranges from 44.37 to 50.63%. The similarity between the genomes is visualized using a BLAST shown in Fig. 1, by aligning the thirty-five genomes against *V. parahaemolyticus* RIMD 2210633 along with *V. cholerae* as an outgroup. The comparison of the



**Fig. 1** Image showing the BLAST results of the 35 genome (along with *V. cholerae* as an outgroup) against the reference genome in the form of circular consecutive ring. The identity percentage is shown by the color

gradient for each ring; the regions appearing gray denote 70% similarity and the regions appearing in solid colors are similar at 90% or above



**Fig. 2** Image showing the core-pan profile of the genomes. The blue boxes represent the number of new genes for each genome and the green boxes represent the core gene count

genomes was carried out with an upper identity threshold of 90% and a lower identity threshold of 70% using BLASTn. The BLAST atlas illustrates that the genome of *V. parahaemolyticus* is more similar to *V. alginolyticus*, *V. campbellii*, and *V. harveyi* than other species of interest, and this similarity was further observed in the phylogeny analysis also.

### Genome annotation

The genomes were annotated using Prokka before subjecting to other analyses in order to standardize and enhance the

annotations. This step was carried out to normalize the anomalies in the annotations if any present, keeping in mind the fact that different annotation approaches could have been followed for each genome sequence that was submitted to GenBank. Prokka predicted the CDSs, RNAs, and CRISPRs using Prodigal, RNAmmer, and MINCED, respectively. The Prokka annotations and statistics of the genome feature revealed that an average of 4609 CDS, 31 genes coding for rRNAs and 117 genes coding for tRNAs, and 1 gene coding for tmRNAs were present in a genome; the genome-wise statistics is given in Supplementary Table 1.

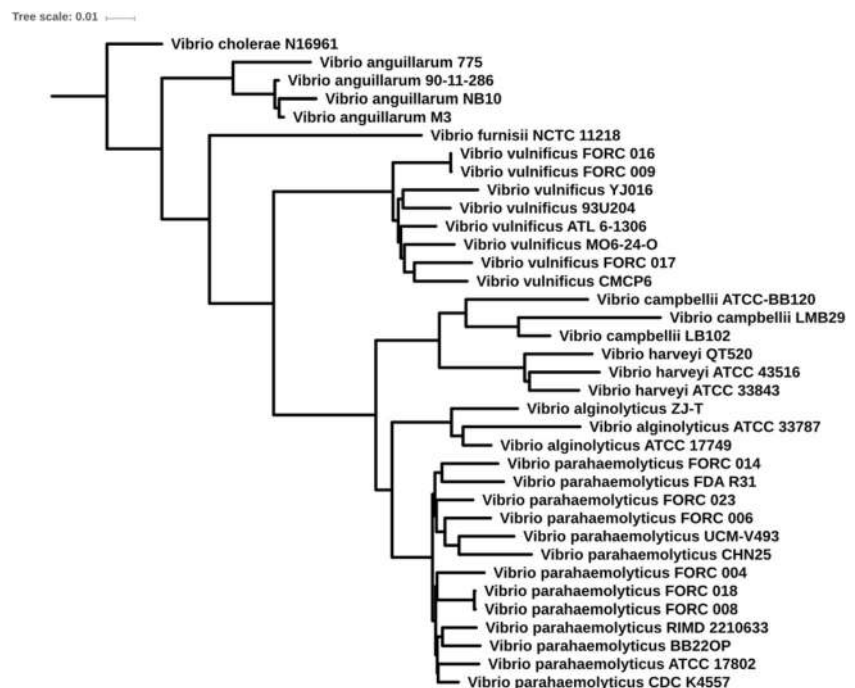
### Pan-genome analysis

The pan-genome of seven *Vibrio* species (35 strains) consisted of 17,033 genes, which included 2004 (11.76%), 8249 (48.43%), and 6780 (39.81%) core, accessory, and unique genes, respectively. The variable content of 88.24% present in the pan-genome reflects the high genetic diversity of *Vibrio* spp. Genome-wise detailed statistics on the core, accessory, unique genes along with exclusively absent genes are given in Supplementary Table 2.

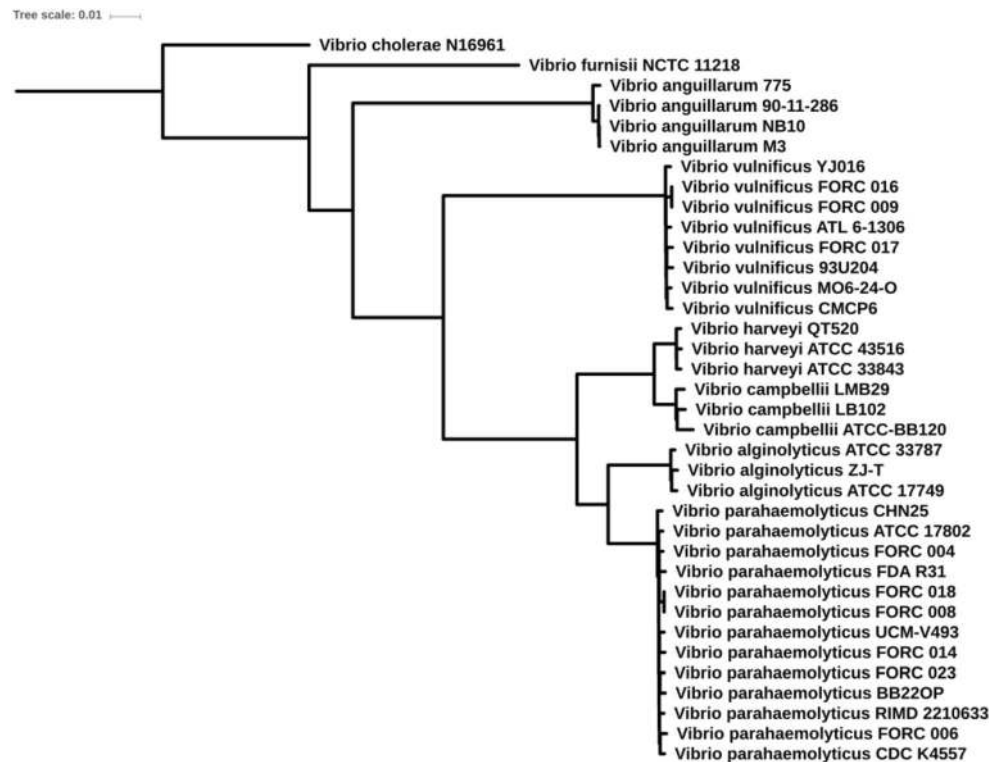
### Core-pan profile

Based on the number of core, accessory, unique, and exclusively absent genes, the core-pan profile is plotted. The core-pan profile plot exhibits the open nature of the pan-genome, i.e., with the sequential addition of genomes, the new genes count tends to increase significantly which again establishes

**Fig. 3** Image showing the phylogenetic tree of the Pan-genome



**Fig. 4** Image showing the phylogenetic tree of Core genome



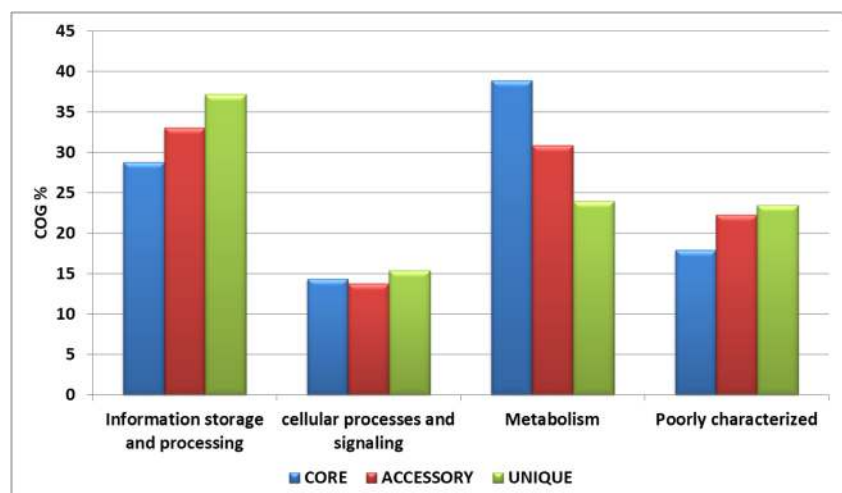
the diverseness of the *Vibrio* genomes. In general, open pan-genomes are predominantly observed in bacteria that are susceptible to horizontal gene transfer (Hurtado et al. 2018). Figure 2 represents the increase in the number of new genes in the pan-genome and the decrease in the number of core genes with the sequential addition of the genomes.

### Pan and Core phylogeny

Both the pan and core phylogenies were computed using the neighbor-joining method with a default combination of 20

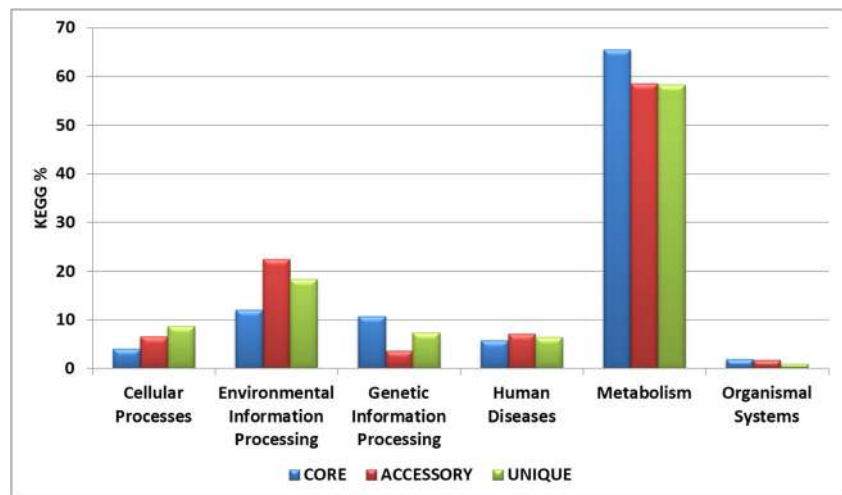
iterations. Both the pan phylogeny represented in Fig. 3 and the core phylogeny represented in Fig. 4 showed similar clustering patterns. Here in both the pan and core phylogeny, all seven species are monophyletic. The strains of *V. parahaemolyticus* and *V. alginolyticus* clusters close to each other, with a similar observation made with respect to *V. campbellii* and *V. harveyi* in both the pan and core phylogeny, whereas *V. furnissii* clusters near *V. vulnificus* in the pan phylogeny and near *V. anguillarum* in core phylogeny. By and large, clear distinction of species was observed in phylogenies made with core and pan genes.

**Fig. 5** Percentage of genes involved in information processing, cellular processes, and metabolism in *Vibrio* spp., based on COG analysis





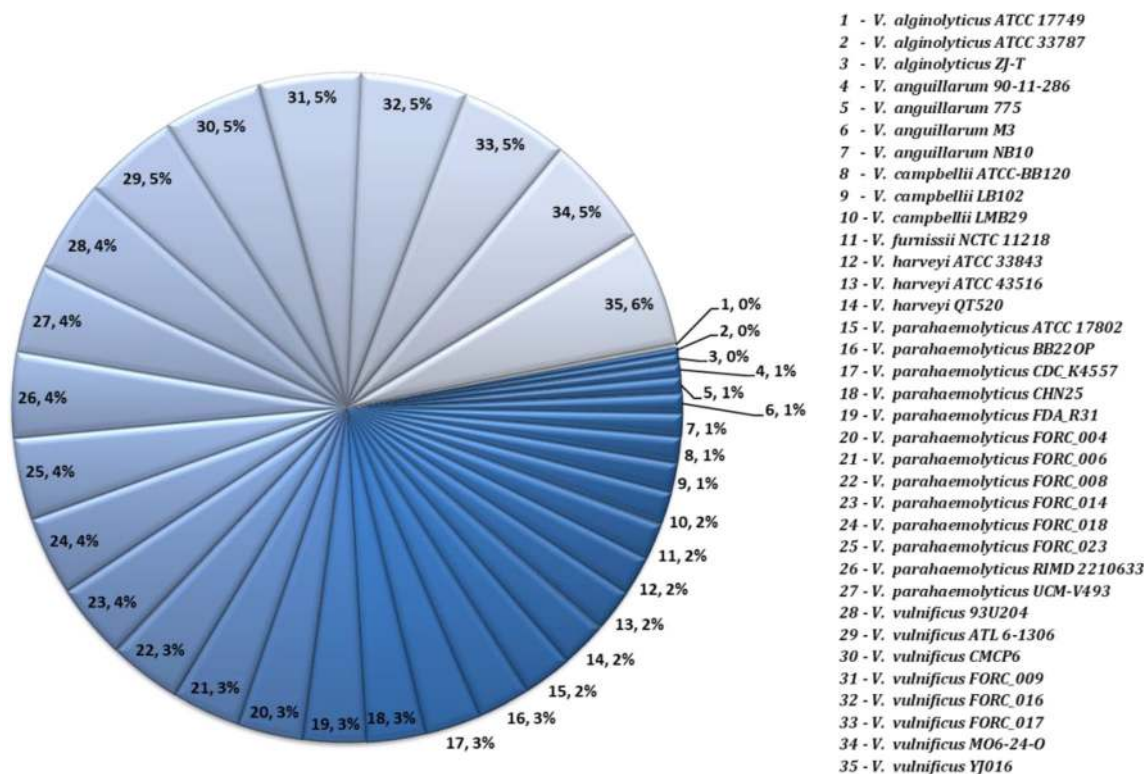
**Fig. 6** Percentage of genes involved in information processing, cellular processes, and metabolism in *Vibrio* spp., based on KEGG analysis



### Functional and pathway analysis

The functional and pathway analyses were carried out using KEGG and COG annotations. The COG annotations as represented in Fig. 5 revealed that a majority of the genes were found to be associated with metabolism, information storage, and processing. Probing further on COG annotations revealed that a large group of genes was involved in transcription followed by signal transduction. Similarly, the KEGG annotations as represented in Fig. 6 revealed that on a broader classification, genes were found to be involved in metabolism

and environmental information storage and processing. On further evaluation, it was known that the genes were found to be involved in carbohydrate, amino acid, and lipid metabolism; signal transduction; replication and repair; translation; cell motility; and in energy metabolism. The core genes are mostly involved in metabolism and the accessory and unique genes are majorly involved in information storage and processing with respect to environment and other factors. From these results, we can infer that the genome is designed in such a way that the genes in some cases belonging to the core gene set and in some cases belonging to the variable gene set are



**Fig. 7** Percentage of genes encoding the toxin-antitoxin families in 35 *Vibrio* genomes

largely involved in the process and mechanism of survival during variable conditions and thereby contributing to the virulence and pathogenicity of the organism.

### Toxin-antitoxin systems

The toxin-antitoxin prediction using the TAFinder revealed the presence of 394 type II toxins and antitoxins on total in the 35 organisms. The list of the TAs present in each genome is given in Supplementary Table 3. Amongst the 35 organisms, *V. harveyi* QT520 contained the highest number of the toxin-antitoxin systems present in them, followed by *V. vulnificus* and *V. campbellii* strains, respectively (Fig. 7). The most commonly observed toxin/antitoxin (T/A) pairs amongst the 35 genomes are relE/relB, relE/PHD, relE/RHH, GNAT/RHH, and HipA/Xre. The TA system is important as it reveals not only pathogenicity of bacteria but also its role in the defense mechanism. Every organism has its own toxin and antitoxin system, the toxin part of the system aids in cell death during unfavorable conditions and the antitoxin part of the system also helps in unfavorable conditions but in this case, it helps the organism to overcome the obstacles caused by various external and internal factors. All the 35 genomes were identified with at least one toxin-antitoxin pair.

### Antibiotic resistance

Antibiotic resistance in bacteria is becoming one of the major threats to human health. With the emergence and progression of diseases on one side and the bacteria becoming resistant to one or more drugs on the other side, it is important to study the bacteria's competence and the mechanism of action it adopts to safeguard itself from one or more drugs. Hence, it is very important to delineate these antibiotic-resistant genes.

The antibiotic resistance prediction revealed the presence of 312 resistant genes in total (Supplementary Table 4). Twenty-one of the 36 genomes were identified with antibiotic resistance and 6 of them had at least one gene coding for multidrug resistance. The predicted antibiotic resistance genes were primarily involved in functions such as target site protection, antibiotic efflux, antibiotic inactivation, and decrease in antibiotic permeability. These genes were found to be coding for resistance against some of the highly important antibiotics such as tetracycline, beta-lactam, chloramphenicol, fluoroquinolone, fosfomycin, and sulfonamides like trimethoprim (Guglielmini and Van Melderden 2011). Six strains, viz., *V. alginolyticus* ATCC 17749, *V. anguillarum*, *V. furnissii*, *V. parahaemolyticus* RIMD 221063, and *V. vulnificus* CMCP6 and YJ016, were identified with the presence of multidrug resistant genes. The multidrug resistant genes present in these 6 strains were of two types, i.e., (i) multidrug resistance transporter and (ii) multidrug efflux pump.

### Virulence factors

One of the important components of a genome is the genes coding for virulence factors. Hence, the prediction of potentially virulent genes is necessary. Genomic islands are regions of the genome that had been horizontally transferred. The genomic islands and the unique genes were subjected to virulence prediction using VirulentPred. As a result, 4802 out of 7098 genes of the genomic islands and 4825 out of 6780 unique genes were predicted to be potentially virulent (Supplementary Table 5). The analysis of the predicted virulent genes revealed that amongst the unique virulent genes, 1885 had proper annotations, whereas 2756 were hypothetical, 64 were unknown, and 120 were annotated as uncharacterized genes. Similarly, amongst the genomic islands, 1773 had proper annotations and 3029 were annotated as hypothetical proteins.

*V. campbellii* strain LB 102's unique genes were studied in detail due to its predominant occurrence in the shrimp hatcheries present in the coastal regions in India. The unique genes from *V. campbellii* LB 102 were extracted and checked for virulence; a total of 131 unique genes were subjected to virulence prediction. The analysis revealed the presence of 19 virulence genes out of which, 13 had known functions and the remaining 6 were categorized as hypothetical proteins. On delving deep into the function of these 13 genes might help us understand the host-pathogen interaction with strain specificity.

### Conclusion

The pan-genome analysis helped us in figuring out the functions of the core genes and the dispensable (accessory and unique) genes. The pan-genome, genomic islands, and the phylogenetic analysis were useful for understanding the genetic diversity, the evolutionary pattern, and relationships. In addition, prediction and identification of genes coding for toxin-antitoxin systems, virulence factors, and antibiotic resistance implicate to the pathogenicity, survival, and the defense mechanisms of *Vibrio* spp. The unique genes could serve as targets for developing diagnostics and help in designing other methods to scale down the epidemic.

**Funding information** The Indian Council of Agricultural Research (ICAR), New Delhi, provided financial support under the "Network Project on Agricultural Bioinformatics and Computational Biology."

### Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Research involving human participants and/or animals** N/A



**Informed consent** N/A

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Alikhan N-F, Petty NK, Ben Zakour NL, Beatson SA (2011) BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12:402. <https://doi.org/10.1186/1471-2164-12-402>
- Baker-Austin C, Oliver JD, Alam M, et al (2018) *Vibrio* spp. infections. *Nat Rev Dis Prim* 4. <https://doi.org/10.1038/s41572-018-0005-8>
- Bertelli C, Laird MR, Williams KP et al (2017) IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res* 45:W30–W35. <https://doi.org/10.1093/nar/gkx343>
- Bland C, Ramsey TL, Sabree F et al (2007) CRISPR Recognition Tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* 8:209. <https://doi.org/10.1186/1471-2105-8-209>
- Chaudhari NM, Gupta VK, Dutta C (2016) BPGA-an ultra-fast pan-genome analysis pipeline. *Sci Rep* 6:1–10. <https://doi.org/10.1038/srep24373>
- Chen L, Yang J, Yu J et al (2004) VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res* 33:D325–D328. <https://doi.org/10.1093/nar/gki008>
- Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>
- Fao (2018) World fisheries and aquaculture
- Garg A, Gupta D (2008) VirulentPred: a SVM based prediction method for virulent proteins in bacterial pathogens. *BMC Bioinformatics* 9: 1–12. <https://doi.org/10.1186/1471-2105-9-62>
- Guérout A-M, Iqbal N, Mine N et al (2013) Characterization of the phd-doc and ccd toxin-antitoxin cassettes from *Vibrio* superintegrons. *J Bacteriol* 195:2270–2283. <https://doi.org/10.1128/JB.01389-12>
- Guglielmini J, Van Melder L (2011) Bacterial toxin-antitoxin systems. *Mob Genet Elements* 1:283–306. <https://doi.org/10.4161/mge.18477>
- Hurtado R, Carhuarica D, Soares S et al (2018) Pan-genomic approach shows insight of genetic divergence and pathogenic-adaptation of *Pasteurella multocida*. *Gene* 670:193–206. <https://doi.org/10.1016/j.gene.2018.05.084>
- Ina-Salwany MY, Al-saari N, Mohamad A et al (2019) Vibriosis in fish: a review on disease development and prevention. *J Aquat Anim Health* 31:3–22. <https://doi.org/10.1002/aah.10045>
- Jia B, Raphenya AR, Alcock B et al (2017) CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res* 45:D566–D573. <https://doi.org/10.1093/nar/gkw1004>
- Kim JY, Lee J-L (2017) Correlation of total bacterial and *Vibrio* spp. populations between fish and water in the aquaculture system. *Front Mar Sci* 4:147. <https://doi.org/10.3389/fmars.2017.00147>
- Lagesen K, Hallin P, Rødland EA et al (2007) RNAMmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 35: 3100–3108. <https://doi.org/10.1093/nar/gkm160>
- Le Roux F, Wegner KM, Baker-Austin C et al (2015) The emergence of *Vibrio* pathogens in Europe: ecology, evolution, and pathogenesis (Paris, 11–12th March 2015). *Front Microbiol* 6:830. <https://doi.org/10.3389/fmicb.2015.00830>
- Lloyd NA, Nazaret S, Barkay T (2019) Genome-facilitated discovery of RND efflux pump-mediated resistance to cephalosporins in *Vibrio* spp. isolated from the mummichog fish gut. *J Glob Antimicrob Resist*. <https://doi.org/10.1016/J.JGAR.2019.05.006>
- Mohamad N, Amal MNA, Saad MZ et al (2019) Virulence-associated genes and antibiotic resistance patterns of *Vibrio* spp. isolated from cultured marine fishes in Malaysia. *BMC Vet Res* 15:176. <https://doi.org/10.1186/s12917-019-1907-8>
- Mok JS, Ryu A, Kwon JY et al (2019) 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* 106:106697. <https://doi.org/10.1016/J.FOODCONT.2019.06.023>
- Seemann T (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>
- Silvester R, Alexander D, Hatha M, Ammanamveetil A (2015) Prevalence, antibiotic resistance, virulence and plasmid profiles of *Vibrio parahaemolyticus* from a tropical estuary and adjoining traditional prawn farm along the southwest coast of India. <https://doi.org/10.1007/s13213-015-1053-x>
- Soto-Rodriguez SA, Lozano-Olvera R, Palacios-Gonzalez DA et al (2019) Characterization and growth conditions of *Vibrio parahaemolyticus* strains with different virulence degrees that cause acute hepatopancreatic necrosis disease in *Litopenaeus vannamei*. *J World Aquac Soc*. <https://doi.org/10.1111/jwas.12617>
- Szekeres S, Dauti M, Wilde C et al (2007) Chromosomal toxin-antitoxin loci can diminish large-scale genome reductions in the absence of selection. *Mol Microbiol* 63:1588–1605. <https://doi.org/10.1111/j.1365-2958.2007.05613.x>
- Wattam AR, Davis JJ, Assaf R et al (2017) Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center. *Nucleic Acids Res* 45:D535–D542. <https://doi.org/10.1093/nar/gkw1017>
- Xie Y, Wei Y, Shen Y et al (2018) TADB 2.0: an updated database of bacterial type II toxin-antitoxin loci. *Nucleic Acids Res* 46:D749–D753. <https://doi.org/10.1093/nar/gkx1033>
- Zankari E, Hasman H, Cosentino S et al (2012) Identification of acquired antimicrobial resistance genes. *J Antimicrob Chemother* 67:2640–2644. <https://doi.org/10.1093/jac/dks261>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.