Genomic mining of *Vibrio parahaemolyticus* highlights prevalence of antimicrobial resistance genes and new genetic markers associated with AHPND and *tdh+/trh+* genotypes

Marieke Vandeputte[1,2*], Sieglinde Coppens[3], Peter Bossier[2†], Nick Vereecke[3†] and Daisy Vanrompay[1†]

Abstract

Background Acute Hepatopancreatic Necrosis Disease (AHPND) causes significant mortality in shrimp aquaculture. The infection is primarily instigated by *Vibrio parahaemolyticus* (*Vp*) strains carrying a plasmid encoding the binary toxin PirAB. Yet, comprehension of supplementary virulence factors associated with this relatively recent disease remains limited. Furthermore, the same holds for gastroenteritis in humans caused by other *Vp* genotypes. Additionally, given the prevalent use of antibiotics to combat bacterial infections, it becomes imperative to illuminate the presence of antimicrobial resistance genes within these bacteria.

Results A subsampled number of 1,036 *Vp* genomes was screened for the presence of antimicrobial resistance genes, revealing an average prevalence of 5 ± 2 (SD) genes. Additional phenotypic antimicrobial susceptibility testing of three *Vp* strains (M0904, TW01, and PV1) sequenced in this study demonstrated resistance to ampicillin by all tested strains. Additionally, *Vp* M0904 showed multidrug resistance (against ampicillin, tetracycline, and trimethoprim-sulfamethoxazole). With a focus on AHPND, a screening of all *Vibrio* spp. for the presence of *pirA* and/or *pirB* indicates an estimated prevalence of 0.6%, including four *V. campbellii*, four *V. owensii*, and a *Vibrio* sp. next to *Vp*. Their *pirAB*-encoding plasmids exhibited a highly conserved backbone, with variations primarily in the region of the Tn3 family transposase. Furthermore, an assessment of the subsampled *Vp* genomes for the presence of known virulence factors showed a correlation between the presence of the Type 3 Secretion System 2 and *tdh*, while the presence of the Type 6 Secretion System 1 was clade dependent. Furthermore, a genome-wide association study (GWAS) unveiled (new) genes associated with *pirA*, *pirB*, *tdh*, and *trh* genotypes. Notable associations with the *pirAB* genotype included outer membrane proteins, immunoglobulin-like domain containing proteins, and toxin-antitoxin systems. For the *tdh+/trh+* genotypes (containing *tdh*, *trh*, or both genes), associations were found with T3SS2 genes, urease-related genes and nickel-transport system genes, and genes involved in a 'minimal' type I-F CRISPR mechanism.

---

†Peter Bossier, Nick Vereecke and Daisy Vanrompay share senior authorship.

*Correspondence:
Marieke Vandeputte
Marieke.Vandeputte@UGent.be
Full list of author information is available at the end of the article

Vandeputte *et al. BMC Genomics*   (2024) 25:178

Page 2 of 14

Conclusions This study highlights the prevalence of antimicrobial resistance and virulence genes in *Vp*, identifying novel genetic markers associated with AHPND and *tdh + /trh +* genotypes. These findings contribute valuable insights into the genomic basis of these genotypes, with implications for shrimp aquaculture and food safety.

Keywords *Vibrio parahaemolyticus*, Acute Hepatopancreatic Necrosis Disease, Whole genome sequencing, Antimicrobial resistance, Genome-wide association, Aquaculture, Shrimp, PirAB toxins

Background

Acute Hepatopancreatic Necrosis Disease (AHPND), previously referred to as Early Mortality Syndrome (EMS), is an emerging bacterial disease in shrimp. It has inflicted substantial economic losses on the shrimp industry, particularly in Asian countries like Thailand, Vietnam, and Malaysia, as well as in South America and the United States [1–4]. AHPND is caused by specific strains of the Gram-negative bacterium *Vibrio parahaemolyticus* carrying a conjugative plasmid (pVA1) of approximately 69 kbp in size, housing the *pirAB^{vp}* genes [5]. The binary pore-forming toxin PirAB is homologous to toxins secreted by *Photorhabdus* spp. (Pir) and is responsible for the characteristic lesions observed in the shrimp's hepatopancreas [4, 5]. More recently, other *Vibrio* (*V.*) species have been reported to carry plasmids homologous to pVA1, that carry the *pirAB* genes, such as *V. harveyi* [6], *V. campbellii* [7–9], *V. owensii* [10] and *V. punensis* [11]. The plasmid-borne nature of these toxins facilitates their mobility and spread to different *Vibrio* species, which makes it more difficult to control the disease [12, 13]. Increased horizontal gene transfer of this mobile genetic element (MGE) is facilitated by the number of genes associated with conjugative transfer and plasmid mobilization, rendering the plasmid self-transmissible [1, 5]. This principle has been demonstrated in vitro by the successful horizontal transfer of a pVA1-type plasmid from *V. parahaemolyticus* to a non-pathogenic *V. campbellii* [14]. Furthermore, the plasmid encompasses a post-segregational killing system, ensuring its consistent inheritance [1, 5, 12, 15]. The plasmid contains a set of transfer-related genes, including the Tn903 (3.5 kbp) transposon where the *pirAB^{vp}* genes are located, and four transposases (ORF15, 48, 55, and 68) able to trigger the transfer of the DNA transposon [12, 16]. This has resulted in plasmid variants such as deletions of the *pirA* gene, total loss of *pirA,* and partial loss of *pirB* coding sequences [15, 17, 18].

Recent research has shown a geographical distinction between plasmids originating from Asian strains and Latin American strains. Specifically, Asian strains lack the Tn3 transposon that is present in all Latin American strains [19, 20]. However, a broader phylogenomic analysis by Yang et al. *(*2019) has suggested that *V. parahaemolyticus* is divided into four diverse populations, VppUS1, VppUS2, VppX, and VppAsia [21]. The VppUS1 and VppUS2 populations are largely restricted to the US and Northern Europe, while VppX and VppAsia are found worldwide, with VppAsia making up the great majority of seawater isolates around Asia [21].

Next to the aforementioned plasmid-encoded toxins, *V. parahaemolyticus* can contain chromosome-encoded virulence factors, such as hemolysins (thermostable direct hemolysin (TDH) and TDH-related hemolysin (TRH)), two Type III secretion systems (T3SS1 and T3SS2) and two T6SSs (T6SS1 and T6SS2). Genotypes carrying *tdh* and/or *trh* genes are highly pathogenic in humans, causing gastroenteritis after consumption of contaminated seafood. Throughout this manuscript, these genotypes are referred to as '*tdh + /trh +* genotypes'. Additionally, T3SS2 was shown to be present in these genotypes. TDH, TRH, and T3SS2 are exclusively found in these human clinical isolates responsible for human acute gastroenteritis and are not related to pathogenesis in shrimp [22, 23]. In contrast, T3SS1 and T6SS2 are highly conserved and are suggested to play a role in virulence in both human infections and AHPND infections in shrimp. However, their working mechanisms are not fully understood yet [24, 25].

Over the last few years, aquaculture has expanded significantly, resulting in an increase in disease prevalence. This, in turn, has led to increased antibiotic usage among farms worldwide [26]. However, the improper, empirical, and extensive application of antibiotics has resulted in the emergence of antimicrobial resistance in a substantial amount of *V. parahaemolyticus* strains in the environment [27–29]. Every year, there is a reported increase in the multidrug resistance of pathogenic *V. parahaemolyticus* to clinically important antimicrobial drugs. This not only reduces the effectiveness of antibiotics used in aquaculture practices but also poses a threat to human health due to the transmission of *V. parahaemolyticus* strains carrying (mobilizable) antimicrobial resistance genes (ARGs) to humans upon consumption [27, 30, 31].

In the scope of this study, three distinct AHPND-associated *V. parahaemolyticus* strains were sequenced, originating from three different geographical sources: Mexico, Thailand, and China. The identification of antimicrobial resistance genes (ARGs) was linked to phenotypic antimicrobial susceptibility testing against a panel of 13 different antibiotics. In addition, using publicly

Vandeputte *et al. BMC Genomics*   (2024) 25:178

Page 3 of 14

available whole genome sequencing (WGS) data, extensive genomic mining, phylogenetics, and genome-wide association studies (GWAS) were performed to elucidate and characterize putative genetic mediators in pathogenic *V. parahaemolyticus* strains with a focus on AHPND and $tdh+/trh+$ genotypes.

Results

Long-read whole genome sequencing statistics and quality check

As summarized in Additional file 1: Table S1, our newly sequenced isolates *Vp* M0904 (CP133891-CP133899), *Vp* PV1 (JAVKPG000000000), and *Vp* TW01 (CP133900-CP133905) showed 100% genome completeness, an average GC content of 45.3%, genome sizes of 5.6 Mbp ($\pm$ 0.2 Mbp), longest contigs of 3.5 Mbp ($\pm$ 0.04 Mbp), and an average of 5,718 ($\pm$ 169) predicted genes. These values are all in line with the RIMD 2210633 *V. parahaemolyticus* (NC_004603.1) NCBI reference strain.

Phenotypic antibiotic susceptibility testing and genomic ARG screening

The results from the antibiotic susceptibility testing can be found in Table 1. All three strains were resistant to ampicillin, with *Vp* M0904 showing additional resistance to piperacillin, tetracycline, and trimethoprim-sulfamethoxazole. Furthermore, both *Vp* TW01 and *Vp* PV1 were intermediately resistant to piperacillin. Interestingly, these phenotypes were supported by chromosome-encoded ARGs. While strains M0904 and PV1 encoded a $bla_{CARB-23}$ beta-lactamase, strain *Vp* TW01 had a $bla_{CARB-21}$ gene, explaining the observed phenotypes against

ampicillin and piperacillin. Comparable, the *Vp* M0904 strain was the only strain carrying a *tet(B)* and *dfrA6* gene, supporting the resistance phenotypes against tetracycline and trimethoprim-sulfamethoxazole, respectively. Also, a *QnrVC5* gene was found in its genome, though no phenotypic resistance against any of the tested fluoroquinolones was observed. As shown in Additional file 2 and Additional file 1: Table S4, the *CRP*, *msbA*, *tet(35)*, and *ugd* genes were found in all three genomes.

Whenever expanding our ARG screening to our sub-sampled dataset, we observed an average presence of 5 $\pm$ 2 (SD) ARGs per *V. parahaemolyticus* genome (Additional file 2 and Additional file 1: Table S4). This also confirmed the overall high prevalence of the *CRP* (99.7%), *msbA* (99.0%), *tet(35)* (99.3%), and *ugd* (84.4%) genes within the *V. parahaemolyticus* species. Notably, these genes showed relatively low nucleotide identity scores to the references as compared to other ARGs; 80.0% $\pm$ 0.1%, 66.5% $\pm$ 0.2%, 82.6% $\pm$ 0.2%, and 69.7% $\pm$ 0.9% nucleotide identity, for *CRP*, *msbA*, *tet(35)*, and *ugd*, respectively. A substantial amount of *V. parahaemolyticus* strains harbored either the $bla_{CARB-23}$ (44.3%) or $bla_{CARB-21}$ (55.1%) gene, suggesting a high resistance prevalence against β-lactams such as ampicillin and piperacillin. Other β-lactamases that were identified included the $bla_{CARB-22}$ (0.2%), $bla_{CTX-M-14}$ (3.1%), $bla_{CTX-M-15}$ (0.1%), bla_{NDM-1} (0.2%), bla_{TEM-1} (0.1%), $bla_{TEM-116}$ (0.1%), bla_{VEB-1} (2.2%). Aversely, resistance towards tetracycline and trimethoprim-sulfamethoxazole is suggested to be rather low as the *tet(B)* and *dfrA6* genes were only found in 3.1% and 4.1% of subsampled strains. Additionally, other tetracycline (*tet(59)* (1.9%),

Table 1 Results of antibiotic susceptibility tests of the *V. parahaemolyticus* (*Vp*) strains. Strains were resistant (R) or susceptible (S) to the antibiotics or showed an intermediate (I) profile. Identified ARGs, supporting the observed resistance phenotypes are included. A visual overview of all identified ARGS is given in Additional file 2

Antibiotic class	Antibiotic	*Vp* M0904	*Vp* TW01	*Vp* PV1
B-lactam				
- Penams	ampicillin (10 μg)	R $bla_{CARB-23}$	R $bla_{CARB-21}$	R $bla_{CARB-23}$
	amoxicillin-clavulanate (20/10 μg)	S	S	S
	ampicillin-sulbactam (10/10 μg)	S	S	S
	piperacillin (100 μg)	R $bla_{CARB-23}$	I $bla_{CARB-21}$	I $bla_{CARB-23}$
- Cephalosporins	cefotaxime (30 μg)	S	S	S
	ceftazidime (30 μg)	S	S	S
Aminoglycosides	amikacin (30 μg)	S	S	S
	gentamicin (10 μg)	S	S	S
Tetracyclines	tetracycline (30 μg)	R *tet(B)*	S	S
Fluoroquinolones	ciprofloxacin (5 μg)	S	S	S
	levofloxacin (5 μg)	S	S	S
	ofloxacin (5 μg)	S	S	S
Folate pathway inhibitors	trimethoprim-sulfamethoxazole (5 μg)	R *dfrA6*	S	S

Vandeputte *et al. BMC Genomics*   (2024) 25:178

Page 4 of 14

tet(A) (0.6%), *tet(C)* (0.1%), *tetI* (0.2%), and *tetM* (0.8%)) and trimethoprim-sulfamethoxazole-associated ARGs (*dfrA1* (0.1%), *dfrA16* (0.1%), *dfrA23* (0.1%), *dffrA27* (0.2%), *dfrA32* (0.1%), *sul1* (1.1%), and *sul2* (3.8%)) were identified. Of note, while a variety of other ARGs were identified at low prevalence (Additional file 1: Table S4), ARGs associated with aminoglycoside resistance ranged from 0.1% up to 3.1% prevalence, including the AAC(3)-IV, AAC(6')-Iia, AAC(6')-Ib9, ANT(2")-Ia, APH(3")-Ib, APH(3')-Ia, APH(6)-Id. These genes seemed to be limited to strains showing a substantially higher (≥ 10 per strain) number of encoded ARGs. Furthermore, as shown in Additional file 2, this data is suggestive of the acquisition of a putative multidrug resistance MGE (*e.g.,* plasmid or integron) that defines a subclade within the VppAsia clade.

Prevalence of AHPND-positive *Vibrio* species and associated plasmids

With a focus on AHPND, all available *Vibrio* spp. genomes from NCBI (24,758 accessed on 05/24/2023) were screened for the presence of *pirA* and *pirB* genes. A total of 142 (0.6%) of all available genome assemblies were *pirA* and/or *pirB* positive, including 133 *V. parahaemolyticus*, four *V. campbellii*, four *V. owensii*, and one *Vibrio* sp. A phylogenetic inference of these plasmids, encoding AHPND-associated genes, showed a highly conserved plasmid backbone (Fig. 1A and Additional file 1: Table S6). The region encoding for a Tn3 family transposase was the most variable one, showing its presence in 30 (21.1%) of the *pirAB*-associated plasmids. While ten (out of a total of 13 VppX strains; 76.9%) of these belonged to the VppX cluster (orange), five (out of a total of 108 VppAsia strains; 4.6%) belonged to the VppAsia cluster (green). For the remaining 15 (out of 21; 71.4%) no data on strain location was available. The plasmids originating from other *Vibrio* species clustered along the Maximum-Likelihood (ML) tree and did not show diverging plasmid compositions, except for strain 15112C from *V. campbellii*, which only had a few matching plasmid genes. Our two Asian strains (TW01 and PV1) and the M0904 Mexican strain showed high sequence concordance and plasmid build-up as compared to the used reference (pVPR14; CP028145) (Fig. 1B). As highlighted in Fig. 1C, our M0904 strain showed the presence of

a Tn3 family transposon comparable to the pVPR14 (CP028145) reference plasmid. This transposon was missing in both Asian strains as complying with the pVA1 (KP324996) reference plasmid. While the conjugative protein operons, mobilization proteins, and the type II toxin-antitoxin system were highly conserved between all plasmids, the PirA/B region showed some heterogeneity (Fig. 1C). An extra IS5 family ISVa2 transposon was found in the pVPR14 (CP028145) plasmid immediately upstream of the *pirA* coding sequence, which was absent in our M0904 strain. This PirA/B region was conserved in our TW01 and PV1 strain, whereas the pVA1 (KP324996) plasmid showed an inverted PirA/B region between the two IS5 family ISVa2 transposons.

Prevalence and genome-wide associations of common virulence-associated genes in *V. parahaemolyticus*

From a total of 301 putative virulence factors, 224 were shown to be present in 90% of our isolates (Additional file 1: Table S5). Hence, these were excluded from further analyses. The remaining 77 genes were visualized in Fig. 2. A relatively low prevalence of *pirA* (4.8%) and *pirB* (4.3%) was observed, with 10% of the *pirA* + genotypes showing the presence of *pirA* only (*i.e.,* complete *pirB* deletion or partial below Abricate cut-offs; –minid 80 –mincov 80). While our three newly sequenced strains were all *pirA*/*pirB* positive, they lacked the chromosome-encoded *tdh* and *trh* genes. The overall prevalence of the TDH hemolysin and TRH hemolysin were estimated at 18.9% and 19.3%, respectively. The presence of the TDH hemolysin showed an apparent correlation with the presence of genes linked to the Type III Secretion System on its second chromosome (T3SS2 on NC_004605 as reference). These genes were found to be present in the *V. parahaemolyticus* population ranging between 13.7% and 15.7%. Of note, for some T3SS2 genes, other *V. parahaemolyticus* strains showed the presence of T3SS2 genes at a lower nucleotide identity, raising its prevalence to 34.1%-36.8% (orange-yellow in Fig. 2). Interestingly, when looking at the Type VI Secretion System on the first chromosome (T6SS1 on NC_004603 as reference), a clear clade-dependent trend can be observed. This results in a clear split within the VppAsia cluster (green), which is also true for the VppX (beige) and VppUS1 (light blue) clusters. The VppUS2 (dark blue) strains do not seem to

(See figure on next page.)
Fig. 1 Maximum-Likelihood (ML) tree of pirAB-associated *V. parahaemolyticus* plasmid sequences and identified plasmid genes (pgenes). **A** All identified genes are classified per region in the plasmid based on the pVPR14 (CP028145). Our three new strains and the reference strains were indicated with blue and red circles, respectively and clades were colored using the color code presented by Yang et al., 2019 [21]. Plasmids originating from *V. campbellii* and *V. owensii* isolates were indicated with a red and blue box, respectively; **B** The genetic landscape of our three new isolates in relation to some reference strains; **C** Zoom-in on the versatile pirAB region within the plasmid highlighting
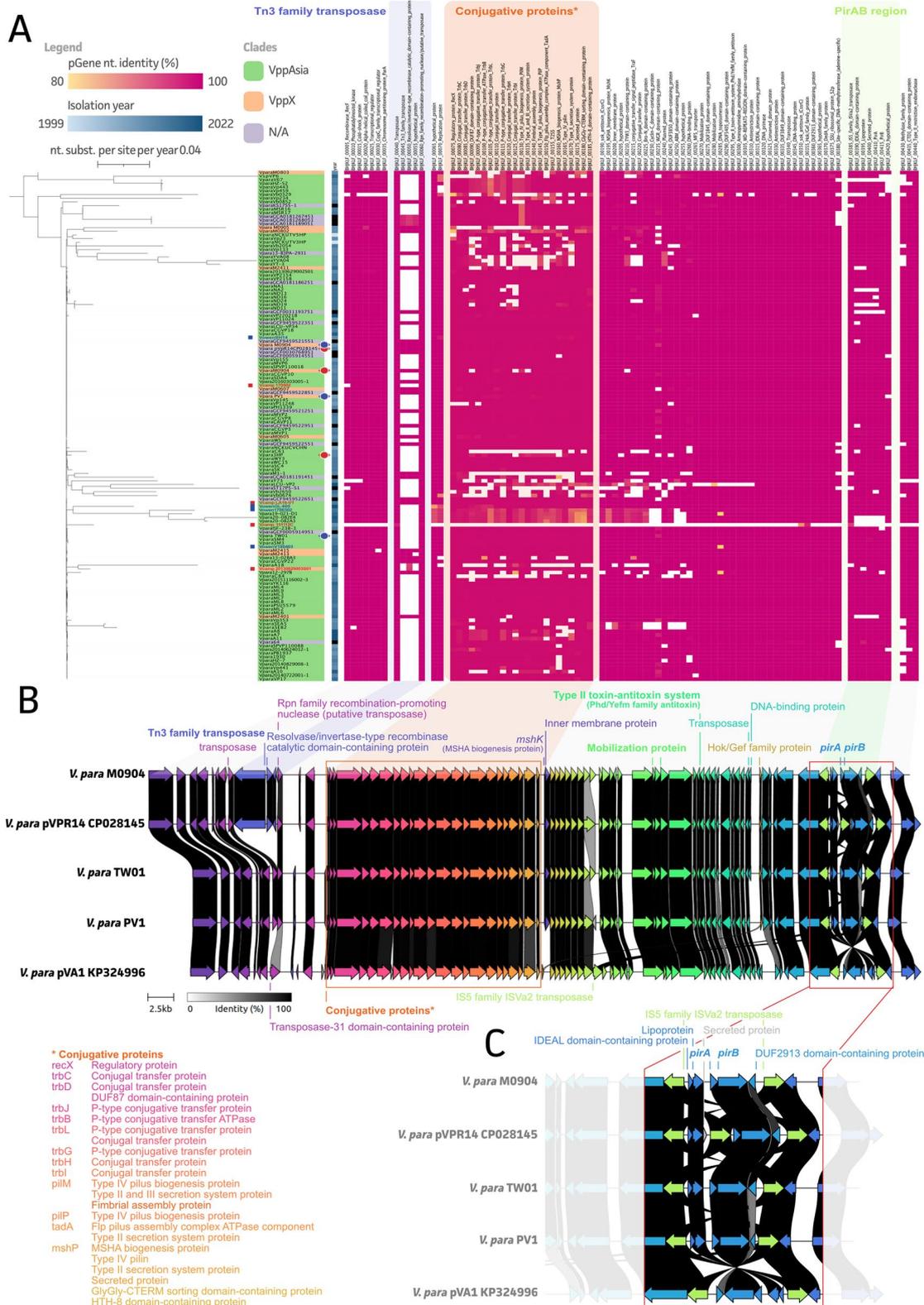
Vandeputte *et al. BMC Genomics*   (2024) 25:178

Page 5 of 14

Fig. 1 (See legend on previous page.)

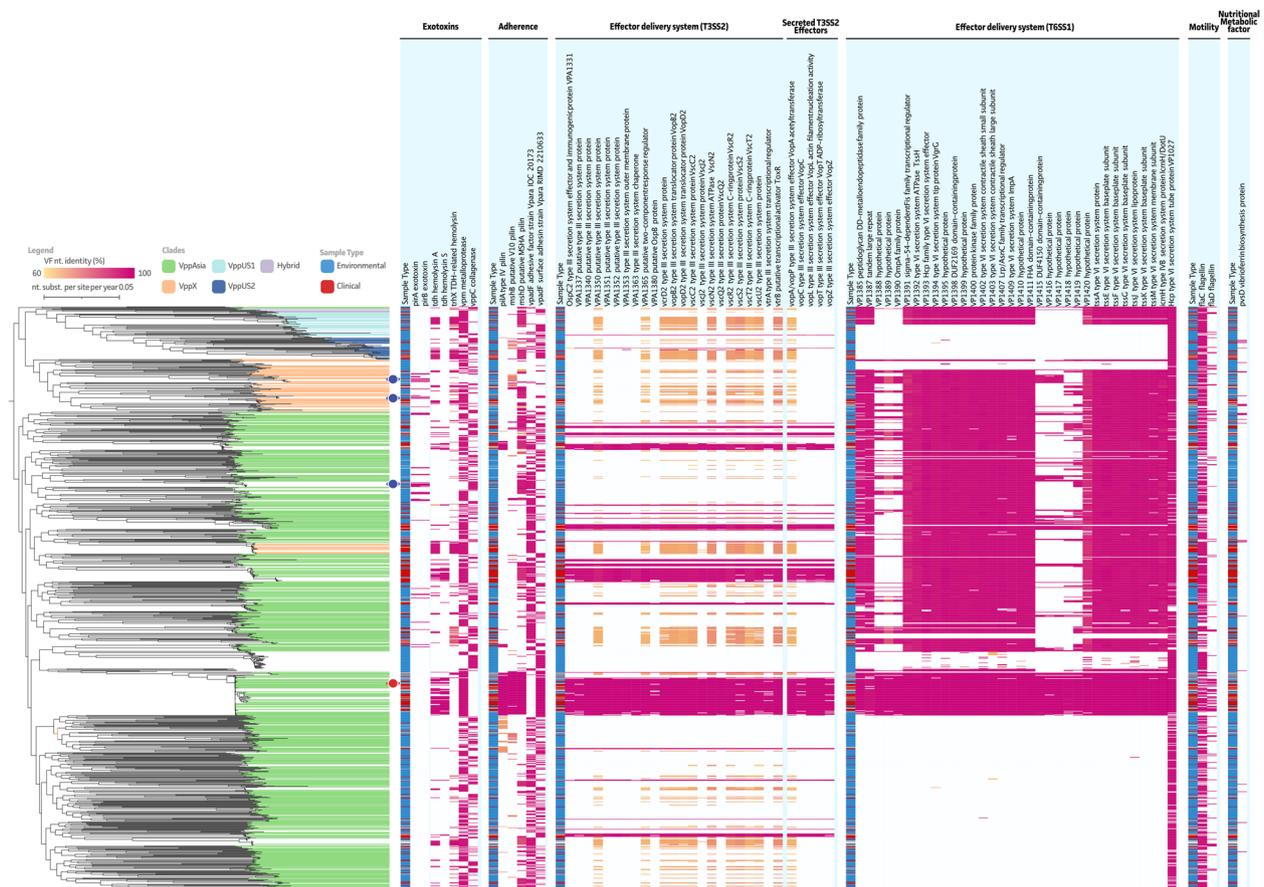Vandeputte *et al. BMC Genomics*   (2024) 25:178

Page 6 of 14

Fig. 2 Maximum-Likelihood (ML) tree of subsampled *V. parahaemolyticus* WGS assemblies and identified virulence factor (VF) genes. All identified VFs are classified per subcategory and clades were colored using the color code presented by Yang et al., 2019 [21]. A repeated metadata band is represented highlighting the clinical (red) or environmental (blue) type of the samples. Our three new strains and the reference strain (*V. parahaemolyticus* RIMD 2210633; NC_004603 and NC_004605) were indicated with blue and red circles, respectively

possess the T6SS. Overall the presence of T6SS-associated genes was estimated to range between 56.5% and 58.6%. Comparable to the exclusiveness of *pirA/pirB* or *tdh/trh* genes within a single *V. parahaemolyticus* genome, exotoxin genes *vpm* (a metalloprotease with an estimated prevalence of 64.5%) and *vppC* (a collagenase with an estimated prevalence of 33.7%) were shown to never occur together in any of the studied strains. Furthermore, only 1.8% of *V. parahaemolyticus* are suggested to be negative for both exotoxins, the genotype to which our new *V. parahaemolyticus* PV1 strain also belongs. To completion, virulence factors involved in adherence, motility, and nutritional metabolism were included (Fig. 2 and Additional file 1: Table S5).

To understand the involvement of other genetic mediators in both AHPND and human acute gastroenteritis phenotypes, a gene-based genome-wide association study (GWAS) was performed for the *pirAB* and *tdh/trh* genes. The top 75 associated genes for *pirA* and *tdh* are highlighted in Table 2, whereas the results for *pirB* and

trh are enclosed as Additional file 1: Table S8 and S10. Of note, for *pirAB* GWAS outputs, the *pirAB*-plasmid associated genes were filtered out, as it is evident that a plasmid-encoded gene will be highly associated with other plasmid-located genes. A complete overview can be found in Additional file 1: Table S7-S10. From our 75 top genes associated with *pirA* presence, 23 were proteins with predicted domains, a domain of unknown function (DUF), or hypothetical proteins. The remaining 52 non-*pirA* plasmid associated genes comprised proteins belonging to the family of transposable elements (*n*=8), outer membrane proteins and fimbriae (*n*=9, including ompA-like and OMP-b-brl proteins), metabolic enzymes (*n*=6, including Glycosyltransferase, maltose ABC transporter permease *MalF*, maltose/maltodextrin ABC transporter substrate-binding protein *MalE*, Acyl-CoA hydrolase, and Glucose-6-phosphate isomerase) phage-associated proteins (*n*=4), recombinases (*n*=3), restriction endonucleases (*n*=2), transcriptional regulators (*n*=2), Immunoglobulin-like proteins (*n*=2), the YjbH

Vandeputte *et al. BMC Genomics*   (2024) 25:178

Page 7 of 14

Table 2 Top 75 associated genes with *pirA* and *tdh* presence. Gene annotations are shown along with their UniParc accession where available. For some genes, only a UniRef accession was available. Bars indicated number of isolates present/absent in positive (*pirA* + or *tdh* +) or negative (*pirA*- or *tdh*-) strains. For *pirA*, the top 75 non-*pirA* plasmid-associated genes are highlighted

Annotation	UniParc	PirA+ Present	PirA- Present	PirA+ Absent	PirA- Absent
PapD-N domain-containing protein	UPI00046EB487	32	40	15	949
Fimbrial biogenesis outer membrane usher protein	UPI00046F009C	27	25	20	964
OmpA family protein	UPI00046F7C87	28	36	19	953
YjbH	UPI000D52F708	21	10	26	979
DUF1460 domain-containing protein	UPI012FB9656	24	25	23	964
Bacterial Ig-like domain (Group 2)	UPI013311BBD	23	24	24	965
Ig-like protein group 2	UPI01331B2CE	23	25	24	964
Transcription-associated protein 1	UPI0004700E33	23	25	24	964
Type II restriction endonuclease	UPI0003AAA711	14	0	33	989
Recombinase family protein	UPI0003C79CAC	24	35	23	954
Fimbrial protein	UPI00046F1961	22	25	25	964
Ycel domain-containing protein	UPI0004718AC2	22	25	25	964
Toxin CcdB	UPI0003A8C183	24	36	23	953
Transposase	UPI000A2FC9B1	22	26	25	963
Resolvase	UPI01E3E5F61	23	32	24	957
OMP-b-brl domain-containing protein	A0A72ZMYF8 (UniRef50)	21	22	26	967
Type II secretion system protein	UPI000461175	21	22	26	967
DUF2281 domain-containing protein	UPI0003A4325B	23	34	24	955
Glycosyltransferase	UPI00044E2A43	20	19	27	970
Plasmid maintenance protein CcdB	UPI0002D42A3E	23	35	24	954
Transposase	UPI0003A9C128	23	35	24	954
recombinase family protein	UPI00B2ADD7B	22	30	25	959
Transposase	UPI0003B926E	21	25	26	964
OmpA-like domain-containing protein	UPI0003A1C579	22	32	25	957
DUF3085 domain-containing protein	UPI0003A4B561	23	38	24	951
Big-5 domain-containing protein	UPI00132F36C3 (UniRef50)	20	22	27	967
Phage protein	UPI0002E7A999	22	33	25	956
Glycos-transf-1 domain-containing protein	UPI000D52F22B	21	28	26	961
maltose ABC transporter permease MalF	UPI000205D732	46	381	1	608
CpmJ protein	A0A72ZMYC9 (UniRef50)	20	24	27	965
Peptidase-M23 domain-containing protein	UPI000472D0C3	22	35	25	954
endoribonuclease MazF	UPI0003C7952F	23	43	24	946
Transposase	UPI0002F57C4F	21	31	26	958
OmpA family protein	UPI0002E3D661	22	37	25	952
Phage tail protein	UPI000308572F	22	37	25	952
Transposase	A0A1Y1BA50 (UniRef50)	20	26	27	963
LHH domain-containing protein	UPI00046F4372	20	26	27	963
AbrB/MazE/SpoVT family DNA-binding domain-containing protein	UPI0003766188	23	44	24	945
P-type conjugative transfer ATPase TrbB	UPI000446C308	14	4	33	985
OmpA family protein	UPI0003A97109	22	38	25	951
HTH-8 domain-containing protein	UPI00046EA87B	13	2	34	987
OMP-b-brl domain-containing protein	UPI0004720026	20	27	27	962
AAA-31 domain-containing protein	UPI0004727E24	15	7	32	982
HTH-8 domain-containing protein	UPI0004472441A	22	40	25	949
dTDP-glucose 4,6-dehydratase	UPI00029A47D	16	11	31	978
Transposase	UPI0132F20BA	19	25	28	964
maltose/maltodextrin ABC transporter substrate-binding protein MalE	UPI00000A8FF	45	383	2	606
DUF5405 domain-containing protein	A0A0C1Z5U8 (UniRef50)	19	26	28	962
Acyl-CoA hydrolase	UPI0019147EB	42	308	5	681
Transcriptional regulator	UPI00039F506E	19	27	28	962
Lipoprotein	UPI00046E999F	18	22	29	967
Phage terminase large subunit	UPI01C4B80B	17	18	30	971
Restriction endonuclease	UPI0044BEE10	17	18	30	971
LHH domain-containing protein	UPI001543196	19	28	28	961
Transmembrane protein	UPI00044A3D6E	14	7	33	982
Integrase	UPI0039C77A7	20	34	27	955
Resolvase	UPI0003A3ED15	18	23	29	966
Resolvase	UPI0003C79669	18	23	29	966
Lipoprotein	UPI0003FD91EC	17	19	30	970
Transposase	UPI004260FA6	17	19	30	970
Cell wall anchor protein	UPI00044E3E60	15	11	32	978
hypothetical protein	UPI0111CD3E9 (UniRef50)	15	11	32	978
Spy/CpxP family protein refolding chaperone	UPI00047057FF	19	30	28	959
Methyltransferase	UPI0003F6690D	15	12	32	977
Tyr recombinase domain-containing protein	UPI003D59876	23	59	24	930
SUFU domain-containing protein	UPI0001543197	17	21	30	968
Acyl-CoA hydrolase	UPI000000A8FB	5	657	42	332
DUF2169 domain-containing protein	UPI000205EB96	47	489	0	500
DUF1566 domain-containing protein	UPI0003A80546	18	27	29	962
DUF6602 domain-containing protein	UPI0039AE3FE	15	13	32	976
Cytochrome b561 domain-containing protein	UPI051D16AD11	22	54	25	935
IS630 family transposase	UPI0003C7C0B9	18	28	29	961
Inovirus Gp2 family protein	UPI0038E2C2A	28	111	19	878
Phage baseplate protein	UPI000A37AD03	16	18	31	971
Glucose-6-phosphate isomerase	UPI00177DBC35	12	4	35	985

Annotation	UniParc	tdh+ Present	tdh- Present	tdh+ Absent	tdh- Absent
Putative integral membrane protein	UPI000409E850	183	92	10	751
Transcriptional regulator VspR	UPI00038E4F99	183	94	10	749
TniQ family protein	A7N6M5 (UniRef50)	182	91	11	752
Integral membrane protein	UPI000000A8EF	183	95	10	748
Superfamily II DNA or RNA helicase	UPI0002A56107	182	92	11	751
type I-F CRISPR-associated endoribonuclease Cas6/Csy4	UPI0046FA303	182	93	11	750
DUF4115 domain-containing protein	UPI000000A8F5	181	91	12	752
CRISPR-Cas system type I-F effector complex large subunit Cas8f	UPI000156475A	176	85	17	758
Secreted protein	UPI000A88A608	94	3	99	840
Replication protein	UPI000205F048	103	17	90	826
thermostable direct hemolysin TDH	UPI0003E26189	96	10	97	833
Transposase	UPI000205CB0B	92	8	101	835
Integrase	UPI000205CB0C	91	8	102	835
ISNCY family ISVbsp4 transposase	UPI000205CAFD	95	12	98	831
Ion-trans domain-containing protein	UPI00038E30DF	96	14	97	829
type I-F CRISPR-associated protein Csy3	UPI000205CB06	102	22	91	821
IS481 family transposase	UPI00038E1F51	94	16	99	822
Endonuclease	UPI000205CB0D	98	22	95	821
thermostable direct hemolysin TDH	UPI01317592E5	73	0	120	843
DUF4347 domain-containing protein	UPI000000A871	112	50	81	793
IS4 family transposase	A0A812FEA4 (UniRef50)	69	2	124	841
BIG2 domain-containing protein	UPI00038E2C19	90	25	103	818
Lytic transglycosylase	UPI000205CAFB	79	13	114	830
Molecular chaperone	UPI013762AA9	100	43	93	800
TDH-related hemolysin TRH1	N/A	103	51	90	792
IS4 family transposase	N/A	63	1	130	842
Peptidase M60 domain-containing protein	UPI000000A8E7	66	3	127	840
Acetyl-CoA acetyltransferase	UPI0003592351	99	45	94	798
3-keto-5-aminohexanoate cleavage protein	UPI00040861E3	94	41	99	802
thermostable direct hemolysin A	P19250 (UniRef50)	57	1	136	842
KTSC domain-containing protein	UPI0001BC70AE	55	1	138	842
GIY-YIG nuclease family protein	UPI0001BC70E3	103	71	90	772
Integrase catalytic domain-containing protein	UPI00038E343E	99	63	94	780
Helix-turn-helix domain-containing protein	UPI0035909C8	95	56	98	787
putative transcriptional regulator, contains an XRE-type HTH domain	UPI0001BC6A67	75	23	118	820
Acyl-CoA hydrolase	UPI0019147E8	0	350	193	493
Phage protein	UPI0035910BB	56	4	137	839
SH2 domain-containing protein	UPI000B776F85	102	77	91	766
Nterm-IS4 domain-containing protein	A0A286SFT5 (UniRef50)	79	33	114	810
Type III secretion protein	UPI000205C720	106	89	87	754
DNA-binding protein	UPI000205C715	106	90	87	753
Secreted effector protein	UPI000205C717	106	90	87	753
AraC family transcriptional regulator	UPI000205C712	107	93	86	750
Type III secretion system, LEE associated	UPI000359191D	106	91	87	752
Competence protein	UPI000205C71E	105	89	88	754
EAL domain-containing protein	UPI000205CAF8	105	89	88	754
Transcriptional regulator	UPI000205C71D	104	87	89	755
Transcriptional regulator	UPI0009C64951	106	92	87	751
ABC-type nickel transporter, ATP-binding protein NikE	UPI000205C711	107	95	86	748
Transposase	UPI001A342A14	81	40	112	803
Urease accessory protein UreF	UPI00035926DD	107	96	86	747
DUF1810 domain-containing protein	UPI001919479B	105	91	88	752
Putative type III secretion system apparatus protein VscR2	UPI000205C724	106	94	87	749
Type III secretion system apparatus protein VscT2	UPI00038E1301	106	94	87	749
ATP-dependent exonuclease	UPI00038E1663	106	94	87	749
urease accessory protein alpha	UPI000205F04E	107	97	86	746
ABC-type nickel transporter, ATP-binding protein NikD	UPI000205C710	107	97	86	746
ABC-type nickel transporter, transmembrane protein NikB	UPI000205C70E	107	97	86	746
ABC-type nickel transporter, periplasmic binding protein NikA	UPI000205C70D	107	97	86	746
urease accessory protein UreD	UPI00000AF28B	107	97	86	746
urease subunit beta	UPI000205F04D	107	97	86	746
Urease accessory protein UreE	UPI000205F04F	107	97	86	746
Putative D,D-dipeptide transport system permease protein DdpC	A0A286SF55 (UniRef50)	107	97	86	746
Putative type III secretion system translocon protein VopB2	UPI000205CAED	103	87	90	756
Dimethyladenosine transferase	UPI000205CAEB	105	92	88	751
CBS domain-containing protein	UPI000205CAF9	62	13	131	830
FliH domain-containing protein	UPI0001914799	105	94	88	749
Putative type III secretion system apparatus protein VscJ2	UPI0003C820A5	105	94	88	749
urease accessory protein UreG	UPI000205F04B	105	94	88	749
urease subunit gamma	UPI0000B5551	106	97	87	746
Putative type III secretion system apparatus protein VcrD2	UPI000205C52B	102	87	91	755
Internalin	UPI0038E37E1	61	11	132	830
t-SNARE coiled-coil-like γ domain-containing protein	D0I6B1 (UniRef50)	105	95	88	748
Regulator	UPI004166C53	106	98	87	745
Multidrug DMT transporter permease	UPI000205C718	102	88	91	755

adaptor protein, the toxin *ccdB*, the endoribonuclease *mazF*, a lipoprotein, a Spy/CpxP family protein refolding chaperone, and methyltransferase. A GWAS analysis for *pirB* resulted in the same set of associated genes (Additional file 1: Table S8), with the addition of some extra associated genes (maltose/maltodextrin ABC transporter substrate-binding protein MalE, a holin (UPI-0001BC6B02), Lipoprotein GfcB, amongst other proteins with predicted/DUF domains). Of note, these genes were also found in the *pirA* non-plasmid associated top 100 list (indicated in red in Additional file 1: Table S7 and S8).

For the *tdh* gene, 18 of the top 75 associated genes were proteins with putative functions, a predicted domain, or DUF. Interestingly, the GWAS confirms the association of Type III secretion genes (*n*=7, including *vcrD2*, *vopB2*, *vscJ2*, *vscR2*, and *vscT2*) with the presence of the *tdh* gene. Next to this also urease-associated genes (*n*=7, including urease subunit alpha, beta, and gamma, along with accessory genes *ureDEFG*), transcriptional regulators (*n*=6, including *vspR*), transposable elements (*n*=6), metabolic enzymes (*n*=4, including acetyl-CoA acetyltransferase, 3-keto-5-aminohexanoate cleavage protein, acyl-CoA hydrolase, dimethyl adenosine transferase), an ABC-type nickel transporter operon (*n*=4, *nikABDE*), a type I-F CRISPR-associated endoribonuclease Cas6/Csy4 and CRISPR-Cas system type I-F effector complex large subunit Cas8f, an uncharacterized secreted (effector) protein (UPI000A88A608 and UPI000205C717), a lytic trans glycosylase, a competence protein, an internalin, and a multidrug DMT transporter permease were found to be associated with *tdh* positive strains. Again, comparable results were obtained for the GWAS using *trh* positivity. Still some extra genes popped up, including seven extra Type III secretion genes, an *OmpA*-like domain-containing protein (UPI0001914797), two transcriptional regulators, a virulence-associated protein (UPI0005B6DDF1), cytotoxin (UPI000205CAF6; cytotoxic necrotizing factor Rho-activating domain-containing protein), and

Vandeputte *et al. BMC Genomics*   (2024) 25:178

Page 8 of 14

hemolysin B subunit protein (UPI000359204B), the Type VII secretion protein EssB, and the RTX toxin-activating lysine-acyltransferase, amongst other proteins with putative functions, a predicted domain, or DUF. Comparable to the *pirAB* GWAS outputs, most of these *trh*-associated genes were also found in the top 100 list of *tdh*-associated genes. A complete overview of the GWAS results and associated statistics is given in Additional file 1: Table S9 and S10, with extra *trh*-associated genes indicated in red.

Discussion

The three strains sequenced in this study were confirmed to be of the species *Vibrio parahaemolyticus*, carrying a pVA1-like plasmid containing the AHPND-related *pirAB* genes. The genomes of these strains, along with a representative subsample of 1,036 (of a total of 8,897) *V. parahaemolyticus* genomes available on NCBI, were analyzed to strengthen our understanding of AHPND and *tdh* + /*trh* + genotypes. Furthermore, we showed that long-read ONT sequencing can deliver complete circular *V. parahaemolyticus* genome and plasmid assemblies.

The sampled genomes were screened for the presence of ARGs and other virulence genes. In general, an average of 5 ± 2 (SD) ARGs per *V. parahaemolyticus* genome could be identified. For the three newly sequenced strains, five ARGs were identified in the genomes of *Vp* TW01 and *Vp* PV1, but eight ARGs for *Vp* M0904. This was also reflected in the phenotypical antimicrobial susceptibility testing (AST), where *Vp* M0904 showed resistance to four antibiotics (out of the panel of 13), whereas *Vp* TW01 and *Vp* PV1 were only resistant against a single antibiotic drug (and one intermediately sensitive) antibiotic. As strain *Vp* M0904 is resistant to antimicrobials of three different categories (β-lactams, tetracyclines, and trimethoprim-sulfamethoxazole), it is regarded as a multidrug-resistant (MDR) organism. Our phenotypic observations against ampicillin and piperacillin could be linked with the presence of specific ARGs, namely the β-lactamases $bla_{CARB-21}$ and $bla_{CARB-23}$. Resistance towards tetracycline and trimethoprim-sulfamethoxazole is suggested to be due to the chromosomal presence of a *tet(B)* and *dfrA6* gene. Unfortunately, substantial phenotypic metadata is lacking for the available genomes on NCBI. However, screening the subsampled genomes revealed a high estimated prevalence of these β-lactamases in the *V. parahaemolyticus* population (54.9% and 44.4% for $bla_{CARB-21}$ and $bla_{CARB-23}$, respectively). Available literature also suggests a high resistance prevalence towards β-lactams, which we phenotypically confirmed in this study by the resistance of all three strains towards ampicillin and piperacillin. Larger-scale (*n* > 36) phenotypical studies are confirming this too, with high reported resistance prevalence towards both penams [*e.g.,* 45–100% to ampicillin [6, 28, 30–33]

and 68–100% to penicillin [30–32]] and cephalosporins [*e.g.,* 49.6–84% to cefazolin [31, 34], 52–73% to cefotaxime [33, 35] and 24–82% to cephalothin [30, 34]].

An estimated 3.1% of strains showed the presence of the *tet(B)* gene, with a lower prevalence of other tetracycline resistance-associated genes such as *tet(59)*, *tet(A)*, *tet(C)*, *tet(E)*, *tetM*. Comparable results were obtained for *dfrA6* prevalence (4.0%), with the rare identification of *dfrA1*, *dfrA16*, *dfrA23*, *dfrA27*, and *dfrA32* potentially conferring strains resistant to trimethoprim-sulfamethoxazole. In addition, even though no phenotypic resistance was observed in our strains, ARG prevalence suggests low prevalence of resistance towards aminoglycosides (aminoglycoside acetyltransferase, nucleotidyltransferase, and phosphotransferase genes) and fluoroquinolones (quinolone resistance genes). This is again reflected in phenotypical studies on *V. parahaemolyticus* clinical and environmental strains, where low prevalence of resistance towards tetracyclines [*e.g.* 0.3% to doxycycline [30, 31, 36], 0–15% to tetracycline [27, 28, 30–33, 35, 36]], trimethoprim-sulfamethoxazole [0–7% [27, 30, 31, 33]], and fluoroquinolones (*e.g.,* 0% to ciprofloxacin [27, 30] and 0–1% to levofloxacin [27, 33]] were reported. However, reported resistance prevalence against aminoglycosides is very variable [*e.g.,* 0–64% to amikacin [27, 28, 30, 33] and 15–50% to kanamycin [30, 33, 34]]. It should be noted that resistance to aminoglycosides is often the result of mutations of the ribosome [37], a factor that was not included in our analysis. As a result, there is a possibility of underestimating the extent of this resistance. Furthermore, most of the phenotypical AST screening studies focused on samples from one geographical region, and there might be large differences between *V. parahaemolyticus* populations of different locations. Furthermore, differences in applied AST screening methods (*i.e.,* disc diffusion testing v*ersus* broth microdilution method), as well as sample types (*e.g.,* clinical, environmental, seafood or surface water samples, among others), differ between studies, possibly resulting in large differences and potential underestimations of actual resistance prevalence towards certain antibiotics. Hence, using WGS-based ARG genotyping in *V. parahaemolyticus* seems to be a good predictive tool to identify, study, understand, and follow up on the dissemination of drug-resistance genotypes in the population. Many studies reported the presence of at least one multidrug-resistant isolate like *Vp* M0904 [27, 28, 38]. These isolates represent an even bigger threat to aquaculture and human health, as they are more difficult to control with antibiotics, and are an important reservoir of ARGs. This is also highlighted in the identification of a *V. parahaemolyticus* cluster showing the presence of a putative plasmid carrying multiple

Vandeputte *et al. BMC Genomics*   (2024) 25:178

Page 9 of 14

ARGs. Most of the genomes presented within this genetic clade originate from China and were part of an unpublished study on the occurrence and genetic environments of $bla_{CTX-M-14}$ among foodborne *Vibrio* spp. (PRJNA622672). Still, researchers should be encouraged to link phenotypic and genotypic observations and be aware of potential ARGs not contributing to resistance phenotypes and/or being the result of contaminating contigs in the assembly. To completion, point mutations (*e.g.,* in rRNA, *gyrA*, *parC* genes) should also be considered as a potential source of resistance, which requires accurate metadata and genomics to establish proper associations [39, 40]). This was highlighted in the very high prevalence of ARGs *CRP* (99.7%), *msbA* (99.0%), *tet(35)* (99.3%), and *ugd* (84.4%). *CRP* is suggested to be involved in macrolide, penam, and fluoroquinolone resistance [41], *msbA* in nitroimidazole resistance [42], *tet(35)* in tetracycline resistance and *ugd* in resistance against peptide antibiotics [43], but do not seem to induce resistance in *V. parahaemolyticus*. It is plausible that these are so-called 'silent' genes, that might only be activated in specific conditions or expressed at low levels [28, 44]. The high prevalences of especially *CRP*, *msbA*, and *tet(35)* ($\geq$99%) suggests that these might be intrinsic resistance genes [45]; however, their relatively low nucleotide identity (*CRP* 80.0%$\pm$0.1%, *msbA* 66.5%$\pm$0.2%, *tet(35)* 82.6%$\pm$0.2%, and *ugd* 69.7%$\pm$0.9%) compared to other ARGs suggests that these genes might be degenerated and have lost functionality over time. In contrast, all other identified ARGs are suggested to be acquired, as these are strain-specific [45].

Next, screening of all available *Vibrio* spp. genomes showed an overall low (0.57%) prevalence of *pirA* and/or *pirB* genes, and a prevalence of 4.8% and 4.3% in *V. parahaemolyticus*, respectively. In accordance with previous research [19], the two Asian strains *Vp* TW01 and PV1 did not possess the Tn3 transposon on their plasmid, while the Latin-American (Mexican) strain *Vp* M0904 did. The *pirAB* region is usually flanked by two identical IS5 family ISVa2 transposases in opposite directions that form a composite transposon called Tn903 or Tn6264 [19]. This region was conserved in the three newly sequenced strains. As observed before [46], an extra IS5 family transposon was found in pVPR14 immediately upstream of the *pirA* coding sequence. Interestingly, this insertion seems to have no effect on the transcription of the genes, but instead inhibits translation of the PirAB proteins [46]. Furthermore, compared to the other analyzed plasmids, the *pirAB* region between the two IS5 family transposons of the pVA1 plasmid is inverted compared to the other plasmids. Here our plasmid screening and characterization showed a highly conserved backbone of the *pirAB*-encoding plasmid, highlighting the

biggest genetic diversity at the Tn3 transposon region. Still, it is important to note that our plasmid analysis is limited to the available data in NCBI as insufficient sequencing depth might result in the lack of data on plasmids. Also, the choice of sequencing platform affects the actual recovery of plasmid information, hence care should be taken when performing this type of analyses on genome assemblies available in the public domain. The choice of sequencing technology might also impact the potential duplication and orientation of transposons due to their repetitive regions [47–50].

The three newly sequenced *Vp* strains, as well as the subsampled genomes, were screened for the presence of virulence-associated genes. As expected, *Vp* M0904, TW01, and PV1 possessed the *pirA* and *pirB* genes, and lacked the *tdh* and *trh* hemolysin genes. Furthermore, the presence of *tdh* was strongly associated with the presence of genes linked to the T3SS2. These observations confirm the suggestions made before, that: (i) AHPND genotypes are not pathogenic to humans since they lack the *tdh* and *trh* genes, along with the absence of the T3SS2 and (ii) there is a correlation between the presence of *tdh* and *trh* genes and the T3SS2 [22, 23]. Genes related to the T6SS on the first chromosome (T6SS1) were absent in the VppUS2 clade, and present in only a subpart of the other three clades VppAsia, VppX, and VppUS1. Although previously it has been suggested that T6SS1 is present in all AHPND and human clinical isolated, and not in non-pathogenic environmental strains [22], this is not supported by our results.

Many *Vp* AHPND strains have been identified and different rates of mortality in infected shrimps were reported [51], indicating that the PirAB toxin is not the only factor contributing to virulence in shrimp. The same holds for gastroenteritis in humans, where illness is linked with the presence of *tdh/trh* and T3SS2, but other factors might contribute to virulence. To gain more insight into contributing factors, a GWAS was performed for the *pirAB* and *tdh/trh* genes. For *pirAB* it is important to note that most associations were with the plasmid genes themselves, as it is evident that a plasmid-encoded gene will be highly associated with other plasmid-located genes. These genes have been filtered out in our analyses so we could focus better on chromosome-encoded associations. Many transposon- (n=8), phage-related (n=4) proteins were identified. Both are often associated with the insertion of new genes primarily by horizontal gene transfer and expressing novel pathogenic properties [51], and CRISPRs are suggested to be able to prevent the invasion of prophages into bacteria, being a barrier of defense to fight against foreign DNA [51]. A previous study indicated the lack of these CRISPR elements in *Vp* AHPND strains, allowing the prophage insertion of

Vandeputte *et al. BMC Genomics*   (2024) 25:178

Page 10 of 14

virulence genes, possibly contributing to the virulence of *Vp* AHPND strains [51]. These observations seem to be confirmed in the present GWAS. Multiple outer membrane proteins (OMPs) were also among the associated genes, of which several were related to OmpA. For many Gram-negative bacteria, this is a key virulence factor involved in bacterial biofilm formation, eukaryotic cell infection, antibiotic resistance, and immunomodulation [52]. In *V. parahaemolyticus*, several OMPs have been shown to be involved in ampicillin resistance [53] and osmoregulation [54]. Furthermore, a Spy/CpxP family protein refolding chaperone was present in the GWAS of *pirAB*, and is demonstrated to protect OMPs against protein-unfolding stress in *E. coli* [55]. Possibly, it is important in the protection against unfolding of the previously mentioned OmpA and other OMPs. The association with two immunoglobulin-like domain-containing proteins is seen, which are often surface proteins involved in cell-to-cell recognition, adhesion, biofilm formation, and conjugative transfer [56, 57]. These domains are widely present in numerous proteins, and several have been identified in *V. parahaemolyticus*, such as adhesive factor VpadF [56, 57]. Association with toxin gene *ccdB* indicated the importance of the toxin-antitoxin (TA) system ccdA-ccdB, consisting of the stable toxin ccdB and less stable antitoxin ccdA to neutralize the toxin. It is defined as a plasmid maintenance system, acting through post-segregational killing of plasmid-free cells [58, 59]. The endoribonuclease *mazF* is also part of the chromosomal type II TA system *mazEF*. It has been identified in multiple bacteria where it regulates virulence factors, such as biofilm production in *E. coli* and in *Staphylococcus* (*S.*) *aureus*, but it may be involved in other functions such as antibiotic resistance and programmed cell death under stressful conditions [59, 60]. However, its role in *Vibrionaceae* has not been studied to the author's knowledge.

The GWAS of the *tdh/trh* genes, related to human gastroenteritis, confirmed the association with the T3SS2, as seven genes were present in the top 75. Additionally, multiple urease-associated genes are linked to *tdh*, which are virulence factors in various pathogenic bacteria [61]. The ability of ureases to raise the pH in the immediate environment inside the host contributes to the survival of the bacteria in the host digestive tract [38]. Although there seems to be an association between *tdh* and urease genes, phenotypically this association seems to be more correct for *trh* since multiple reports have mentioned that all *trh* genotypes were urease positive, which was not the case for *tdh* [62, 63]. The association with *trh* was also confirmed in our GWAS, and the urease-related genes were ranked higher in the top 75 than for *tdh* (7 genes ranked between place 27 and 41 for *trh* compared to 7 genes ranked between place 51 and 70 for *tdh*). Furthermore,

it was demonstrated that the genes for an ATP-binding cassette-type nickel transport system, which may play a role in nickel transport through the bacterial cytoplasmic membrane, are located adjacent to the urease gene cluster on the genome of *V. parahaemolyticus*, and that this region is in close proximity to the *trh* gene [63, 64]. Interestingly, this nickel transport system is indeed associated with *tdh* and *trh* according to the current GWAS. Possibly, this system aids in providing nickel for incorporation into the metallocentre within the active sites of the ureases [63]. Another interesting finding from the *tdh*-based GWAS was the association of a type I-F CRISPR-associated endoribonuclease Cas6/Csy4. This is related to a variant of the subtype I-F CRISPR-Cas system called the 'minimal' type, which showed a strong association with *tdh* in previous reports as well [65, 66]. Furthermore, genes for TniQ and a Cas8f-like protein also returned in our GWAS, which are all possibly linked with the minimal type I-F system. Interestingly, it was shown that this system was present within a Tn7-like transposon in *V. parahaemolyticus* and that the pathogenicity island containing the T3SS2 was correlated [67]. Association with *trh* on the other hand with genes of this minimal type I-F CRISPR-Cas system was a lot lower. Furthermore, the presence of many transposable elements ($n=6$) as well as a competence protein is underlining the importance of genetic diversity that is associated with these genes.

Conclusion

This study has provided a comprehensive characterization of three AHPND-inducing *V. parahaemolyticus* strains and their associated PirAB-encoding plasmids. Additionally, to the author's knowledge, this is the first in-depth and comprehensive genome mining study on *V. parahaemolyticus* ARGs and virulence genes. The presence of ARGs and important virulence genes were examined and correlated with phylogenomic analysis of an extensive dataset of *V. parahaemolyticus* genomes. We recommend this approach to fellow researchers for rapid antibiotic resistance screening and to contain dissemination of the ARGs and resistance plasmids, particularly in light of the emergence of multidrug-resistant strains, as observed not only in this study but also in related studies. Furthermore, through GWAS, we have successfully identified known and novel genetic markers associated with the AHPND and *tdh*+/*trh*+ genotypes of *V. parahaemolyticus*. This discovery holds promise for the development of rapid and precise discriminatory tests targeting these genotypic variations. Collectively, these findings significantly enhance our comprehension of pathogenic *V. parahaemolyticus*, underscoring the importance of close monitoring of antimicrobial resistance of the bacteria.

Vandeputte *et al. BMC Genomics*   (2024) 25:178

Page 11 of 14

Methods

Bacterial strains and growth conditions

Three new *V. parahaemolyticus* strains were used in this study: M0904, isolated in northwestern Mexico (originally received from A.C. Mazatlàn unit of Aquaculture), TW01, isolated in Southern Thailand, and PV1, isolated in China (both received from Robins McIntosh). All three strains were isolated from AHPND-affected shrimp and were previously confirmed to produce PirAB by Western Blot [68]. One colony of each strain was inoculated in Marine Broth (MB) (Carl Roth – CP73.1) and grown overnight (6 ± 2 h) at 28°C with shaking at 140 rpm.

Antibiotic susceptibility testing

The three *V. parahaemolyticus* strains and the reference strain *Escherichia* (*E.*) *coli* ATCC 25922 (strain LMG8223 from the Belgian Coordinated Collection of Microorganisms BCCM) were tested for antibiotic resistance using the disc diffusion method, in compliance with the Clinical and Laboratory Standards Institute (CLSI) M45 guidelines for *Vibrio* spp. A total of 13 different antibiotics (all purchased from Oxoid Limited, UK) were tested: ampicillin (10 μg), amoxicillin-clavulanate (20/10 μg), ampicillin-sulbactam (10/10), piperacillin (100 μg), cefotaxime (30 μg), ceftazidime (30 μg), amikacin (30 μg), gentamicin (10 μg), tetracycline (30 μg), ciprofloxacin (5 μg), levofloxacin (5 μg), ofloxacin (5 μg) and trimethoprim-sulfamethoxazole (1.25/23.75 μg). Briefly, the bacteria were grown on Mueller–Hinton Agar (MHA, Carl Roth, Germany) plates after which a direct colony suspension was prepared in a 0.85% NaCl solution, and turbidity was adjusted to 0.5 McFarland standard. The MHA plates were inoculated with 100 μL of this suspension which was spread with a sterile triangle rod. Plates were allowed to dry for 5 to 10 min prior to applying the antibiotic disks onto the agar with sterile tweezers. The plates were incubated inverted at 35°C for 16 ± 2 h. The reference strain *E. coli* ATCC 25922 was used as a control to monitor the accuracy of the disk diffusion tests.

Genomic DNA extraction and whole genome long-read sequencing

A freshly grown overnight culture of each of the three strains was subjected to the isolation of High Molecular Weight (HMW) DNA using the DNA MiniPrep Kit (Zymo Research, USA) as described before [40]. Manufacturer's instructions were followed with the addition of a 30-min proteinase K (500 ng μL^{-1}; Promega, USA) treatment at 56°C after two cycles of bead bashing in a TissueLyzer (30 oscillations per minute for 5 min; Qiagen, Germany). The quality of the resulting DNA was confirmed on a NanoDrop device. Whenever A_{260}/A_{280} and/or A_{260}/A_{230} measures did not reach 1.7 or 1.5 cut-offs,

an extra DNA clean-up was performed using CleanNGS (CleanNA, The Netherlands) magnetic beads at a 1:1 ratio. For subsequent long-read Nanopore sequencing, 400 ng DNA was used in rapid library preparation (RBK-004), barcoding each sample for 48h sequencing on an R9.4.1 MinION flow cell. Data were acquired on a Grid-ION device (Oxford Nanopore Technologies Ltd., UK), supporting real-time, super accurate base calling, and demultiplexing with guppy (v6.1.5; ONT).

Genome assembly and annotation

The obtained reads were used in an *in-house* established bacterial genome assembly pipeline, which includes read quality checking (NanoComp v.1.10.0; [69]), read filtering (filtlong v0.2.1; –min_length 1,000 –keep_percent 95; https://github.com/rrwick/Filtlong), Trycycler subsampling (v0.5.3; –min_read_depth 50 –count 10 –genome_size 5 M; [47]) for independent genome assemblies using Flye (v2.9; –nano-hq -g 3m; [70]), raven (v1.8.1; [71]), wtdbg2 (v1.12; [72]), and miniasm_miniplish.sh (v0.3; -x ont -g 3m; https://github.com/rrwick/Minipolish) with *default* settings if not depicted. Subsequent trycycler commands were run as instructed using *default* settings, followed by read polishing using minimap2 (v.2.20; -a -x map-ont; [73]) and medaka (v.1.5.0; consensus –model r941_min_sup_g507 –batch_size 50; stitch –no-fillgaps; ONT). The quality of the resulting genomes was checked using CheckM (v.1.1.0; [74]), including 1,084 markers from 70 *Vibrio* spp. genomes. Species classification was done using rMLST via the pubMLST web interface [75]. Average Nucleotide Identity (ANI) of the three new strains against the *V. parahaemolyticus* reference strain RIMD 2210633 (NC_004605.1) was calculated using the ANI tool on the EZBioCloud web interface [76]. An overview of raw read statistics, quality checks, and NCBI accession numbers can be found in Additional file 1: Table S1.

Screening for ARGs and virulence factor, and phylogenomic analysis

A total of 8,897 *V. parahaemolyticus* genome assemblies were available on NCBI (accessed on 24/06/2023), which were subsampled to 1,036 genomes maintaining genomic diversity with minimal loss of resolution (Additional file 1: Table S2). Subsampling was done using a dendrogram, which was generated based on genomic blast (*default* settings), containing the sequences from *V. parahaemolyticus* based on its taxonomy ID 223926. While iteratively going through the nodes of this dendrogram, the average distance of each node to its leaves was calculated. Whenever this average distance dropped below 0.01, only one representative sequence from that node was retained. All genomes,

Vandeputte *et al. BMC Genomics*   (2024) 25:178

Page 12 of 14

including the newly obtained sequences were subjected to a virulence factor and ARG screening using Abricate (v.1.0.1; –minid 80 –mincov 80; https://github.com/tseemann/abricate) using the pre-built CARD [77] and a custom virulence factor database, which included a more extensive list of putative *V. parahaemolyticus* virulence-associated genes ($n = 301$) based on available literature. This list can be accessed in Additional file 1: Table S3. Complete Abricate outputs can be accessed in Additional file 1: Table S4 and S5. The subsampled dataset was also subjected to a single nucleotide polymorphism (SNP) based phylogenetic analysis as described by Kaas et al. [78]. In short, SNPs were identified by aligning all the sequences to the reference strain *V. parahaemolyticus* RIMD 2210633 (NC_004605.1), using the nucmer aligner from the MUMmer package (v3.1, -CIlrT; [79]). From this output, a concatenated alignment of the SNPs was produced, which was subjected to maximum-likelihood (ML) phylogenetic inference using IQtree (v.1.6.1; -m GTR+I+R -nt AUTO -bb 1000; [80]) applying the GTR+F+I model and 1,000 ultra-fast bootstraps (–ufboot). Final tree visualizations were done in iTOL (v.5; [81]).

Plasmid characterization and genome-wide association study

All *pirA* and/or *pirB* positive strains, as obtained by virulence factor screening, were analyzed separately, to study their relatedness and genomic landscape. First, all *pirA* and/or *pirB* positive plasmid sequences were subjected to an SNP-based alignment as described before to obtain an ML phylogenetic inference. Next, our newly sequenced strains, along with some type strains (pVPR14: CP028145 and pVA1: KP324996) were used for plasmid annotation using Bakta (v1.8.2.; –db db-full –compliant –genus Vibrio –species parahaemolyticus –gram -; [82]). This allowed to generate a custom plasmid gene database for an Abricate-based search (–minid 80 –mincov 80) of all *pirA* and/or *pirB*-associated plasmid sequences. Database entries and Abricate output can be found in Additional file 1: Table S6. In addition, a visualization could be made of the *pirA/pirB* plasmid landscape in our new isolates. Gene maps were obtained using clinker (v.0.0.28; [83]). In addition, all subsampled genomes and new isolates were subjected to whole genome annotation using Bakta, with identical settings as depicted before, for use in a genome-wide association study (GWAS). The latter was done using Roary (v.3.13.0; -p 8 -e -n -v; [84]) and Scoary (v.3.6.16; -e 1000 -n tree.newick; [85]), allowing the association of coding sequences based on any pheno- or genotypic feature. Here, *pirA/pirB/tdh/trh* presence (addressed 1) and absence (addressed 0) were used in the GWAS metadata file as obtained from the Abricate output and highlighted in green in Additional file 1:

Table S5. A complete overview of GWAS outputs can be found in Additional file 1: Table S7-S10.

Supplementary Information

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Laboratory of Immunology and Animal Biotechnology, Department of Animal Production and Aquatic Ecology, Faculty of Bioscience Engineering, Ghent University, Ghent, Belgium. [2]Laboratory of Aquaculture & Artemia Reference Center, Department of Animal Production and Aquatic Ecology, Faculty of Bioscience Engineering, Ghent University, Ghent, Belgium. [3]PathoSense BV, Lier, Belgium.

References
1. Kumar R, Ng TH, Wang HC. Acute hepatopancreatic necrosis disease in penaeid shrimp. Rev Aquac. 2020;12(3):1867–80.
2. Kumar V, Roy S, Behera BK, Bossier P, Das BK. Acute hepatopancreatic necrosis disease (Ahpnd): Virulence, pathogenesis and mitigation strategies in Shrimp aquaculture. Toxins. 2021;13(8):1–28.
3. Soto-Rodriguez SA, Lozano-Olvera R, Montfort GRC, Zenteno E, Sánchez-Salgado JL, Vibanco-Pérez N, Aguilar Rendón KG. New Insights into the Mechanism of Action of PirAB from Vibrio Parahaemolyticus. Toxins. 2022;14(4):243.
4. Han JE, Tang KFJ, Tran LH, Lightner DV. Photorhabdus insect-related (Pir) toxin-like genes in a plasmid of Vibrio parahaemolyticus, the causative agent of acute hepatopancreatic necrosis disease (AHPND) of shrimp. Dis Aquat Org. 2015;113(1):33–40.
5. Lee CT, Chen IT, Yang YT, Ko TP, Huang YT, Huang JY, et al. The opportunistic marine pathogen Vibrio parahaemolyticus becomes virulent by

Vandeputte *et al. BMC Genomics*   (2024) 25:178

Page 13 of 14

acquiring a plasmid that expresses a deadly toxin. Proc Natl Acad Sci USA. 2015;112(34):10798–803.

6. Muthukrishnan S, Defoirdt T, Ina-Salwany MY, Yusoff FM, Shariff M, Ismail SI, Natrah I. Vibrio parahaemolyticus and Vibrio harveyi causing Acute Hepatopancreatic Necrosis Disease (AHPND) in Penaeus vannamei (Boone, 1931) isolated from Malaysian shrimp ponds. Aquaculture. 2019;511:734227.

7. Dong X, Bi D, Wang H, Zou P, Xie G, Wan X, et al. pirABvp-Bearing Vibrio parahaemolyticus and Vibrio campbellii pathogens isolated from the Same AHPND-affected pond possess highly similar pathogenic plasmids. Front Microbiol. 2017;8:1859.

8. Ke HM, Prachumwat A, Yu CP, Yang YT, Promsri S, Liu KF, et al. Comparative genomics of Vibrio campbellii strains and core species of the Vibrio Harveyi clade. Sci Rep. 2017;7:41394.

9. Kondo H, Van PT, Dang LT, Hirono I. Draft genome sequence of non-Vibrio parahaemolyticus acute hepatopancreatic necrosis disease strain KC13.17.5, isolated from diseased shrimp in Vietnam. Genome Announc. 2015;3(5):e00978-15.

10. Liu L, Xiao J, Zhang M, Zhu W, Xia X, Dai X, et al. A Vibrio owensii strain as the causative agent of AHPND in cultured shrimp, Litopenaeus vannamei. J Invertebr Pathol. 2018;153:156–64.

11. Restrepo L, Bayot B, Arciniegas S, Bajaña L, Betancourt I, Panchana F, Reyes Muñoz A. PirVP genes causing AHPND identified in a new Vibrio species (Vibrio punensis) within the commensal Orientalis clade. Sci Rep. 2018;8(1):13080.

12. Wang H-C, Lin S-J, Mohapatra A, Kumar R, Wang H-C. A review of the functional annotations of important genes in the AHPND-Causing pVA1 Plasmid. Microorganisms. 2020;8(7):996.

13. Soto-Rodriguez SA, Gomez-Gil B, Lozano-Olvera R, Aguilar-Rendón KG, González-Gómez JP. Identification of new Vibrio campbellii strains harboring the pVA1 plasmid isolated from Penaeus vannamei postlarvae affected by outbreaks of acute hepatopancreatic necrosis disease (AHPND) in Mexico. Aquaculture. 2024;579:740221.

14. Dong X, Song J, Chen J, Bi D, Wang W, Ren Y, et al. Conjugative transfer of the PVA1-type plasmid carrying the PirABVP genes results in the formation of new AHPND-causing vibrio. Front Cell Infect Microbiol. 2019;9:195.

15. Aguilar-Rendón KG, Soto-Rodriguez SA, Gomez-Gil B, Lozano-Olvera R, Yáñez-Rivera B. Water microbiome dynamics of Pacific white shrimp Penaeus vannamei infected with Vibrio parahaemolyticus strains responsible for acute hepatopancreatic necrosis disease. Aquaculture. 2022;551:737871.

16. Xiao J, Liu L, Ke Y, Li X, Liu Y, Pan Y, Yan S. Shrimp AHPND-causing plasmids encoding the PirAB toxins as mediated by pirAB -Tn903 are prevalent in various Vibrio species. Nat Publ Group. 2016;2017:1–11.

17. Han JE, Tang KFJ, Aranguren LF, Piamsomboon P. Characterization and pathogenicity of acute hepatopancreatic necrosis disease natural mutants, pirABvp (-) V. parahaemolyticus, and pirABvp (+) V. campbellii strains. Aquaculture. 2017;470:84–90.

18. Kanrar S, Dhar AK. Complete Genome Sequence of a Novel Mutant Strain of Vibrio parahaemolyticus from Pacific White Shrimp (Penaeus vannamei). Genome Announc. 2018;6(24):e00497-18.

19. González-Gómez JP, Soto-Rodriguez S, López-Cuevas O, Castro-del Campo N, Chaidez C, Gomez-Gil B. Phylogenomic Analysis Supports Two Possible Origins for Latin American Strains of Vibrio parahaemolyticus Associated with Acute Hepatopancreatic Necrosis Disease (AHPND). Curr Microbiol. 2020;77(12):3851–60.

20. Han JE, Tang KFJ, Lightner DV. Genotyping of virulence plasmid from Vibrio parahaemolyticus isolates causing acute hepatopancreatic necrosis disease in shrimp. Dis Aquat Org. 2015;115(3):245–51.

21. Yang C, Pei X, Wu Y, Yan L, Yan Y, Song Y, et al. Recent mixing of Vibrio parahaemolyticus populations. ISME J. 2019;13(10):2578–88.

22. Li P, Kinch LN, Ray A, Dalia AB, Cong Q, Nunan LM, et al. Acute hepatopancreatic necrosis disease-causing Vibrio parahaemolyticus strains maintain an antibacterial type VI secretion system with versatile effector repertoires. Appl Environ Microbiol. 2017;83(13):e00737-17.

23. Osorio CR. T3SS effectors in Vibrios: Homology in sequence, diversity in biological functions? Virulence. 2018;9:721–3.

24. Okada N, Iida T, Park KS, Goto N, Yasunaga T, Hiyoshi H, et al. Identification and characterization of a novel type III secretion system in trh-positive vibrio parahaemolyticus strain TH3996 reveal genetic lineage and diversity of pathogenic machinery beyond the species level. Infect Immun. 2009;77(2):904–13.

25. Salomon D, Gonzalez H, Updegraff BL, Orth K. Vibrio parahaemolyticus Type VI secretion system 1 is activated in marine conditions to target bacteria, and is differentially regulated from system 2. PLoS ONE. 2013;8(4):e61086.

26. Lulijwa R, Rupia EJ, Alfaro AC. Antibiotic use in aquaculture, policies and regulation, health and environmental risks: a review of the top 15 major producers. Rev Aquac. 2019;12(2):640–63.

27. Stratev D, Fasulkova R, Krumova-Valcheva G. Incidence, virulence genes and antimicrobial resistance of Vibrio parahaemolyticus isolated from seafood. Microb Pathog. 2023;177:106050.

28. Changsen C, Likhitrattanapisal S, Lunha K, Chumpol W, Jiemsup S, Prachumwat A, et al. Incidence, genetic diversity, and antimicrobial resistance profiles of Vibrio parahaemolyticus in seafood in Bangkok and eastern Thailand. PeerJ. 2023;11:e15283.

29. Zaafrane S, Maatouk K, Alibi S, Mansour HB. Occurrence and antibiotic resistance of Vibrio parahaemolyticus isolated from the Tunisian coastal seawater. J Water Health. 2022;20(2):369–84.

30. Elmahdi S, DaSilva LV, Parveen S. Antibiotic resistance of Vibrio parahaemolyticus and Vibrio vulnificus in various countries: a review. Food Microbiol. 2016;57:128–34.

31. Tan CW, Rukayadi Y, Hasan H, Thung TY, Lee E, Rollon WD, et al. Prevalence and antibiotic resistance patterns of Vibrio parahaemolyticus isolated from different types of seafood in Selangor. Malaysia. Saudi J Biol Sci. 2020;27(6):1602–8.

32. Amalina NZ, Santha S, Zulperi D, Amal MNA, Yusof MT, Zamri-Saad M, Ina-Salwany MY. Prevalence, antimicrobial susceptibility and plasmid profiling of Vibrio spp. isolated from cultured groupers in Peninsular Malaysia. BMC Microbiol. 2019;19(1):251.

33. Lee LH, Ab Mutalib NS, Law JW, Wong SH, Letchumanan V. Discovery on antibiotic resistance patterns of Vibrio parahaemolyticus in Selangor reveals carbapenemase producing Vibrio parahaemolyticus in marine and freshwater fish. Front Microbiol. 2018;9:2513.

34. Xu X, Cheng J, Wu Q, Zhang J, Xie T. Prevalence, characterization, and antibiotic susceptibility of Vibrio parahaemolyticus isolated from retail aquatic products in North China. BMC Microbiol. 2016;16(1):1–9.

35. Letchumanan V, Pusparajah P, Tan LTH, Yin WF, Lee LH, Chan KG. Occurrence and antibiotic resistance of Vibrio parahaemolyticus from Shellfish in Selangor, Malaysia. Front Microbiol. 2015;6:1417.

36. Tan CW, Malcolm TTH, Kuan CH, Thung TY, Chang WS, Loo YY, et al. Prevalence and antimicrobial susceptibility of Vibrio parahaemolyticus isolated from short mackerels (Rastrelliger brachysoma) in Malaysia. Front Microbiol. 2017;8:1087.

37. Garneau-Tsodikova S, Labby KJ. Mechanisms of resistance to aminoglycoside antibiotics: Overview and perspectives. MedChemComm: Royal Society of Chemistry; 2016. p. 11–27.

38. Igbinosa EO, Beshiru A, Igbinosa IH, Ogofure AG, Uwhuba KE. Prevalence and Characterization of Food-Borne Vibrio parahaemolyticus From African Salad in Southern Nigeria. Front Microbiol. 2021;12:632266.

39. Bokma J, Vereecke N, Nauwynck H, Haesebrouck F, Theuns S, Pardon B, Boyen F. Genome-wide association study reveals genetic markers for antimicrobial resistance in mycoplasma bovis. Microbiol Spectr. 2021;9(2):e0026221.

40. Vereecke N, Botteldoorn N, Brossé C, Bonckaert C, Nauwynck H, Haesebrouck F, et al. Predictive power of long-read whole-genome sequencing for rapid diagnostics of multidrug-resistant brachyspira hyodysenteriae strains. Microbiol Spectr. 2023;11(1):e0412322.

41. Nishino K, Senda Y, Yamaguchi A. CRP regulator modulates multidruf resistance of escherichia coli by repressing the mdtEF Multidrug Efflux Genes. J Antibiot. 2008;61(3):120–7.

42. Singh H, Velamakanni S, Deery MJ, Howard J, Wei SL, Van Veen HW. ATP-dependent substrate transport by the ABC transporter MsbA is proton-coupled. Nat Commun. 2016;7:12387.

43. Lee H, Hsu FF, Turk J, Groisman EA. The PmrA-regulated pmrC gene mediates phosphoethanolamine modification of lipid A and polymyxin resistance in Salmonella enterica. J Bacteriol. 2004;186(13):4124–33.

44. Deekshit VK, Srikumar S. 'To be, or not to be'—The dilemma of 'silent' antimicrobial resistance genes in bacteria. J Appl Microbiol. 2022;133:2902–14.

Vandeputte *et al. BMC Genomics*   (2024) 25:178

Page 14 of 14

45. Munita JM, Arias CA. Mechanisms of Antibiotic Resistance. Microbiol Spectr. 2016;4(2):10.1128/microbiolspec.VMBF-0016-2015. https://doi.org/10.1128/microbiolspec.VMBF-0016-2015.

46. Caro LFA, Mai HN, Kanrar S, Cruz-Flores R, Dhar AK. A mutant of vibrio parahaemolyticus pirABvp (+) that carries binary toxin genes but does not cause acute hepatopancreatic necrosis disease. Microorganisms. 2020;8(10):1–13.

47. Wick RR, Judd LM, Wyres KL, Holt KE. Recovery of small plasmid sequences via oxford nanopore sequencing. Microb Genom. 2021;7(8):000631.

48. Arumugam K, Bessarab I, Haryono MAS, Liu X, Zuniga-Montanez RE, Roy S, et al. Recovery of complete genomes and non-chromosomal replicons from activated sludge enrichment microbial communities with long read metagenome sequencing. NPJ Biofilms Microbiomes. 2021;7(1):23.

49. Arredondo-Alonso S, Pöntinen AK, Cléon F, Gladstone RA, Schürch AC, Johnsen PJ, et al. A high-throughput multiplexing and selection strategy to complete bacterial genomes. Gigascience. 2021;10(12):giab079.

50. Johnson J, Soehnlen M, Blankenship HM. Long read genome assemblers struggle with small plasmids. Microb Genom. 2023;9(5):mgen001024.

51. Yu LH, Teh CSJ, Yap KP, Ung EH, Thong KL. Comparative genomic provides insight into the virulence and genetic diversity of Vibrio parahaemolyticus associated with shrimp acute hepatopancreatic necrosis disease. Infect Genet Evol. 2020;83:104347.

52. Nie D, Hu Y, Chen Z, Li M, Hou Z, Luo X, et al. Outer membrane protein A (OmpA) as a potential therapeutic target for Acinetobacter baumannii infection. J Biomed Sci. 2020;27:1–8.

53. Meng X, Huang D, Zhou Q, Ji F, Tan X, Wang J, Wang X. The influence of outer membrane protein on ampicillin resistance of vibrio parahaemolyticus. Can J Infect Dis Med Microbiol. 2023;2023:8079091.

54. Xu C, Ren H, Wang S, Peng X. Proteomic analysis of salt-sensitive outer membrane proteins of Vibrio parahaemolyticus. Res Microbiol. 2004;155(10):835–42.

55. He W, Yu G, Li T, Bai L, Yang Y, Xue Z, et al. Chaperone spy protects outer membrane proteins from folding stress via dynamic complex formation. mBio. 2021;12(5):e0213021.

56. Liu M, Chen S. A novel adhesive factor contributing to the virulence of Vibrio parahaemolyticus. Sci Rep. 2015;5:1–10.

57. Wang D, Wang H. In silico approach gives insights into ig-like fold containing proteins in vibrio parahaemolyticus: a focus on the fibrillar adhesins. Toxins. 2022;14(2):133.

58. Song X, Lin Z, Yuan W. Toxin–antitoxin systems in pathogenic Vibrio species: a mini review from a structure perspective. 3 Biotech. 2022;12(6):125 Springer Science and Business Media Deutschland GmbH.

59. Kamruzzaman M, Wu AY, Iredell JR. Biological functions of type ii toxin-antitoxin systems in bacteria. Microorganisms. 2021;9(6):1276.

60. Amitai S, Yassin Y, Engelberg-Kulka H. MazF-mediated cell death in Escherichia coli: A point of no return. J Bacteriol. 2004;186(24):8295–300.

61. Konieczna I, Arnowiec P, Kwinkowski M, Kolesiska B, Frczyk J, Kamiski Z, Kaca W. Bacterial urease and its role in long-lasting human diseases. Curr Protein Pept Sci. 2012;13(8):789–806.

62. Sujeewa AKW, Norrakiah AS, Laina M. Prevalence of toxic genes of Vibrio parahaemolyticus in shrimps (Penaeus monodon) and culture environment. Int Food Res J. 2009;16:89–95.

63. Park K-S, Iida T, Yamaichi Y, Oyagi T, Yamamoto K, Honda T. Genetic Characterization of DNA Region Containing the trh and ure Genes of Vibrio parahaemolyticus. Infect Immun. 2000;68(10):5742–8.

64. Ellingsen AB, Olsen JS, Granum PE, Rørvik LM, González-Escalona N. Genetic characterization of trh positive Vibrio spp. isolated from Norway. Front Cell Infect Microbiol. 2013;3:107.

65. Baliga P, Shekar M, Venugopal MN. Investigation of direct repeats, spacers and proteins associated with clustered regularly interspaced short palindromic repeat (CRISPR) system of Vibrio parahaemolyticus. Mol Genet Genomics. 2019;294(1):253–62.

66. Sun H, Li Y, Shi X, Lin Y, Qiu Y, Zhang J, et al. Association of CRISPR/cas evolution with vibrio parahaemolyticus virulence factors and genotypes. Foodborne Pathog Dis. 2015;12(1):68–73.

67. McDonald ND, Regmi A, Morreale DP, Borowski JD, Fidelma Boyd E. CRISPR-Cas systems are present predominantly on mobile genetic elements in Vibrio species. BMC Genomics. 2019;20(1):105.

68. Vandeputte M, Verhaeghe M, Willocx L, Bossier P, Vanrompay D. Bovine Lactoferrin and Hen Ovotransferrin Affect Virulence Factors of Acute Hepatopancreatic Necrosis Disease (AHPND)-Inducing Vibrio parahaemolyticus Strains. Microorganisms. 2023;11(12):2912.

69. De Coster W, Rademakers R. NanoPack2: population-scale evaluation of long-read sequencing data. Bioinformatics. 2023;39(5):btad311.

70. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019;37(5):540–6.

71. Vaser R, Šikić M. Time- and memory-efficient genome assembly with Raven. Nature Computational Science. 2021;1(5):332–6.

72. Ruan J, Li H. Fast and accurate long-read assembly with wtdbg2. Nat Methods. 2020;17(2):155–8.

73. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34(18):3094–100.

74. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25(7):1043–55.

75. Jolley KA, Bliss CM, Bennett JS, Bratcher HB, Brehony C, Colles FM, et al. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. Microbiology. 2012;158(4):1005–15.

76. Yoon SH, Ha SM, Lim J, Kwon S, Chun J. A large-scale evaluation of algorithms to calculate average nucleotide identity. Antonie Van Leeuwenhoek. 2017;110(10):1281–6.

77. Alcock BP, Huynh W, Chalil R, Smith KW, Raphenya AR, Wlodarski MA, et al. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. Nucleic Acids Res. 2023;51(1 D):D690–9.

78. Kaas RS, Leekitcharoenphon P, Aarestrup FM, Lund O. Solving the problem of comparing whole bacterial genomes across different sequencing platforms. PLoS ONE. 2014;9(8):e104984.

79. Delcher AL, Salzberg SL, Phillippy AM. Using MUMmer to Identify Similar Regions in Large Sequence Sets. Curr Protoc Bioinformatics. 2003;00:10.3.1–10.3.18. https://doi.org/10.1002/0471250953.bi1003s00.

80. Vereecke N, Van Hoorde S, Sperling D, Theuns S, Devriendt B, Cox E. Virotyping and genetic antimicrobial susceptibility testing of porcine ETEC/STEC strains and associated plasmid types. Front Microbiol. 2023;14:1139312.

81. Letunic I, Bork P. Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49(W1):W293–6.

82. Schwengers O, Jelonek L, Dieckmann MA, Beyvers S, Blom J, Goesmann A. Bakta: Rapid and standardized annotation of bacterial genomes via alignment-free sequence identification. Microb Genom. 2021;7(11):000685.

83. Gilchrist CLM, Chooi YH. Clinker & clustermap.js: Automatic generation of gene cluster comparison figures. Bioinformatics. 2021;37(16):2473–5.

84. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015;31(22):3691–3.

85. Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. Genome Biol. 2016;17(1):238.

Publisher's Note