Identification of a family of *Vibrio* type III secretion system effectors that contain a conserved serine/threonine kinase domain

Plaza N[1], Garcia K[1], Waldor MK[2,3,4], Blondel CJ[5]*

[1] Instituto de Ciencias Biomédicas, Facultad de Ciencias de la Salud, Universidad Autónoma de Chile, Santiago, Chile

[2] Division of Infectious Diseases, Brigham & Women's Hospital, Boston, USA

[3] Department of Microbiology, Harvard Medical School, Boston, USA

[4] Howard Hughes Medical Institute, Boston, USA

[5] Instituto de Ciencias Biomédicas, Facultad de Medicina y Facultad de Ciencias de la Vida, Universidad Andrés Bello, 8320000, Santiago, Chile.

Key words: *Vibrio parahaemolyticus,* NleH*,* VPA1328, VopG, T3SS2, T3SS, Type III Secretion System, foodborne pathogen.

*Corresponding author. carlos.blondel@unab.cl

Mailing address, Instituto de Ciencias Biomédicas, Universidad Andrés Bello. Echaurren 183, Santiago, 8320000, Chile.

ABSTRACT

Vibrio parahaemolyticus is a marine Gram-negative bacterium that is a leading cause of seafood-borne gastroenteritis. Pandemic strains of *V. parahaemolyticus* rely on a specialized protein secretion machinery known as the type III secretion system 2 (T3SS2) to cause disease. The T3SS2 mediates the delivery of 10 known effector proteins into the cytosol of infected cells, where they subvert multiple cellular pathways. Here, we identify a new T3SS2 effector protein encoded by VPA1328 (VP_RS21530) in *V. parahaemolyticus* strain RIMD2210633. Bioinformatic analysis revealed that VPA1328 is part of a larger family of uncharacterized T3SS effector proteins with homology to the VopG effector protein in *V. cholerae* AM-19226. These VopG-like proteins are found in many but not all T3SS2 gene clusters and are distributed among diverse vibrio species including *V. parahaemolyticus*, *V. cholerae*, *V. mimicus*, and *V. diabolicus* and also in *Shewanella baltica*. Three clades of VopG sequences were identified by phylogenetic analysis, but these clades did not correspond to T3SS2 phylotypes, suggesting that *vopG* genes and T3SS2 clusters are evolving somewhat independently. Structure-based prediction analyses uncovered the presence of a conserved C-terminal kinase domain in VopG orthologs that is similar to the serine/threonine kinase domain found in the NleH family of T3SS effector proteins. The presence of this conserved kinase domain suggests that VopG effector proteins correspond to a new family of serine/threonine kinases included in the T3SS2 effector armamentarium.

IMPORTANCE

Vibrio parahaemolyticus is the leading bacterial cause of seafood-borne gastroenteritis worldwide. The pathogen relies on a type III secretion system to deliver a variety of effector proteins into the cytosol of infected cells to subvert cellular function. In this study, we identified a novel *Vibrio parahaemolyticus* effector protein that is similar to the VopG effector of *Vibrio cholerae.* VopG-like effectors were found in diverse vibrio species and contain a conserved serine/threonine kinase domain that bears similarity to the kinase domain in the EHEC *and* Shigella NleH effectors that manipulate the NF-κb signaling pathway. Together our findings identify a new family of vibrio effector proteins and highlight the role of horizontal gene transfer events among marine bacteria in shaping T3SS gene clusters.

INTRODUCTION

Vibrio parahaemolyticus is a marine Gram-negative bacterium that is the leading bacterial cause of seafood-borne gastroenteritis worldwide (1). In 1996, a new clonal *V. parahaemolyticus* strain of the O3:K6 serotype, now known as the pandemic clone, emerged and has been responsible for major outbreaks of gastroenteritis in diverse locations around the globe (2).

In addition to the presence of the characterized virulence factors thermostable direct hemolysin (TDH) and the *tdh*-related hemolysin (TRH), genome sequencing revealed that all *V. parahaemolyticus* strains encode a type III secretion system on chromosome 1 (T3SS1)(3). Furthermore, strains related to the pandemic clone harbor an evolutionarily distinct T3SS known as T3SS2 (4–6) encoded within an 80kb *Vibrio parahaemolyticus* pathogenicity island 7 (VPaI-7) on chromosome 2 (3). T3SSs are multicomponent nanomachines that enable Gram-negative bacteria to deliver proteins known as effectors directly from the bacterial cytosol into the cytosol of eukaryotic cells. Translocation of effectors into host cells enables pathogens to hijack host-cell signaling, thereby manipulating a variety of host cell functions (reviewed in (7). Indeed, the virulence of many human, animal, and plant pathogens depends on the activity of the T3SS injectisome and the repertoire of effector proteins delivered to their respective hosts' cells (8, 9).

Notably, most *V. parahaemolyticus* strains isolated from human clinical samples harbor T3SS2 and studies in animal models have shown that T3SS2 is essential for *V. parahaemolyticus* to colonize the intestine and to cause enteritis and diarrhea (10–12). Therefore, T3SS2 is considered a key *V. parahaemolyticus* virulence factor. Several T3SS2 related gene clusters have been identified in other *Vibrio* species and are referred to as T3SS2 phylotypes (T3SS2α, T3SS2β and T3SS2γ) (13). T3SS2α include T3SS2 gene clusters related to those found in the *tdh*-positive *V. parahae*molyticus

pandemic strain RIMD2210633 and in *V. cholerae* strain AM-19226. T3SS2β include T3SS2 gene clusters related to those found in *V. parahaemolyticus* strain TH3996 and *V. cholerae* strain 1587 (14). Finally, T3SS2γ include T3SS2 gene clusters related to those encoded in *V. parahaemolyticus* strain MAVP-Q, which has features found in the T3SS2α and T3SS2β gene clusters (15).

Ten T3SS2 effector proteins in *V. parahaemolyticus* have been identified to date (VopA, VopT, VopL, VopV, VopC, VopZ, VPA1380, VopO and the recently described "gatekeepers" VgpA and VgpB) (12, 16–22). These effectors subvert several cellular pathways including those controlling actin cytoskeleton dynamics and innate inflammatory responses (reviewed in (23–25)). These ten proteins are classified as either core or accessory T3SS2 effector proteins based on their distribution among the T3SS2 phylotypes (13). Notably, the presence of multiple uncharacterized genes in the VPaI-7 region raises the possibility that there are additional T3SS2 effector proteins yet to be identified.

In this study, we found that VPA1328, an ORF in the *V. parahaemolyticus* VPaI-7, encodes a novel T3SS2 effector protein. VPA1328, re-named here VopG, due to its similarity to the uncharacterized *V. cholerae* effector VopG, is secreted in a T3SS2-dependent fashion. Comparative genomic and phylogenetic analyses revealed that VPA1328 and VopG are members of a larger family of T3SS2 effector proteins encoded within the T3SS2 clusters of vibrios outside of *V. parahaemolyticus* and *V. cholerae* including in *V. mimicus* and *V. diabolicus* and the marine bacterium *Shewanella baltica*. The association of *vopG* genes with insertion sequence elements in several of these clusters suggests independent horizontal gene transfer or rearrangement events of these loci. Furthermore, VopG proteins have a conserved domain that exhibits sequence and predicted structural similarity to the serine/threonine kinase domain in the well-

characterized NleH family of T3SS effector proteins, suggesting a biochemical activity and potential biological functions for the VopG family of effectors.

RESULTS

VPA1328 is a VopG homolog that depends on *Vibrio parahaemolyticus* T3SS2 for its secretion.

We carried out BLASTp-based homology searches of the *V. parahaemolyticus* VPaI-7 genomic island as a way to identify candidate new T3SS2 effector proteins. This approach suggested that VPA1328 (VP_RS21530 in the latest genome annotation of strain RIMD2210633) is a putative T3SS2 effector protein (**Table S2**). VPA1328 is predicted to encode a 260 amino acid protein that shares ~42% amino acid sequence identity with the T3SS effector protein VopG, encoded in the phylogenetically related T3SS2 in *V. cholerae* AM-19226 (26) (**Fig. 1A and 1B**). The function of VopG remains unknown, but it is secreted and translocated by the *V. cholerae* T3SS2 and contributes to host cell cytotoxicity and colonization in a mouse model of infection (26), suggesting an important role for this effector in virulence (26). Even though VPA1328 and VopG are located in different locations within their respective T3SS2 clusters, their sequence similarity and presence in phylogenetically related T3SSs suggests that VPA1328 is a VopG homolog that functions as a *V. parahaemolyticus* T3SS2 effector protein. Below we refer to VPA1328 as VopG.

Next, we tested if VopG (VPA1328) is secreted and whether its secretion requires the *V. parahaemolyticus* T3SS2. In these experiments, *V. parahaemolyticus* WT strain RIMD2210633 and isogenic T3SS1, T3SS2, and T3SS1/T3SS2-deficient mutant strains (Δ*vscn1,* Δ*vscn2* and Δ*vscn1*Δ*vscn2*, respectively) were grown under conditions (27) that induce expression of T3SS2 (LB 0.04% bile) (27). To detect VopG, these

strains were transformed with pVPA1328-CyA, a plasmid that harbors a translational fusion between the VPA1328 ORF (VopG) and the adenylate cyclase domain (CyA) of plasmid pCyA. This construct enables immunoblot detection of VopG in cell lysates and culture supernatants using anti-CyA antibodies. A VopV-CyA fusion (pVopV-CyA) was included as a positive control for T3SS2-dependent secretion (19).

A band corresponding to the predicted size of the VopG-Cya fusion (~74kDa, along with some lower molecular weight species likely corresponding to degradation products) was observed in cell lysates from the WT strain harboring pVPA1328-CyA but not a control strain harboring the empty vector pCyA (**Fig. 2A**). VopG was only detected in supernatants when the WT (pVPA1328-CyA) strain was grown under T3SS2 inducing (LB 0.04% bile), suggesting that its secretion requires T3SS2 activity. Interestingly, previous transcriptomic analysis showed that expression of VPA1328 was increased by the presence of bile and controlled by VtrB, the master regulator of T3SS2 expression, suggesting that it is part of the VtrB regulon (28).

Analyses of VopG secretion from Δ*vscn1* (T3SS1-deficient) and Δ*vscn2* (T3SS2-deficient) strains strongly support the idea that VopG secretion requires T3SS2 and not T3SS1. When secretion by T3SS1 or T3SS2 or both T3SS was disabled by deletion of their respective ATPases, there was similar expression of VopG in cell lysates (**Fig. 2B**); however, VopG was only detected in supernatants from the strain where T3SS1 was inactivated but not when T3SS2 was inactive. An identical pattern was observed with VopV, a known T3SS2 substrate (**Fig. 2B**). The cytosolic RNA polymerase beta subunit (RNAP) was not detected in any of the culture supernatant samples, indicating that detection of VPA1328 in culture supernatants was not a consequence of bacterial lysis. Together, these observations demonstrate that VopG is secreted in a T3SS2-dependent fashion and given its similarity to the *V. cholerae* VopG effector, strongly support the notion that VopG is a novel *V. parahaemolyticus* T3SS2 effector protein.

VopG homologs are widely distributed in vibrios harboring T3SS2 clusters.

The presence of a VopG homolog encoded within the T3SS2 gene cluster in *V. parahaemolyticus* RIMD2210633 prompted us to investigate if additional VopG homologs are present among distinct T3SS2 phylotypes. The VPA1328 sequence was used as a query to identify potential VopG homologs by sequential BLASTn, BLASTp and tBLASTx searches, using publicly available bacterial genome sequences. With cut-off values of 60% sequence coverage and 40% sequence identity, 2044 candidate VopG homologs were identified, including 122 non-redundant protein sequences (**Fig. 3A, 3B and Table S3**). The majority of the VopG homologs (86%, n=1764) were encoded in *V. parahaemolyticus* strains and in *V. cholerae* strains (12.5%, n=256), but homologs were also identified in *V. mimicus* (0.2%, n=5), *V. diabolicus* (0.04%, n=1), *Vibrio sp* (0.5%, n=11) and in *Shewanella* strains (0.3%, n=7); i.e., in most species known to harbor T3SS2 gene clusters. Interestingly, a T3SS2 gene cluster was not previously identified in *V. diabolicus*, a marine organism. However, it is important to note that not all vibrio species, e.g., *Vibrio anguillarum* (29) which harbor T3SS2 gene clusters, encode VopG homologs. Thus, even though VopG is widely distributed, this putative effector protein is not a universal component of the T3SS2.

Next, we evaluated the sequence relatedness of VopG homologs using phylogenetic analysis of the 122 non-redundant VopG sequences. Notably, three distinct clades (A, B and C) of VopG proteins were identified **(Fig. 3C)**. No correlation was found between these clades and T3SS2 phylotypes. For example, the VopG homologs of *V. cholerae* strain AM-19226 and *V. parahaemolyticus* RIMD2210633 (VPA1328) clustered in different clades (B and A, respectively) despite the fact that both these T3SS2 belong to the T3SS2α phylotype. The lack of correlation between the VopG clades and T3SS2

phylotypes suggests that VopG effectors have to some extent evolved independently of the T3SS2 machinery that delivers them to host cells.

Comparative genomic analyses were carried out to gain insights into variation of the genomic contexts of *vopG* genes within different T3SS2 gene clusters. Genome sequences from representatives of each clade of the VopG phylogenetic tree, including at least one genome for each different *Vibrio* species were used for these comparisons. As shown in **Fig. 4**, the overall genetic structure of these T3SS2 gene clusters is highly conserved, particularly in the regions encoding structural components of the T3SS2 apparatus. In most T3SS2 gene clusters the relative position of *vopG* was similar with the exception of *V. parahaemolyticus* RIMD2210633. However, the nucleotide sequences and ORFs that are adjacent to the *vopG* homologs differed in most of the 7 clusters analyzed in Fig 4. In several of these cases, *vopG* was found to be close to sequences with similarity to insertion sequence (IS) elements. This association raises the possibility that IS can promote the mobility of *vopG* loci and potentially account for the variations in the genetic contexts of these loci within different T3SS2 gene clusters.

Consistent with this idea, we identified two *vopG* homologs (FORC14_RS05860 and FORC14_RS06170) encoded in the T3SS2 gene cluster of *V. parahaemolyticus* strain FORC014. Analysis of their respective genetic contexts revealed that one of these *vopG* genes (FORC14_RS05860) is located at the end of the T3SS gene cluster and is flanked by IS200-like mobile genetic elements (**Fig. S1A)**. These elements have high sequence identity to the ISVpa3 insertion sequence. ISVpa3 is an insertion sequence located adjacent to each copy of the TDH gene in *V. parahaemolyticus* strain RIMD2210633 and linked in some strains to deletion of TDH (30). While the T3SS2 gene cluster of *V. parahaemolyticus* RIMD2210633 has 3 copies of these ISVpa3 elements, *V. parahaemolyticus* strain FORC014 has 6 of these elements, 2 of them flanking one of the *vopG* homologs at the end of the cluster (**Fig. S1B**). Sequence analysis showed that

the 2 *vopG* homologs in strain FORC014 share 69% nucleotide identity (**Fig. S2**). Both the sequence divergence of these 2 *vopG* genes and the mobile genetic elements flanking FORC14_RS05860 suggest that this *vopG* homolog was independently acquired, potentially via a horizontal gene transfer event, and not a duplication of FORC14_RS06170.

While the presence of a T3SS2 gene cluster in *Shewanella baltica* species has been inferred due to the presence of the *vscn2* gene in strain *Shewanella baltica* BA175 and *Shewanella baltica* OS183 (31), information regarding the distribution and genetic context of the T3SS2 gene cluster in this genus has not been reported. We found that the *Shewanella* T3SS2 is located within a genomic island inserted between the SBAL678_RS45345 and SBAL678_RS45350 ORFs of reference strain OS678 (**Fig. S1B**). This genomic island includes 45 ORFs. The majority of these genes encode structural components of the T3SS2 apparatus. Interestingly, not every T3SS2 gene cluster identified in *Shewanella* harbors a VopG-encoding gene (**Fig. S1B**).

VopG proteins have sequence and predicted structural similarity to the NleH family of serine/threonine kinases

The amino acid sequence conservation of the 122 VopG homologs was analyzed and depicted using WebLogo. The analysis showed a particularly striking conservation in the C-termini of these amino acid sequences (**Fig. S3**), suggesting that this region of VopG includes a functional domain. To gain clues regarding the function of VopG proteins, we used the structure-based homology tools HHpred (32) and pGenTHREADER (33). For these analyses, the amino acid sequence of *V. parahaemolyticus* VPA1328 was used as a representative of the VopG family of effectors. Both of these algorithms detected a region in the VopG C-terminus with similarity to NleH effectors, e.g., HHpred analysis uncovered the presence of a region of

22 amino acids in VPA1328 (position 190-212) with identity to the T3SS effector proteins NleH1 from *Escherichia coli* O157:H7 strain Sakai (PDB: 4LRJ_B) and OspG from *Shigella flexneri* strain 301 (PDB: 4Q5E_A).

NleH1 and OspG are members of the NleH of family of T3SS effectors (34–36). These proteins are translocated by the T3SSs of different bacterial species and act as serine/threonine kinases in host cells (35). There are 6 members of this family of effector proteins, including the NleH1 and NleH2 proteins of *E. coli* O157:H7 str. Sakai, OspG from *Shigella flexneri* strain 301, NleH of *Citrobacter rodentium* strain DBS100, SboH of *Salmonella bongori* NCTC 12419 and YspK of *Yersinia enterocolitica* strain 8081 (37–42). The serine/threonine kinase domain of these effectors is distantly related to eukaryotic regulatory kinases; moreover, functional studies have suggested that each of these effectors perturbs the NF-κb pathway and inhibits apoptosis in infected host cells (37, 42, 43).

Multiple sequence alignment of representatives of the NleH and VopG family of T3SS effector proteins were carried out to gain further insight into their similarity. Representatives from each clade of VopG proteins were included in these analyses. The analysis showed that the greatest similarity between VopG and NleH proteins is found in their C-termini in the region that includes the characterized NleH serine/threonine kinase domain; in contrast, their N-termini differ in both length and sequence (**Fig. 5A and Fig S4**). VopG proteins contain all the critical amino acid residues and motifs important for kinase activity, including the conserved catalytic residues glycine of the G-rich loop, the aspartic acid (D) and the asparagine (N) of the catalytic loop (alignment position 200–205 in VPA1328) and the PID motif of the activation loop (alignment position 220–233 in VPA1328). In addition, VopG proteins also share the invariant lysine (alignment position 109) involved in the autophosphorylation of the NleH family, and which has been used as a proxy to measure kinase activity (43) (**Fig. 5A and 5B**). Thus, VopG family effector

proteins harbor a NleH-like C-terminal serine/threonine kinase domain. Phylogenetic analysis of bacterial serine/threonine kinase domains also revealed the similarity of the kinase domains of VopG and NleH proteins (**Fig. 6A**). The VopG proteins clustered closer to the NleH proteins on this tree than to non-NleH serine threonine kinases from *Legionella pneumophila, Yersinia pestis*, and *Salmonella enterica*.

We derived a 3D structural model of the C-terminal domain of VPA1328 using comparative homology modeling with I-TASSER (44) to gain further insight into the serine/threonine kinase domain of VopG proteins. In accord with the HHPred and pGenTHREADER analyses, I-TASSER identified NleH1, NleH2 and OspG as suitable models for comparative homology models using the crystal structures available for these proteins. Five models were obtained, and model 1 was chosen based on its error estimation, TM-score and RMSD values (**Fig. S5**). As shown in **Fig. 6B**, this model revealed the remarkable similarity of the predicted structure of the VPA1328 kinase domain with the NleH kinase domain. The structure of the catalytic pocket, including the positions of the predicted catalytic amino acid side chains (K109, D201 and N205) in VPA1328 and OspG structure overlap **(Fig. 6B)**, strongly supporting the notion that the VopG family of proteins encode a NleH-like serine/threonine kinase domain.

DISCUSSION

While all *Vibrio parahaemolyticus* strains harbor a T3SS1, a hallmark of the pandemic *Vibrio parahaemolyticus* O3:K6 clone and most human clinical *Vibrio parahaemolyticus* isolates is the presence of a second and phylogenetically distinct T3SS2. The latter T3SS is essential for both intestinal colonization and virulence in some animal models of disease (10, 12). Here, we found that a T3SS2 ORF (VPA1328) likely corresponds to a novel *V. parahaemolyticus* T3SS2 effector protein. This ORF, which is secreted in a T3SS2-dependent fashion, bears similarity to the VopG effector found in the *V. cholerae* AM-19226 T3SS2. The function of the latter VopG protein is unknown, but it has been shown to be translocated to host cells and linked to *V. cholerae* AM-19226's pathogenicity. Bioinformatic analyses uncovered 122 non-redundant VopG-like proteins encoded in all 3 phylotypes of T3SS2 clusters in diverse vibrio species and in *S. baltica*. Interestingly, the evolutionary history of the T3SS2 phylotypes does not appear to correspond with the evolution of the 3 clades of VopG proteins that were uncovered by phylogenetic analysis. We found that the highly conserved C-terminal domains of VopG proteins bear striking structural similarity to the serine/threonine kinase domain of the NleH family of effectors found in enteric pathogens such as EHEC and Shigella (OspG). Thus, our findings support the idea that VopG effectors function as serine/threonine kinases in host cells.

The *V. cholerae* AM-19226 effector VopG had been classified as a *V. cholerae* specific T3SS effector (26, 45) but our analyses showed that VopG homologs belong to a larger family of putative effector proteins that is widely distributed among *Vibrio* species including *V. parahaemolyticus*, *V. cholerae*, *V. mimicus* and *V. diabolicus* as well as in strains of *Shewanella baltica*. Recently, Matsuda et al proposed to classify T3SS2 effectors proteins as "core" effectors if they are conserved in both *V. parahaemolyticus* and non-O1/non-O139 *V. cholerae* (13) and as "accessory" effectors (13) if they are not.

According to this classification, our work suggests that VopG corresponds to a core effector protein due to its presence in multiple vibrio species. However, VopG homologs are not present in the T3SS2 gene clusters identified in all vibrio species; e.g, the T3SS2 cluster in *Vibrio anguillarum* (29) lack a VopG homolog and not all clusters in–*Vibrio mimicus* (46) encode a recognizable VopG.

T3SS2 gene clusters are classified into three phylotypes (T3SS2α, T3SS2β and T3SS2γ) that are believed to be have been acquired through horizontal gene transfer events (13, 15, 46). Even though VopG homologs are not universally found in all T3SS2 gene clusters, we identified VopG homologs in all three T3SS2 phylotypes. Phylogenetic analysis identified three distinct VopG clades (**Fig. 3C**). These VopG clades did not correlate with T3SS2 phylotypes; i. e., all three clades were found in each T3SS2 phylotype. The apparent independent evolution of T3SS2 phylotypes and VopG clades supports the possibility that *vopG* genes have been independently acquired by different T3SS2 lineages.

The absence of *vopG* genes from certain T3SS2 clusters could be explained either by loss of *vopG* loci due to deletion event(s) or independent acquisition of *vopG* in some T3SS2 clusters. The presence of a second *vopG* homolog flanked by IS elements in *V. parahaemolyticus* strain FORC014 suggests insertion sequences may play a role in mobilizing *vopG* genes. These sequences bear similarity to the ISVpa3 insertion sequence first described in *V. parahaemolyticus* RIMD2210633 (30). Since insertion sequences have been shown to shape bacterial genomic islands through rearrangements, insertion and deletion events, it is plausible that ISVpa3-like elements have shaped the evolution of T3SS2 gene clusters through similar mechanisms. The apparent mobility of *vopG* loci adds an additional layer of complexity to our understanding of T3SS2 clusters. That is, these clusters appear to have been spread via horizontal gene transfer events among marine bacteria and their repertoire of effector

proteins appears to be “tunable” through independent horizontal gene transfer or rearrangement events.

Analysis of the amino acid sequences of the 122 non-redundant VopG homologs identified here revealed a particularly high degree of conservation in their C-termini. This region of VopG proteins was found to be very similar to the conserved serine/threonine kinase domain in the NleH family of T3SS effector proteins. Thus, the conservation of this part of VopG effectors is likely explained by the presence of functional kinase domain. Structural predictions, which showed that VopG proteins contain all the residues that constitute that catalytic pocket of NleH proteins, strongly support this hypothesis.

The NleH family of effectors contain a eukaryotic-like serine/threonine kinase domain that independently evolved in bacteria (34, 35). VopG homologs harbor each of the key residues described in the NleH family of protein kinases. The classification of the NleH proteins as a distinct bacterial kinase family was made through structure-based phylogenetic analysis (34). Phylogenetic analysis showed that the C-termini of VopG proteins have more similarity to the kinase domain of NleH proteins than to other bacterial protein kinases (**Fig. 6A**), but further structural information is required to determine if VopG proteins are novel members of the NleH family or represent a distinct family on their own.

Although the conservation of key catalytic residues in VopG and NleH proteins provides evidence that supports the notion that VopG proteins are functional serine/threonine kinases, it is more problematic to speculate that their biological function is conserved as well. To date, NleH proteins have been linked to their ability to perturb the NF-κb pathway and impact cellular survival during infection through different molecular mechanisms (37, 38, 42, 43, 47). Both NleH1 and NleH2 proteins bind the host protein RPS3 leading to inhibition or activation of the NF-κb pathway respectively (42, 48). Despite these apparently different effects on the NF-κb pathway, both were

shown to inhibit apoptosis through interaction with the Bax Inhibitor 1 protein (43). The *Shigella* OspG protein can inhibit the NF-κb pathway by inhibiting the proteasomal destruction of IκBα (40) and the SboH protein of *Salmonella bongori* blocks intrinsic apoptotic pathways (37). It will be interesting to test whether VopG proteins also perturb the NF-κb pathway. Furthermore, the N-terminal region of NleH proteins has been linked to substrate recognition and the observed functional differences between the NleH1 and NleH2 proteins (34, 38, 42). In this context, the sequence divergence observed within the N-terminal region of VopG homologs (**Fig. S3 and S4**) may have functional implications.

In summary, our work identifies a new family of VopG proteins that are likely T3SS2 effectors. These proteins contain a distinctive NleH-like serine/threonine kinase domain. Future biochemical and structural studies are required to corroborate these predictions. Moreover, defining the role(s) of these effectors in the pathogenicity and/or environmental adaptation of the diverse Vibrio and Shewanella species that encode them will be fruitful.

MATERIALS AND METHODS

Bacterial strains and growth conditions.

All bacterial strains and plasmids used in this study are listed in **Table S1**. *V. parahaemolyticus* RIMD2210633 (3) and its *ΔvscN1*, *ΔvscN2* and *ΔvscN1 ΔvscN2* derivatives (19) were used in this study. Bacterial strains were routinely cultured in LB medium or on LB agar plates at 37°C. Culture media was supplemented with the following antibiotics and chemicals: 0.04% bovine and ovine bile (Sigma Cat No. B8381); 100µg/ml ampicillin, 5µg/ml and 20µg/ml chloramphenicol for *V. parahaemolyticus* and *E. coli* strains, respectively; 1ug/ml IPTG to induce expression vector pCYA.

T3SS secretion assays

To determine if VPA1328 was secreted in a T3SS2-dependent fashion, a reporter fusion was constructed between VPA1328 and the CyA reporter encoded in plasmid pCyA (a pMMB207 derivative) **(49)** generating plasmid pVPA1328-CyA. As a positive control for T3SS2-dependent secretion, a VopV-CyA fusion was used (pVopV-CyA) **(19)**. As a negative control, the empty plasmid pCyA was used. Each plasmid was transformed into V. parahaemolyticus strains by electroporation as previously described **(19)**. Secretion assays were performed by growing each V. parahaemolyticus strain for 1.5 h in LB medium supplemented with 0.04% bile. When cultures reached an OD600nm of 0.6, 1mM IPTG was added to induce expression of the CyA reporter constructs. After 1.5 h, culture supernatants were collected by centrifugation at 6000 RPM for 20 min, filtered sterilized through a 0.22-µm filter and concentrated 50-70-fold by filtration with an Amicon Ultra-15 Centrifugal Filter Unit (Millipore) with a 10 kDa molecular weight cut-off. Prior to concentrating the culture supernatant, BSA (1 mg/ml) was added to serve as a concentration/loading control. Whole-cell lysates were prepared by solubilizing the

bacterial pellet in 1X Laemmli buffer. Lysate and supernatant samples were processed for SDS-PAGE analysis by mixing them with loading buffer, boiling for 5 min, and run on 10% SDS-PAGE gels. For immunoblot analysis, gels were transferred to PVDF membranes. The Pierce Coomassie Plus (Bradford) Assay Kit (Thermo Fisher Cat No. 23236) was used for determination of protein concentrations. Antibodies were used at the following dilutions: anti-CyA (rabbit polyclonal, 1:2,000), anti-RNA polymerase (mouse monoclonal, 1:2,000; Santa Cruz Biotechnology Cat. No. Sc-101597), goat anti-mouse IgG-HRP (1:10,000; Santa Cruz Biotechnology Cat No. sc-2031) and goat anti-Rabbit IgG (H+L) Secondary Antibody, HRP (1:10,000; Thermo Cat No.31460). The blots were developed with SuperSignal West Pico ECL substrate (Thermo Fisher Cat No. 35060) and imaging was performed on a C-DiGit Blot Scanner (LI-COR Biosciences). All blots are representative of at least 3 biological replicates.

Sequence and phylogenetic analysis

Identification of VopG orthologs was carried out using the VPA1328 amino acid and nucleotide sequences as queries in BLASTp, BLASTn, BLASTx, tBLASTn and tBLASTx analyses using publicly available bacterial genome sequences of the NCBI database (December 2020). A 94% sequence length, 40% identity and 60% sequence coverage threshold were used to select positive matches. Sequence conservation was analyzed by multiple sequence alignments using MAFFT (50) and T-Coffee Expresso (51) and visualized by ESPript 3 (52). WebLogo analysis was performed using multiple sequence alignments (53). Comparative genomic analysis of the T3SS2 gene clusters was performed using the multiple aligner Mauve (54) and the IslandViewer 4 pipeline (55) and EasyFig v2.2.2. Nucleotide sequences were analyzed by the sequence visualization and annotation tool Artemis version 18.1 (56). Multiple sequence alignments were used for phylogenetic analyses that were performed with the Molecular

Evolutionary Genetics Analysis (MEGA) software version 7.0 (57) and visualized by iTOL (58). Phylogenetic trees were built from the alignments by the bootstrap test of phylogeny (2000 replications) using the neighbor-joining (NJ) method with a Jones-Taylor-Thornton (JTT) correction model.

Remote homology prediction and homology modeling

Remote homology prediction of VPA1328 was performed using HHpred (32) and pGENTHREADER (33) on the PSIPRED server (59). Protein structure models of the VPA1328 C-terminal domain were obtained using I-TASSER (44), a protein structure homology-modelling server. Protein structure visualization and template alignment and superposition were performed using MAESTRO (60).

ACKNOWLEDGMENTS

We thank members of the Blondel Laboratory for helpful discussions on all aspects of this project and for their comments on the manuscript.

AUTHOR CONTRIBUTIONS

Nicolas Plaza: Conceptualization, Methodology, Formal analysis, Investigation, Visualization, Data curation, Writing-Original Draft and review and editing.

Katherine Garcia: Conceptualization, review and editing final manuscript.

Matthew K. Waldor: Conceptualization, Resources, Writing - Review and editing.

Carlos J Blondel: Conceptualization, Methodology, Formal analysis, Investigation, Visualization, Resources, Data curation, Writing - Review and editing, Funding acquisition, Supervision and Project administration.

CONFLICT OF INTERESTS

The authors declare that there are no conflicts of interest.

FUNDING INFORMATION

This work was funded by the Howard Hughes Medical Institute (HHMI)-Gulbenkian International Research Scholar Grant #55008749, FONDECYT Grant 1201805 (ANID) and REDI170269 (ANID) to CJB. KG is supported by FONDECYT Grant 1190957

462 (ANID). MKW is a Howard Hughes Medical Institute (HHMI) Investigator and is
463 supported by NIAID grant R01-AI-043247.

464 **REFERENCES**

465

466 1. Letchumanan V, Chan K-G, Lee L-H. 2014. Vibrio parahaemolyticus: a review on the
467 pathogenesis, prevalence, and advance molecular identification techniques. Frontiers in
468 microbiology 5:705.

469 2. Velazquez-Roman J, León-Sicairos N, Hernández-Díaz L de J, Canizalez-Roman A.
470 2013. Pandemic Vibrio parahaemolyticus O3:K6 on the American continent. Frontiers in
471 cellular and infection microbiology 3:110.

472 3. Makino K, Oshima K, Kurokawa K, Yokoyama K, Uda T, Tagomori K, Iijima Y, Najima
473 M, Nakano M, Yamashita A, Kubota Y, Kimura S, Yasunaga T, Honda T, Shinagawa H,
474 Hattori M, Iida T. 2003. Genome sequence of Vibrio parahaemolyticus: a pathogenic
475 mechanism distinct from that of V cholerae. Lancet 361:743–749.

476 4. Hazen TH, Lafon PC, Garrett NM, Lowe TM, Silberger DJ, Rowe LA, Frace M,
477 Parsons MB, Bopp CA, Rasko DA, Sobecky PA. 2015. Insights into the environmental
478 reservoir of pathogenic Vibrio parahaemolyticus using comparative genomics. Frontiers
479 in microbiology 6:204.

480 5. Abby SS, Rocha EPC. 2012. The non-flagellar type III secretion system evolved from
481 the bacterial flagellum and diversified into host-cell adapted systems. PLoS genetics
482 8:e1002983.

483 6. Park K-S, Ono T, Rokuda M, Jang M-H, Okada K, Iida T, Honda T. 2004. Functional
484 characterization of two type III secretion systems of Vibrio parahaemolyticus. Infection
485 and immunity 72:6659–6665.

486 7. Portaliou AG, Tsolis KC, Loos MS, Zorzini V, Economou A. 2016. Type III Secretion:
487 Building and Operating a Remarkable Nanomachine. Trends in biochemical sciences
488 41:175–189.

489 8. Lara-Tejero M, Galán JE. 2019. Protein Secretion in Bacteria. Ecosal Plus 8:245–259.

490 9. Galán JE, Lara-Tejero M, Marlovits TC, Wagner S. 2014. Bacterial Type III Secretion
491 Systems: Specialized Nanomachines for Protein Delivery into Target Cells. Annu Rev
492 Microbiol 68:1–24.

493 10. Hubbard TP, Chao MC, Abel S, Blondel CJ, Wiesch PAZ, Zhou X, Davis BM, Waldor
494 MK. 2016. Genetic analysis of Vibrio parahaemolyticus intestinal colonization. Proc
495 National Acad Sci 113:201601718.

496 11. Piñeyro P, Zhou X, Orfe LH, Friel PJ, Lahmers K, Call DR. 2010. Development of
497 two animal models to study the function of Vibrio parahaemolyticus type III secretion
498 systems. Infection and immunity 78:4551–4559.

12. Ritchie JM, Rui H, Zhou X, Iida T, Kodoma T, Ito S, Davis BM, Bronson RT, Waldor MK. 2012. Inflammation and disintegration of intestinal villi in an experimental model for Vibrio parahaemolyticus-induced diarrhea. PLoS pathogens 8:e1002593.

13. Matsuda S, Hiyoshi H, Tandhavanant S, Kodama T. 2020. Advances on Vibrio parahaemolyticus research in the postgenomic era. Microbiol Immunol 64:167–181.

14. Okada N, Iida T, Park K-S, Goto N, Yasunaga T, Hiyoshi H, Matsuda S, Kodama T, Honda T. 2009. 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. Infection and immunity 77:904–913.

15. Xu F, Gonzalez-Escalona N, Drees KP, Sebra RP, Cooper VS, Jones SH, Whistler CA. 2017. Parallel Evolution of Two Clades of an Atlantic-Endemic Pathogenic Lineage of Vibrio parahaemolyticus by Independent Acquisition of Related Pathogenicity Islands. Appl Environ Microb 83:e01168-17.

16. Hiyoshi H, Okada R, Matsuda S, Gotoh K, Akeda Y, Iida T, Kodama T. 2015. Interaction between the Type III Effector VopO and GEF-H1 Activates the RhoA-ROCK Pathway. Plos Pathog 11:e1004694.

17. Hiyoshi H, Kodama T, Saito K, Gotoh K, Matsuda S, Akeda Y, Honda T, Iida T. 2011. VopV, an F-Actin-Binding Type III Secretion Effector, Is Required for Vibrio parahaemolyticus-Induced Enterotoxicity. Cell Host Microbe 10:401–409.

18. Kodama T, Rokuda M, Park K-S, Cantarelli VV, Matsuda S, Iida T, Honda T. 2007. Identification and characterization of VopT, a novel ADP-ribosyltransferase effector protein secreted via the Vibrio parahaemolyticus type III secretion system 2. Cellular Microbiology 9:2598–2609.

19. Zhou X, Gewurz BE, Ritchie JM, Takasaki K, Greenfeld H, Kieff E, Davis BM, Waldor MK. 2013. A Vibrio parahaemolyticus T3SS effector mediates pathogenesis by independently enabling intestinal colonization and inhibiting TAK1 activation. Cell reports 3:1690–1702.

20. Zhang L, Krachler AM, Broberg CA, Li Y, Mirzaei H, Gilpin CJ, Orth K. 2012. Type III effector VopC mediates invasion for Vibrio species. Cell reports 1:453–460.

21. Trosky JE, Mukherjee S, Burdette DL, Roberts M, McCarter L, Siegel RM, Orth K. 2004. Inhibition of MAPK Signaling Pathways by VopA from Vibrio parahaemolyticus. J Biol Chem 279:51953–51957.

22. Tandhavanant S, Matsuda S, Hiyoshi H, Iida T, Kodama T. 2018. Vibrio parahaemolyticus Senses Intracellular K+ To Translocate Type III Secretion System 2 Effectors Effectively. Mbio 9:e01366-18.

23. Matsuda S, Hiyoshi H, Tandhavanant S, Kodama T. 2020. Advances on Vibrio parahaemolyticus research in the postgenomic era. Microbiol Immunol 64:167–181.

536 24. Lingzhi L, Meng H, Dan G, Yang L, Mengdie J. 2019. Molecular mechanisms of
537 Vibrio parahaemolyticus pathogenesis. Microbiol Res 222:43–51.

538 25. O'Boyle N, Boyd A. 2014. Manipulation of intestinal epithelial cell function by the cell
539 contact-dependent type III secretion systems of Vibrio parahaemolyticus. Frontiers in
540 cellular and infection microbiology 3:114.

541 26. Chaand M, Miller KA, Sofia MK, Schlesener C, Weaver JWA, Sood V, Dziejman M.
542 2015. Type Three Secretion System Island-Encoded Proteins Required for Colonization
543 by Non-O1/Non-O139 Serogroup Vibrio cholerae. Infect Immun 83:2862–2869.

544 27. Gotoh K, Kodama T, Hiyoshi H, Izutsu K, Park K-S, Dryselius R, Akeda Y, Honda T,
545 Iida T. 2010. Bile acid-induced virulence gene expression of Vibrio parahaemolyticus
546 reveals a novel therapeutic potential for bile acid sequestrants. PloS one 5:e13365.

547 28. Livny J, Zhou X, Mandlik A, Hubbard T, Davis BM, Waldor MK. 2014. Comparative
548 RNA-Seq based dissection of the regulatory networks and environmental stimuli
549 underlying Vibrio parahaemolyticus gene expression during infection. Nucleic Acids Res
550 42:12212–12223.

551 29. Naka H, Dias GM, Thompson CC, Dubay C, Thompson FL, Crosa JH. 2011.
552 Complete genome sequence of the marine fish pathogen Vibrio anguillarum harboring
553 the pJM1 virulence plasmid and genomic comparison with other virulent strains of V.
554 anguillarum and V. ordalii. Infection and immunity 79:2889–2900.

555 30. Kamruzzaman M, Bhoopong P, Vuddhakul V, Nishibuchi M. 2008. Detection of a
556 functional insertion sequence responsible for deletion of the thermostable direct
557 hemolysin gene (tdh) in Vibrio parahaemolyticus. Gene 421:67–73.

558 31. Dong X, Li N, Liu Z, Lv X, Shen Y, Li J, Du G, Wang M, Liu L. 2020. CRISPRi-
559 Guided Multiplexed Fine-Tuning of Metabolic Flux for Enhanced Lacto- N -neotetraose
560 Production in Bacillus subtilis. J Agr Food Chem 68:2477–2484.

561 32. Zimmermann L, Stephens A, Nam S-Z, Rau D, Kübler J, Lozajic M, Gabler F, Söding
562 J, Lupas AN, Alva V. 2017. A Completely Reimplemented MPI Bioinformatics Toolkit
563 with a New HHpred Server at its Core. J Mol Biol 430:2237–2243.

564 33. Lobley A, Sadowski MI, Jones DT. 2009. pGenTHREADER and pDomTHREADER:
565 new methods for improved protein fold recognition and superfamily discrimination.
566 Bioinformatics 25:1761–1767.

567 34. Grishin AM, Cherney M, Anderson DH, Phanse S, Babu M, Cygler M. 2014. NleH
568 Defines a New Family of Bacterial Effector Kinases. Structure 22:250–259.

569 35. Grishin AM, Beyrakhova KA, Cygler M. 2015. Structural insight into effector proteins
570 of Gram-negative bacterial pathogens that modulate the phosphoproteome of their host.
571 Protein Sci 24:604–620.

572 36. Grishin A, Cherney M, Condos T, Barber K, Anderson D, Phanse S, Babu M, Shaw
573 G, Cygler M. 2014. Bacterial Effector Kinases. Acta Crystallogr Sect Found Adv
574 70:C428–C428.

575 37. Fookes M, Schroeder GN, Langridge GC, Blondel CJ, Mammina C, Connor TR,
576 Seth-Smith H, Vernikos GS, Robinson KS, Sanders M, Petty NK, Kingsley RA, Bäumler
577 AJ, Nuccio S-P, Contreras I, Santiviago CA, Maskell D, Barrow P, Humphrey T, Nastasi
578 A, Roberts M, Frankel G, Parkhill J, Dougan G, Thomson NR. 2011. Salmonella bongori
579 provides insights into the evolution of the Salmonellae. PLoS pathogens 7:e1002191.

580 38. Pham TH, Gao X, Tsai K, Olsen R, Wan F, Hardwidge PR. 2012. Functional
581 Differences and Interactions between the Escherichia coli Type III Secretion System
582 Effectors NleH1 and NleH2. Infect Immun 80:2133–2140.

583 39. García-Angulo VA, Deng W, Thomas NA, Finlay BB, Puente JL. 2008. Regulation of
584 Expression and Secretion of NleH, a New Non-Locus of Enterocyte Effacement-
585 Encoded Effector in Citrobacter rodentium□. J Bacteriol 190:2388–2399.

586 40. Kim DW, Lenzen G, Page A-L, Legrain P, Sansonetti PJ, Parsot C. 2005. The
587 Shigella flexneri effector OspG interferes with innate immune responses by targeting
588 ubiquitin-conjugating enzymes. P Natl Acad Sci Usa 102:14046–14051.

589 41. Matsumoto H, Young GM. 2006. Proteomic and functional analysis of the suite of
590 Ysp proteins exported by the Ysa type III secretion system of Yersinia enterocolitica
591 Biovar 1B. Mol Microbiol 59:689–706.

592 42. Gao X, Wan F, Mateo K, Callegari E, Wang D, Deng W, Puente J, Li F, Chaussee
593 MS, Finlay BB, Lenardo MJ, Hardwidge PR. 2009. Bacterial Effector Binding to
594 Ribosomal Protein S3 Subverts NF-κB Function. Plos Pathog 5:e1000708.

595 43. Hemrajani C, Berger CN, Robinson KS, Marchès O, Mousnier A, Frankel G. 2010.
596 NleH effectors interact with Bax inhibitor-1 to block apoptosis during enteropathogenic
597 Escherichia coli infection. Proc National Acad Sci 107:3129–3134.

598 44. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. 2015. The I-TASSER Suite:
599 protein structure and function prediction. Nat Methods 12:7–8.

600 45. Alam A, Miller KA, Chaand M, Butler JS, Dziejman M. 2011. Identification of Vibrio
601 cholerae Type III Secretion System Effector Proteins †. Infect Immun 79:1728–1740.

602 46. Okada N, Matsuda S, Matsuyama J, Park K-S, Reyes C de los, Kogure K, Honda T,
603 Iida T. 2010. Presence of genes for type III secretion system 2 in Vibrio mimicus strains.
604 BMC microbiology 10:302.

605 47. Grishin AM, Barber KR, Gu R-X, Tieleman DP, Shaw GS, Cygler M. 2018.
606 Regulation of Shigella Effector Kinase OspG through Modulation of its Dynamic
607 Properties. J Mol Biol 430:2096–2112.

608 48. Pham TH, Gao X, Singh G, Hardwidge PR. 2013. Escherichia coli Virulence Protein
609 NleH1 Interaction with the v-Crk Sarcoma Virus CT10 Oncogene-like Protein (CRKL)
610 Governs NleH1 Inhibition of the Ribosomal Protein S3 (RPS3)/Nuclear Factor κB (NF-
611 κB) Pathway. J Biol Chem 288:34567–34574.

612 49. Zhou X, Gewurz BE, Ritchie JM, Takasaki K, Greenfeld H, Kieff E, Davis BM, Waldor
613 MK. 2013. A Vibrio parahaemolyticus T3SS Effector Mediates Pathogenesis by
614 Independently Enabling Intestinal Colonization and Inhibiting TAK1 Activation. Cell
615 Reports 3:1690–1702.

616 50. Katoh K, Rozewicki J, Yamada KD. 2017. MAFFT online service: multiple sequence
617 alignment, interactive sequence choice and visualization. Brief Bioinform 20:bbx108-.

618 51. Notredame C, Higgins DG, Heringa J. 2000. T-Coffee: A Novel Method for Fast and
619 Accurate Multiple Sequence Alignment. JMB 302:205–217.

620 52. Robert X, Gouet P. 2014. Deciphering key features in protein structures with the new
621 ENDscript server. Nucleic Acids Res 42:W320–W324.

622 53. Crooks GE, Hon G, Chandonia J-M, Brenner SE. 2004. WebLogo: A Sequence Logo
623 Generator. Genome Res 14:1188–1190.

624 54. Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: Multiple Alignment of
625 Conserved Genomic Sequence With Rearrangements. Genome Res 14:1394–1403.

626 55. Bertelli C, Laird MR, Williams KP,
627 Simon Fraser University Research Computing Group, Lau BY, Hoad G, Winsor GL,
628 Brinkman FS. 2017. IslandViewer 4: expanded prediction of genomic islands for larger-
629 scale datasets. Nucleic Acids Res 45:W30–W35.

630 56. Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. 2012. Artemis: an
631 integrated platform for visualization and analysis of high-throughput sequence-based
632 experimental data. Bioinformatics 28:464–469.

633 57. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics
634 Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol 33:1870–1874.

635 58. Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new
636 developments. Nucleic Acids Res 47:gkz239-.

637 59. Buchan DWA, Jones DT. 2019. The PSIPRED Protein Analysis Workbench: 20
638 years on. Nucleic Acids Res 47:W402–W407.

639 60. 2020-4 SR. 2020. MAESTRO. New York.

640

641

FIGURE LEGENDS

FIG 1. VPA1328, an ORF in the *V. parahaemolyticus* RIMD2210633 T3SS2 gene cluster, is similar to the VopG effector protein encoded within the *V. cholerae* AM-19226 T3SS2 gene cluster. (A) Schematic depiction of a comparison of the T3SS2 gene clusters in *V. parahaemolyticus* RIMD2210633 and *V. cholerae* AM-19226. BLASTn alignment was performed and visualized using EasyFig. (B) Multiple Sequence alignment of VPA1328 and VopG homologs. BLASTp alignments were performed using T-Coffee Expresso and visualized by ESPript 3. Amino acids with a red background correspond to positions with 100% identity, amino acids with a yellow background correspond to positions with >70% identity.

FIG 2. VPA1328 is secreted in a T3SS2 dependent fashion. Immunoblots of culture supernatants and whole cell lysates using anti-CyA antibodies for CyA tagged VPA1328 (VopG) and VPA1357 (VopV) in WT (A) and (B) isogenic Δ*vscn1,* Δ*vscn2* and Δ*vscn1*Δ*vscn2* mutant strains. Immunoblots for RNAP was used as a control for cell lysis.

FIG 3. VopG homologs are widely distributed among *Vibrio* and *Shewanella* species. (A) Number of *Vibrio* sp and *Shewanella* sp isolates where VopG homologs were identified. (B) Distribution of the 122 non-redundant VopG protein sequences in different *Vibrio* and *Shewanella* species. (C) Phylogenetic analysis of the 122 non-redundant VopG homologs identified in this study. Phylogenetic analysis was performed with MEGA and visualized by iTOL. Distinct bacterial species are highlighted in different colors. VopG

666 homologs used in the multiple sequence alignment of Fig 5A are highlighted with a black
667 dot.
668
669 **FIG 4.** Location of *vopG* homologs within T3SS2 gene clusters. BLASTn alignments
670 were performed and visualized using EasyFig.
671
672 **FIG 5**. Conservation of the C-terminal domains of VopG homologs. (A) Multiple
673 sequence alignment and Weblogo analysis of the C-terminal domains of VopG proteins
674 (aminoacids 84-260 in VPA1328) with the serine/threonine kinase domains of NleH
675 proteins. G-rich, catalytic and activation loops of the kinase are highlighted in colored
676 boxes matching the schematic diagram in (B). BLASTp alignment was performed using
677 T-Coffee Expresso and visualized by ESPript 3. Amino acids within a red background
678 correspond to positions with 100% identity, amino acids with a yellow background
679 correspond to positions with >70% identity. The secondary structure of NleH1 and OspG
680 is shown flanking the alignment (α, alpha helices; β, beta sheets; T, turns). (B)
681 Schematic representation of VPA1328 highlighting the presence of the T3S signal as
682 well as conserved regions and key catalytic residues of the putative serine/threonine
683 kinase domain identified in (A).
684
685 **FIG 6**. VopG contain a NleH-like serine/threonine kinase domain. (A) Phylogenetic
686 analysis of bacterial serine/threonine kinases. The analysis was performed with MEGA
687 and visualized by iTOL. (B) Comparative homology model of the C-terminal domain of
688 VPA1328 (aminoacids 84-260) superimposed on the known structure OspG (PDB
689 4bvU). The inset depicts the catalytic domain of OspG and the superimposed predicted
690 structure of this region in VPA1328. Homology modelling was performed using the I-
691 TASSER pipeline and visualized with MAESTRO.

692

693 **FIG S1.** Genomic context of *vopG* in *Shewanella baltica* BA175 and *V. parahaemolyticus*
694 FORC014. (A) Schematic of the *vopG* genomic context within T3SS2 gene clusters of
695 *Shewanella* species and within the T3SS gene cluster of *V. parahaemolyticus* FORC014
696 (B). BLASTn alignment was performed and visualized using EasyFig.

697

698 **FIG S2.** Multiple sequence alignment of the DNA sequence of the *vopG* homologs
699 encoded within the T3SS2 gene cluster of *V. parahaemolyticus* strain FORC014.
700 BLASTn alignment was performed using T-Coffee and visualized by ESPript 3.
701 Nucleotides within a red background correspond to positions with 100% identity.

702

703 **FIG S3.** Weblogo analysis of the multiple sequence alignment of the 122 non-redundant
704 VopG proteins identified in this study.

705

706 **FIG S4.** Multiple sequence alignment of the amino acid sequences of the VopG
707 homologs from *V. parahaemolyticus* RIMD2210633 and *V. cholerae* AM-19226 with the
708 NleH1, NleH2 proteins of *E. coli* O157:H7 and OspG protein of *S. sonnei* strain Ss046.
709 BLASTp alignment was performed using M-Coffee and visualized by ESPript 3. Amino
710 acids within a red background corresponds to positions with 100% identity and amino
711 acids in a yellow background correspond to positions with over 70% of identity.

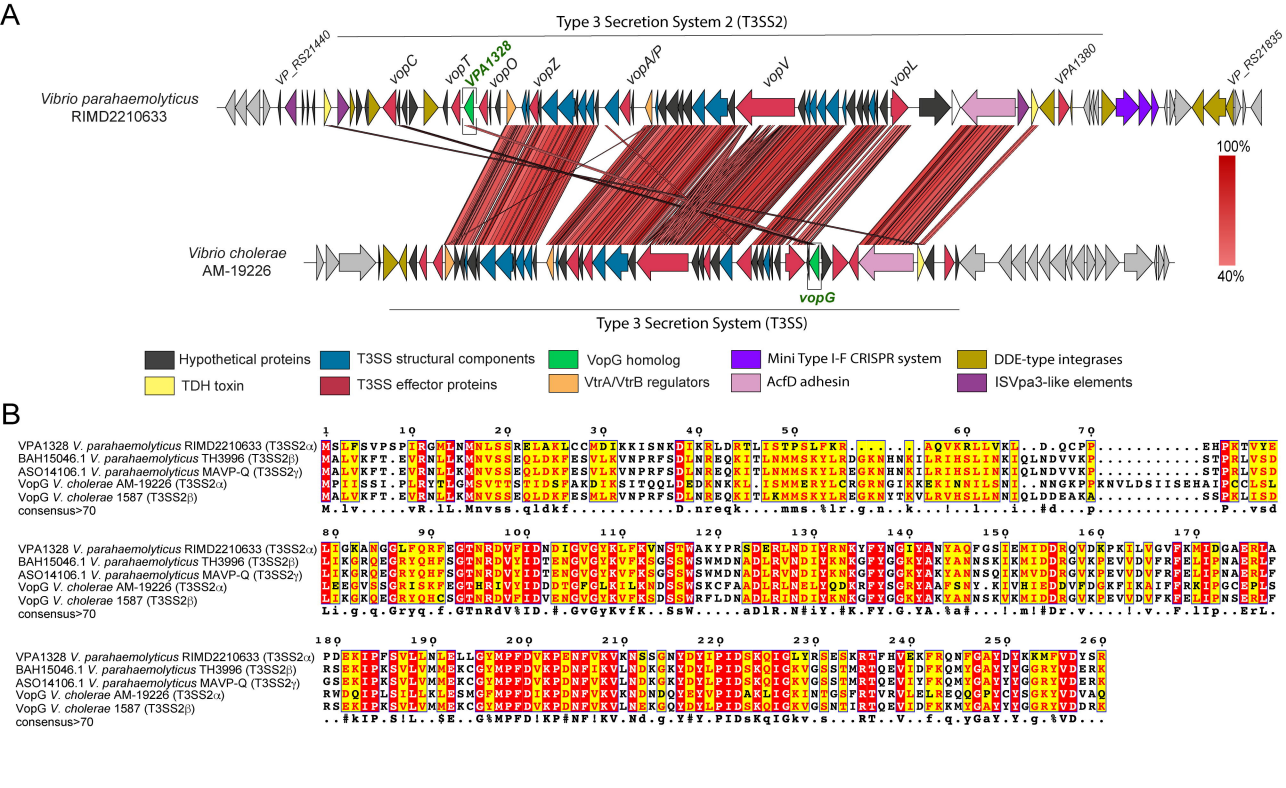
712

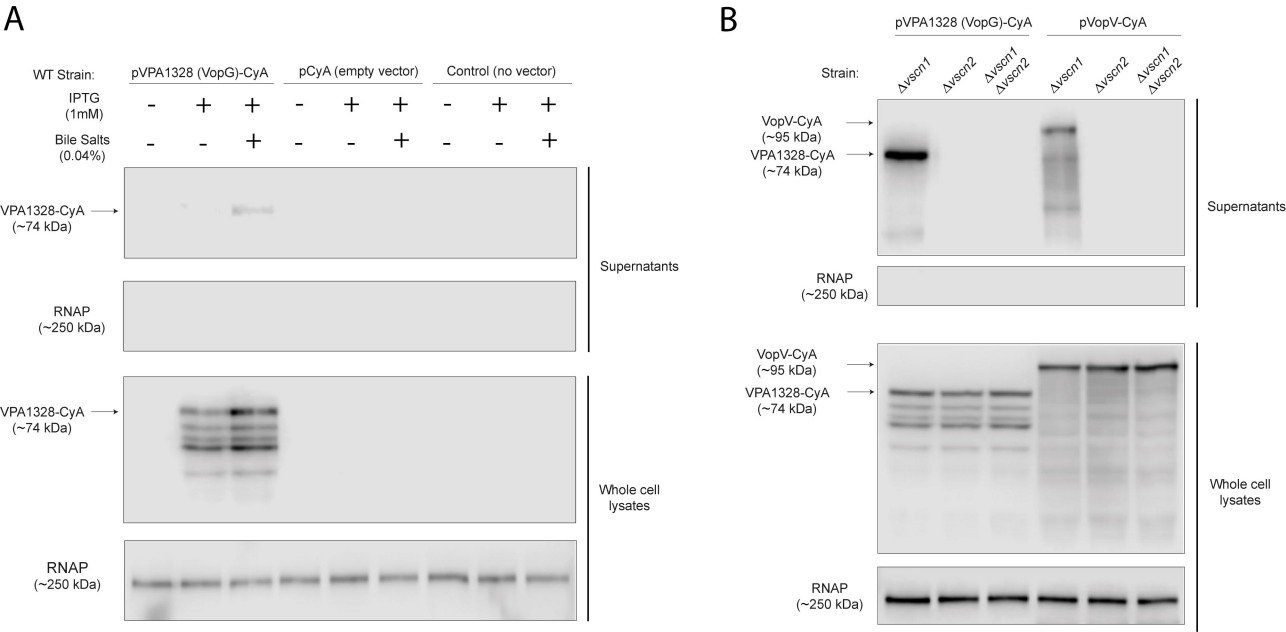
713 **FIG S5.** Estimated accuracy of the five protein structures of the C-terminal (amino acids
714 84-260) domain of VPA1328 obtained through comparative homology modelling using
715 the I-TASSER pipeline. The locations of the predicted G-rich, catalytic and activation
716 loop are highlighted.

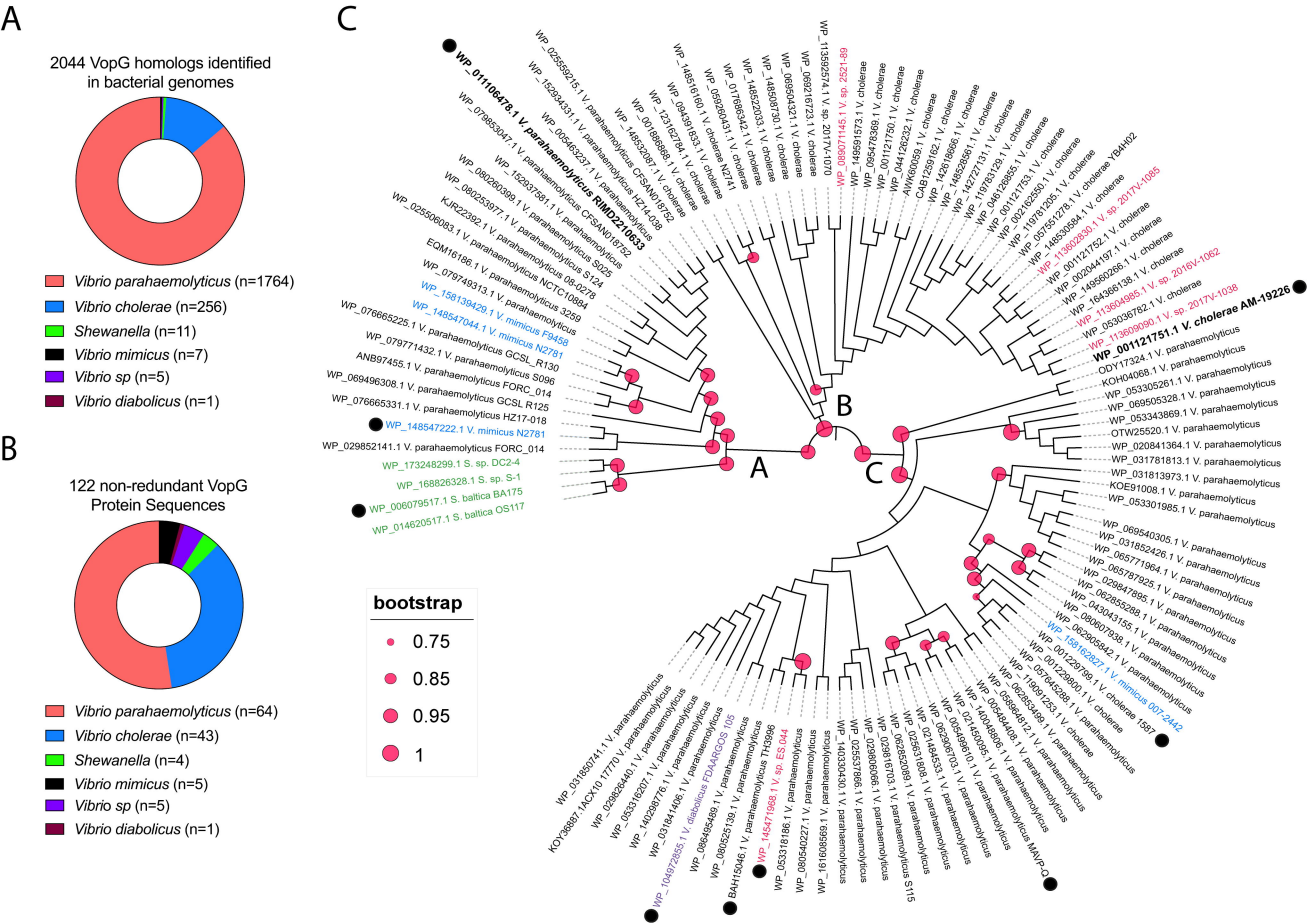
717 **Table S1.** Bacterial strains and plasmids used in this study.

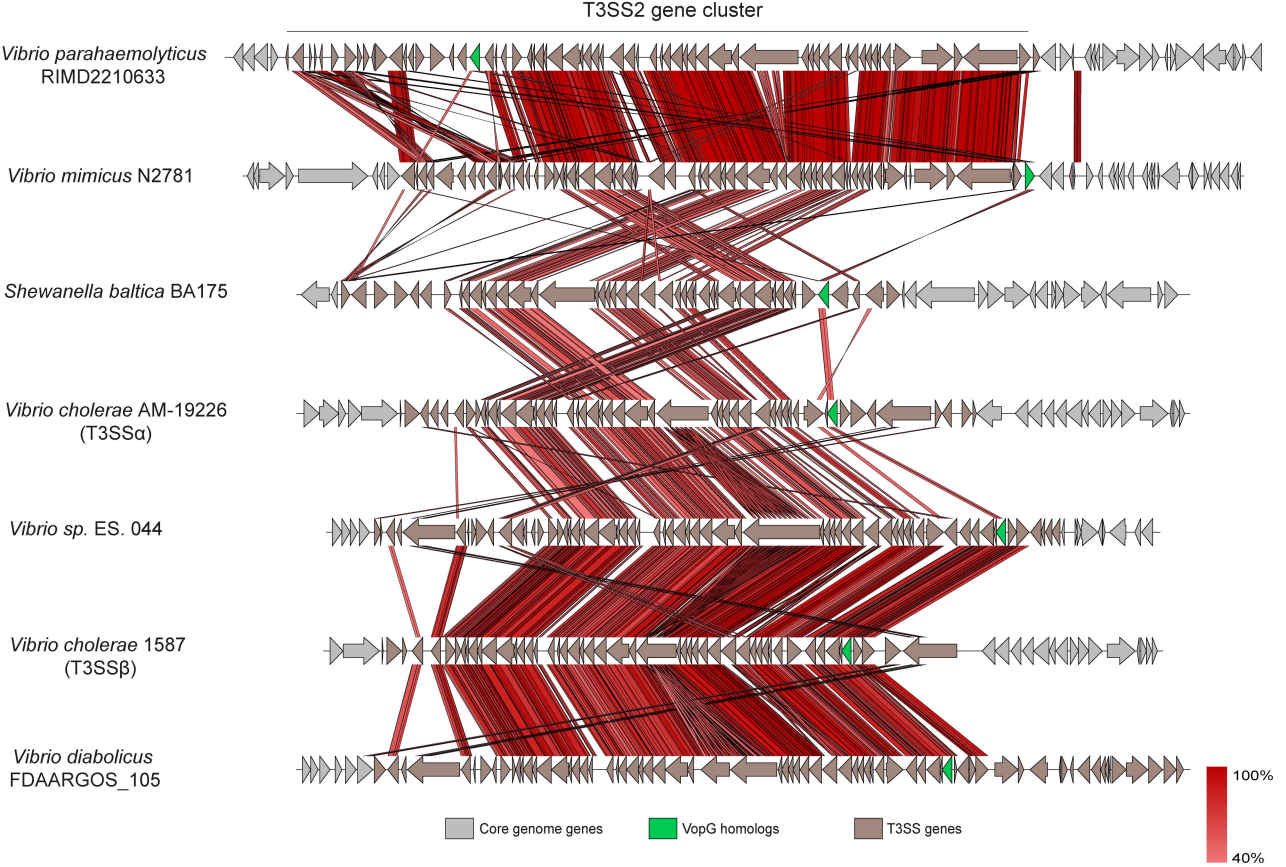
718 **Table S2.** Analysis of the VPA1328 ORF using different T3SS effector prediction
719 software.

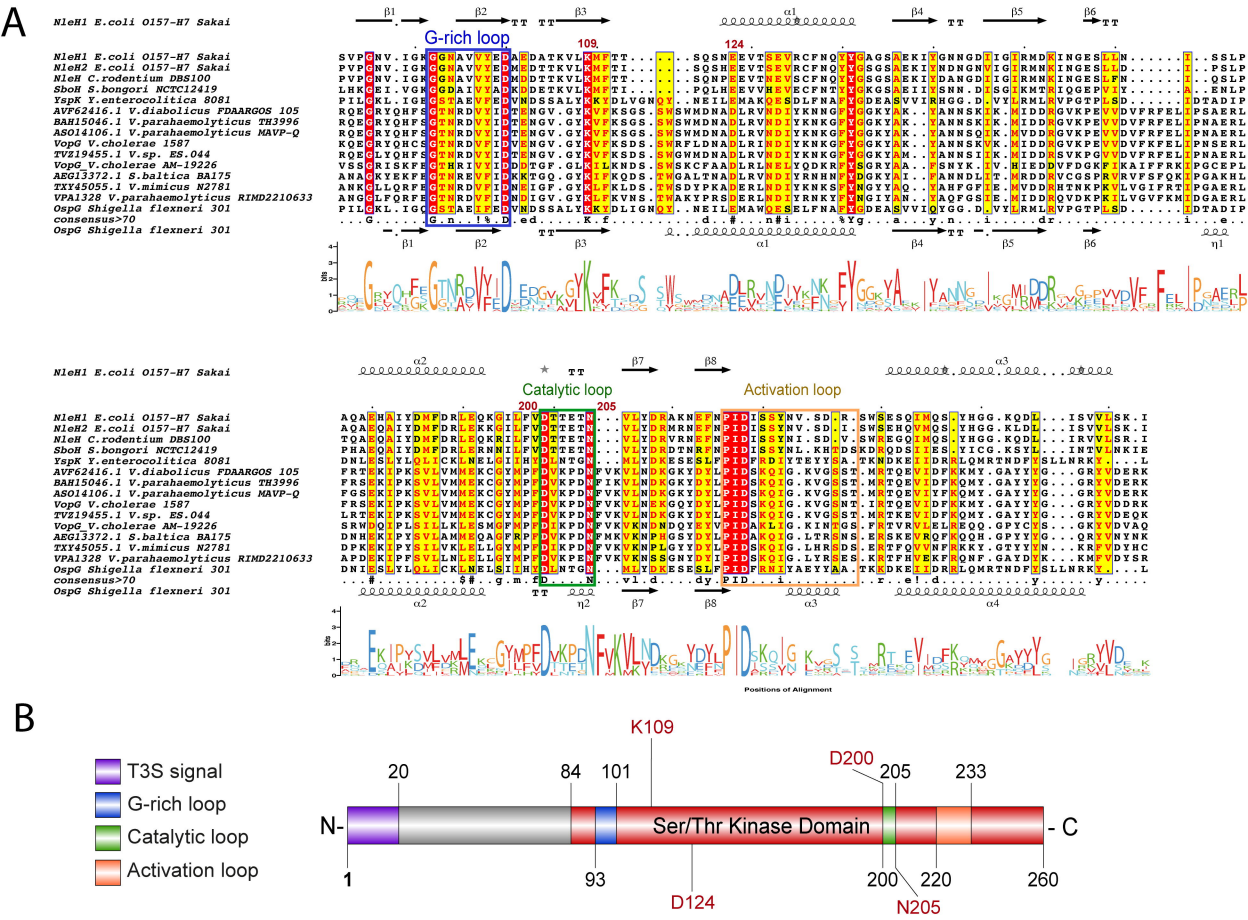
720 **Table S3.** List of the 2044 total VopG protein sequences and the 122 non-redundant
721 VopG protein sequences identified in bacterial genome databases.


A
Type 3 Secretion System 2 (T3SS2)
VP_RS21440
vopC
vopT
VPA1328
vopO
vopZ
vopA/P
vopV
vopL
VPA1380
VP_RS21835
Vibrio parahaemolyticus
RIMD2210633
100%
40%
Vibrio cholerae
AM-19226
vopG
Type 3 Secretion System (T3SS)
Hypothetical proteins
T3SS structural components
VopG homolog
Mini Type I-F CRISPR system
DDE-type integrases
TDH toxin
T3SS effector proteins
VtrA/VtrB regulators
AcfD adhesin
ISVpa3-like elements
B
VPA1328 V. parahaemolyticus RIMD2210633 (T3SS2α)
BAH15046.1 V. parahaemolyticus TH3996 (T3SS2β)
ASO14106.1 V. parahaemolyticus MAVP-Q (T3SS2γ)
VopG V. cholerae AM-19226 (T3SS2α)
VopG V. cholerae 1587 (T3SS2β)
consensus>70

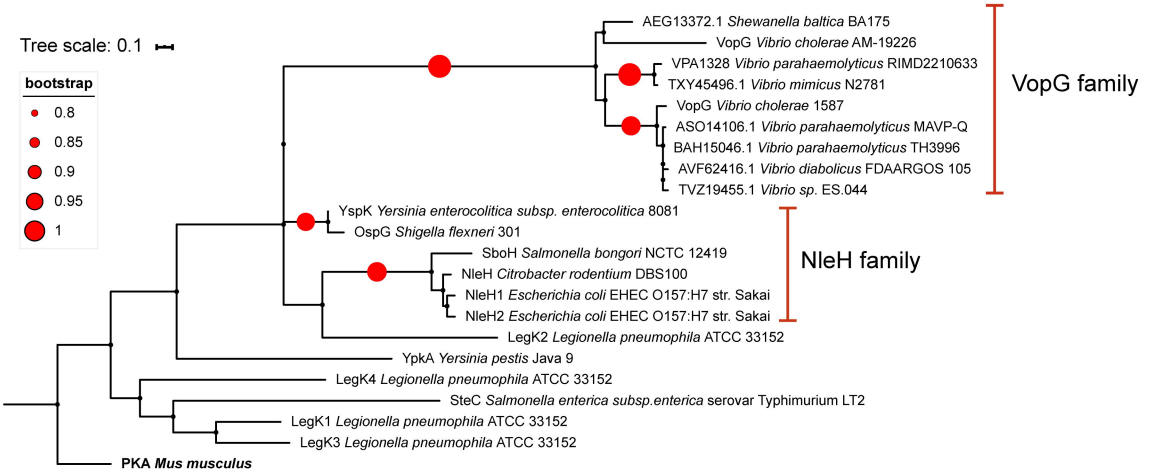

A
WT Strain:
pVPA1328 (VopG)-CyA
pCyA (empty vector)
Control (no vector)
IPTG (1mM)
Bile Salts (0.04%)
- + + - + + - + +
- - + - - + - - +
VPA1328-CyA (~74 kDa)
RNAP (~250 kDa)
Supernatants
VPA1328-CyA (~74 kDa)
RNAP (~250 kDa)
Whole cell lysates
B
pVPA1328 (VopG)-CyA
pVopV-CyA
Strain:
Δvscn1
Δvscn2
Δvscn1 Δvscn2
Δvscn1
Δvscn2
Δvscn1 Δvscn2
VopV-CyA (~95 kDa)
VPA1328-CyA (~74 kDa)
RNAP (~250 kDa)
Supernatants
VopV-CyA (~95 kDa)
VPA1328-CyA (~74 kDa)
RNAP (~250 kDa)
Whole cell lysates


A
2044 VopG homologs identified in bacterial genomes
Vibrio parahaemolyticus (n=1764)
Vibrio cholerae (n=256)
Shewanella (n=11)
Vibrio mimicus (n=7)
Vibrio sp (n=5)
Vibrio diabolicus (n=1)
B
122 non-redundant VopG Protein Sequences
Vibrio parahaemolyticus (n=64)
Vibrio cholerae (n=43)
Shewanella (n=4)
Vibrio mimicus (n=5)
Vibrio sp (n=5)
Vibrio diabolicus (n=1)
C
bootstrap
0.75
0.85
0.95
1
A
B
C

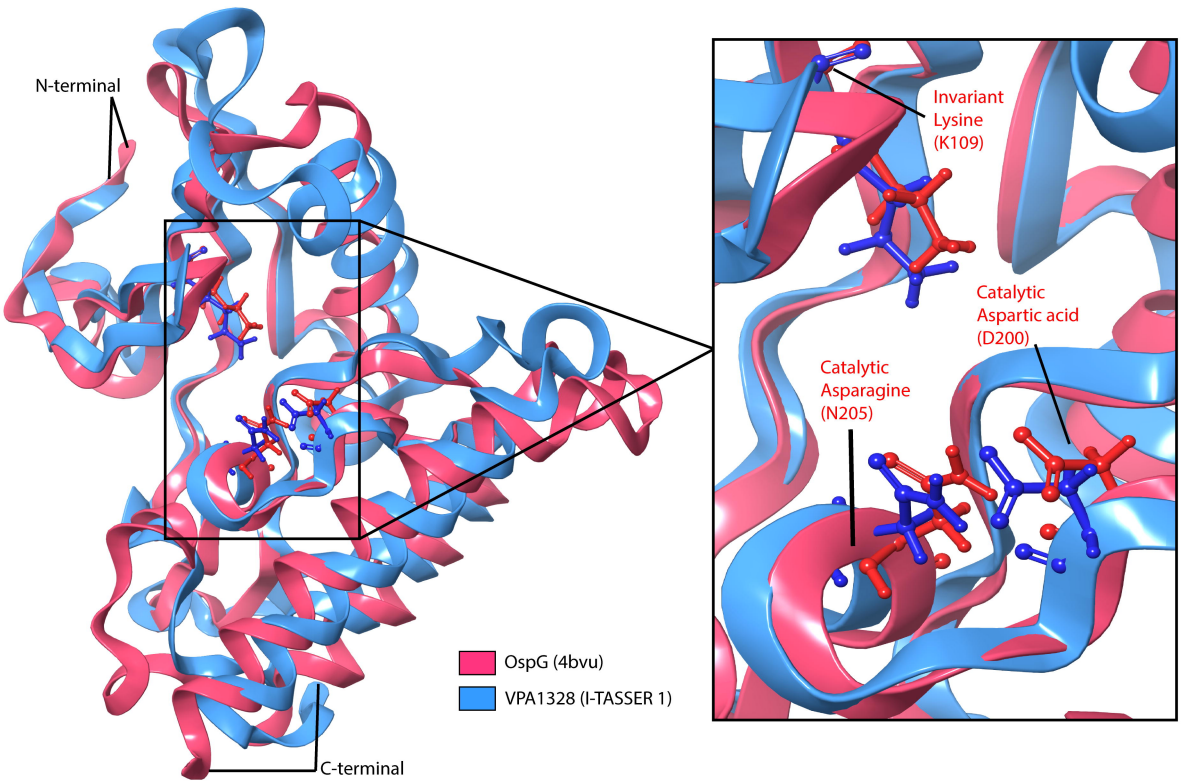

T3SS2 gene cluster
Vibrio parahaemolyticus RIMD2210633
Vibrio mimicus N2781
Shewanella baltica BA175
Vibrio cholerae AM-19226 (T3SSα)
Vibrio sp. ES. 044
Vibrio cholerae 1587 (T3SSβ)
Vibrio diabolicus FDAARGOS_105
Core genome genes
VopG homologs
T3SS genes
100%
40%


A
NleH1 E.coli O157-H7 Sakai
G-rich loop
109
124
NleH1 E.coli O157-H7 Sakai
NleH2 E.coli O157-H7 Sakai
NleH C.rodentium DBS100
SboH S.bongori NCTC12419
YspK Y.enterocolitica 8081
AVF62416.1 V.diabolicus FDAARGOS 105
BAH15046.1 V.parahaemolyticus TH3996
ASO14106.1 V.parahaemolyticus MAVP-Q
VopG V.cholerae 1587
TVZ19455.1 V.sp. ES.044
VopG_V.cholerae AM-19226
AEG13372.1 S.baltica BA175
TXY45055.1 V.mimicus N2781
VPA1328 V.parahaemolyticus RIMD2210633
OspG Shigella flexneri 301
consensus>70
OspG Shigella flexneri 301
NleH1 E.coli O157-H7 Sakai
Catalytic loop
Activation loop
200
205
NleH1 E.coli O157-H7 Sakai
NleH2 E.coli O157-H7 Sakai
NleH C.rodentium DBS100
SboH S.bongori NCTC12419
YspK Y.enterocolitica 8081
AVF62416.1 V.diabolicus FDAARGOS 105
BAH15046.1 V.parahaemolyticus TH3996
ASO14106.1 V.parahaemolyticus MAVP-Q
VopG V.cholerae 1587
TVZ19455.1 V.sp. ES.044
VopG_V.cholerae AM-19226
AEG13372.1 S.baltica BA175
TXY45055.1 V.mimicus N2781
VPA1328 V.parahaemolyticus RIMD2210633
OspG Shigella flexneri 301
consensus>70
OspG Shigella flexneri 301
Positions of Alignment
B
T3S signal
G-rich loop
Catalytic loop
Activation loop
K109
D200
20
84
101
205
233
N-
Ser/Thr Kinase Domain
- C
1
93
200
220
260
D124
N205

A


Tree scale: 0.1
bootstrap
0.8
0.85
0.9
0.95
1
AEG13372.1 Shewanella baltica BA175
VopG Vibrio cholerae AM-19226
VPA1328 Vibrio parahaemolyticus RIMD2210633
TXY45496.1 Vibrio mimicus N2781
VopG Vibrio cholerae 1587
ASO14106.1 Vibrio parahaemolyticus MAVP-Q
BAH15046.1 Vibrio parahaemolyticus TH3996
AVF62416.1 Vibrio diabolicus FDAARGOS 105
TVZ19455.1 Vibrio sp. ES.044
VopG family
YspK Yersinia enterocolitica subsp. enterocolitica 8081
OspG Shigella flexneri 301
SboH Salmonella bongori NCTC 12419
NleH Citrobacter rodentium DBS100
NleH1 Escherichia coli EHEC O157:H7 str. Sakai
NleH2 Escherichia coli EHEC O157:H7 str. Sakai
NleH family
LegK2 Legionella pneumophila ATCC 33152
YpkA Yersinia pestis Java 9
LegK4 Legionella pneumophila ATCC 33152
SteC Salmonella enterica subsp.enterica serovar Typhimurium LT2
LegK1 Legionella pneumophila ATCC 33152
LegK3 Legionella pneumophila ATCC 33152
PKA Mus musculus

B


N-terminal
C-terminal
OspG (4bvu)
VPA1328 (I-TASSER 1)
Invariant Lysine (K109)
Catalytic Aspartic acid (D200)
Catalytic Asparagine (N205)