gbpA and chiA genes are not uniformly distributed amongst diverse Vibrio cholerae

Thea G. Fennell1,2†, Grace A. Blackwell1,3, Nicholas R. Thomson1,4 and Matthew J. Dorman1,2,*

Abstract
Members of the bacterial genus Vibrio utilize chitin both as a metabolic substrate and a signal to activate natural competence. Vibrio cholerae is a bacterial enteric pathogen, sub-lineages of which can cause pandemic cholera. However, the chitin metabolic pathway in V. cholerae has been dissected using only a limited number of laboratory strains of this species. Here, we survey the complement of key chitin metabolism genes amongst 195 diverse V. cholerae. We show that the gene encoding GbpA, known to be an important colonization and virulence factor in pandemic isolates, is not ubiquitous amongst V. cholerae. We also identify a putatively novel chitinase, and present experimental evidence in support of its functionality. Our data indicate that the chitin metabolic pathway within V. cholerae is more complex than previously thought, and emphasize the importance of considering genes and functions in the context of a species in its entirety, rather than simply relying on traditional reference strains.

DATA SUMMARY
The authors confirm that all supporting data, code and protocols have been provided within the article or through supplementary data files.

(1) No whole-genome sequencing data were generated in this study. Accession numbers for the publicly available sequences used for these analyses are listed in Tables 2 and S1 (available in the online version of this article), and the Methods.

(2) Metadata, accession numbers and references for each genome included in this study are provided in Table S1, deposited in Figshare: (https://dx.doi.org/10.6084/m9.figshare.14398688).

(3) Supplementary tables, figures and references are available in the Supplementary Material file, deposited in Figshare: (https://dx.doi.org/10.6084/m9.figshare.14398688).

(4) All other data which underpin the figures in this paper, including pan-genome data matrices, modified and unmodified sequence alignments and phylogenetic trees, original images of gels and immunoblots, raw fluorescence data, amplicon sequencing reads, and the R code used to generate Fig. 5, are available in Figshare: https://dx.doi.org/10.6084/m9.figshare.13169189 (DOI).

INTRODUCTION
The genus Vibrio of marine gammaproteobacteria contains a number of virulent human pathogens, of significant public health concern [1]. Most notorious of these pathogens is Vibrio cholerae, members of which are the aetiological agent of cholera in humans [2, 3]. Two biochemically defined and distinct V. cholerae biotypes are associated with cholera pandemics. Classical biotype V. cholerae are believed to have caused the first six pandemics [2–4], whilst the current seventh pandemic (1961 to the present) is attributed to El Tor biotype V. cholerae [5, 6]. Genomic evidence has shown that classical V. cholerae form a discrete phylogenetic lineage from the lineage causing the seventh cholera pandemic, dubbed
Impact Statement

It is thought that the ability to metabolize chitin is ubiquitous amongst Vibrio species, and that this enables these species to survive in aqueous and estuarine environmental contexts. Although chitin metabolism pathways have been detailed in several members of this genus, little is known about how these processes vary within a single Vibrio species. Here, we present the distribution of genes encoding key chitinase and chitin-binding proteins across diverse Vibrio cholerae, and show that our canonical understanding of this pathway in this species is challenged when isolates from non-pandemic V. cholerae lineages are considered alongside those linked to pandemics. Furthermore, we show that genes previously thought to be species core genes are not in fact ubiquitous, we identify novel components of the chitin metabolic cascade in this species, and we present functional validation for these observations.

The pathways by which chitin is degraded and utilized by V. cholerae have been described in detail [31], as it has been in other members of the genus (e.g. [32–39]). Although a comprehensive review of the chitin utilization pathway is beyond the scope of this paper, it is important to appreciate the complexity of this pathway. Chitin degradation, import and metabolism in V. cholerae involves at least 27 proteins, 24 encoded by genes on chromosome 1, and three by genes on the smaller chromosome 2 [20]. Here, focus will be directed to the initial stages of chitin metabolism - adhesion to a chitinous substrate, and expression of extracellular degradative chitinase enzymes.

The first step in chitin metabolism is the attachment of V. cholerae to chitinous surfaces through interactions with GlcNAc. This is mediated both by the mannose-sensitive haemagglutinin (MSHA) pilus and the chitin adhesin GbpA (encoded by VC_A0811, accession no. AAF96709.1) [24, 40]. Although GbpA was initially identified as a putative chitinase enzyme [24], it was shown to be an adhesin induced by GlcNAc which enabled V. cholerae to attach to chitinous substrates [24]. Subsequently, it was found that as well as mediating attachment of V. cholerae to chitin, GbpA is also required for the successful colonization of the intestine [40]. This is thought to occur through interactions with mucin – GbpA interacts with mucin in the intestine, and gbpA transcription increases upon exposure of V. cholerae to mucin [41]. The crystal structure and domain architecture of GbpA have been determined [42], and the fourth domain of GbpA is structurally similar to the chitin-binding domain of known chitinases [42]. Evidence also suggests that GbpA has lytic polysaccharide monooxygenase activity [43], and that GbpA activity is higher at low population densities due to the activity of quorum-sensing-regulated proteases [44].

Once V. cholerae adheres to a chitinous surface, extracellular endochitinase enzymes are required for the bacterium to hydrolyse complex chitin polymers into oligosaccharides which can be imported into the cell for further metabolism [35]. As many as seven putative endochitinases have been identified in V. cholerae [20, 45, 46], two of which, ChiA-1 (encoded by VC_1952, accession no. AAF95100.1) and ChiA-2 (encoded by VC_A0027, accession no. AAF95941.1), are the principal chitinases required for V. cholerae chitin catabolism [24, 31, 45, 47]. ChiA-1 was first shown to be an extracellular chitinase in 1998 [48]; subsequently, ChiA-2 was shown to be important for intestinal colonization and for metabolizing mucin in the intestine by V. cholerae strain N16961 (N16961) [30]. ChiA-2 is also the most highly expressed chitinase in El Tor biotype V. cholerae strain E7946 [45]. Both ChiA-1 and ChiA-2 are essential for V. cholerae to grow in media supplemented with colloidal chitin [24]. Once chitin oligomers have been digested by extracellular chitinases, the resultant oligosaccharides are thought to enter the bacterial periplasm via the chitoporin ChiP (encoded by VC_0972, accession no. AAF94134.1) and by other as-yet-uncharacterized porins [24, 36, 49], and subsequently transported to the cytoplasm
via phosphotransferase (PTS) and ABC-type transporters [20, 36] (Fig. 1).

Previous work used the genomes of 20 diverse Vibrionaceae (including seven V. cholerae) to determine the presence and absence of genes involved in metabolizing chitin across this family of bacteria [20]. However, it is important to note that the chitin degradation pathway of V. cholerae has been described using reference strains of the species (particularly N16961 [24]), and although data exist on how the chitin catabolism pathway varies amongst members of the genus Vibrio [20], less is known about how this pathway varies within a single species. This is particularly relevant because emerging evidence suggests that non-7PET lineages of V. cholerae cause different patterns of disease, even if they harbour some or all of the canonical pathogenicity determinants associated with cholera cases [8]. However, because the chitin metabolic pathway has principally been studied in N16961, a 7PET strain, we know little about the extent to which it varies amongst non-pandemic members of V. cholerae.

In this study, we focused specifically on genes that encode components of the initial steps of the chitin degradation pathway across V. cholerae. We focused on these because the functions of many of these genes have been characterized experimentally, and we sought to determine how well the observations in the literature reflect the true distribution of these genes, and their functions, across a diverse species. We generated a pangenome from 195 annotated V. cholerae genome sequences, which were chosen to obtain as balanced and unbiased a view of V. cholerae as possible (i.e. without focusing solely on epidemic and pandemic lineages). We find that the distribution of these genes is not uniform within V. cholerae, and we identify variation amongst the chitinases encoded by diverse V. cholerae.

Fig. 1. Initial steps in V. cholerae chitin uptake and catabolism. Schematic summarizing the principal stages in chitin degradation and import by V. cholerae. Comprehensive descriptions of this pathway are reported in the literature [20, 36]. The MSHA adhesin has not been included in this diagram.
We also identify a putatively novel chitinase gene, and present experimental evidence in support of its functional classification.

**METHODS**

**Strains, plasmids and oligonucleotides**

Strains, plasmids and oligonucleotide primers (Sigma-Aldrich) used for experimental work in this study are listed in Table 1. Bacteria were cultured routinely on LB media supplemented with chloramphenicol (10 µg ml⁻¹; LB-Cm) where appropriate.

**Genome sequences and accession numbers**

The 198 genome sequences used to calculate the pangenome described in this paper are listed in Table S1. Accession numbers for additional chromosome sequences to which the text refers are as follows: *V. harveyi* chromosome 2 (accession no. CP009468.1); *V. parahaemolyticus* chromosome 2 (accession no. BA000032.2). Accession numbers for the chitinase protein sequences referred to in Hunt et al. [20] and used for blastp comparisons are listed in Table 2.

**Genome assemblies**

*V. cholerae* genome sequences were assembled from short-read data using SPAdes v3.8.2 [50], as part of a high-throughput pipeline [51]. Assemblies were annotated automatically using Prokka v1.5 [52] and a genus-specific reference database [53]. If raw sequencing reads were unavailable for genome sequences, assemblies were downloaded and similarly annotated using the automated Prokka-based pipeline.

**Pangenome and phylogenetic calculations**

A pangenome was produced from 198 Prokka-annotated genome assemblies using Roary v3.12.0 [54] (parameters: ‘-p 10 -e --mafft -s -cd 97’). A core-gene alignment of 2520 genes and 1096140 nucleotides was produced from this pangenome calculation. The alignment was trimmed using trimAl v1.4.1 [55] and used to produce an alignment of 183896 single
nucleotide variants (SNVs) using SNP-sites v2.5.1 [56]. A maximum-likelihood phylogeny was produced using IQ-Tree v1.6.10 [57] from the SNV-only alignment (options: '- nt 10 -m GTR+ASC -bb 5000 -alrt 5000').

Protein sequence alignments, domain prediction and comparative genomics
Protein sequences were aligned using blast p [58] and were annotated using the InterProScan web server [59]. Comparative genomic figures were generated using blast n [58] sequence alignments and visualized using ACT v13 and v18.0.2 [60], and Easyfig v2.2.2 [61].

Confirmation of gene presence/absence by mapping
Reads were mapped to reference sequences using SMALT v0.7.4 (https://www.sanger.ac.uk/tool/smalt-0/) and the method described by Harris et al. [62], as part of automated analysis pipelines run by Wellcome Sanger Institute Pathogen Informatics. All of the software developed by Pathogen Informatics is freely available for download from GitHub under an open source licence, GNU GPL 3 (https://github.com/sanger-pathogens/vr-codebase). Ordered BAM files were visualized against reference sequences using Artemis v16 and v18.0.2, which incorporate BamView [63, 64].

Molecular cloning
Plasmid DNA was extracted from Escherichia coli using the QIAprep Spin Miniprep kit (Qiagen; #27104). Genomic DNA (gDNA) was extracted from NCTC 30 as described previously [65]. Cloning intermediates were purified using the QIAquick PCR Purification kit (Qiagen; #28104). gDNA from NCTC 30 was used as a template from which to amplify chiA-3 using primers oMJD202 and oMJD203, high-fidelity Phusion Hot Start Flex polymerase [New England Biolabs (NEB); #M0535S] using the supplied GC buffer, DMSO (3%, v/v, final concentration) and dNTPs (Thermo Scientific; #R0191). Twenty-nine PCR cycles were performed using the manufacturer’s protocol (annealing temperature: 55°C, extension time: 2 min). The amplicon was purified and digested using 30 units of SacI-HF and SalI-HF (NEB; #R3156S and R3138S respectively) at 37°C for 45 min. pBAD33 was similarly treated with SacI-HF and SalI-HF, and after 15 min incubation at 37°C, the plasmid digestion was supplemented with 1.5 units of recombinant shrimp alkaline phosphatase (rSAP; NEB; #M0371S).

Digested insert and vector were purified and ligated together at room temperature for 30 min using T4 DNA ligase (NEB; #M0202S) in approximately a 3:1 molar ratio. Chemically competent 10-beta E. coli (NEB; #C3019I) were transformed with ligated DNA according to the manufacturer’s instructions, and transformants were selected on LB agar supplemented with chloramphenicol (10 µg ml⁻¹).

Chloramphenicol-resistant colonies were resuspended in 30 µl PBS. A screen for clones containing an insert into pBAD33 was carried out using 1 µl of this suspension as a PCR template using primers oMJD204 and oMJD205 and OneTaq Quickload 2X Master Mix (NEB; #M0486S), according to the manufacturer’s instructions (annealing temperature 45°C, extension time 3 min). Plasmids were extracted from overnight cultures of clones from which PCR produced an amplicon of the expected size (1548 bp). The presence of an insertion into pBAD33 was verified by digesting purified plasmid DNA with SacI-HF and SalI-HF as described above. Plasmids were then sequence-confirmed by amplicon sequencing (GATC/Eurofins Genomics) in both directions across the pBAD33 multiple cloning site using primers oMJD204 and oMJD205. Sequence-verified plasmids were transformed into chemically competent NiCo21(DE3) cells (NEB; #C2529H) following the manufacturer’s instructions, and these transformants were used for protein expression purposes.

Table 2. Pairwise BLASTp alignments between chiA-3 and chitinases from Hunt et al. [20]

<table>
<thead>
<tr>
<th>Chitinase gene</th>
<th>Species</th>
<th>Accession no.</th>
<th>Subject length (aa)</th>
<th>e-value</th>
<th>Covered by query (%)</th>
<th>Identity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>chiA-3 (self)</td>
<td>Vibrio cholerae</td>
<td>n/a</td>
<td>431</td>
<td>0</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>SKA34_14935</td>
<td>Photobacterium sp.</td>
<td>EAR55415.1</td>
<td>441</td>
<td>5e-116</td>
<td>99</td>
<td>39.63</td>
</tr>
<tr>
<td>SKA34_13330</td>
<td>Photobacterium sp.</td>
<td>EAR55445.1</td>
<td>399</td>
<td>6e-08</td>
<td>21</td>
<td>28.97</td>
</tr>
<tr>
<td>P3TCK_21620</td>
<td>Photobacterium profundum</td>
<td>EAS45126.1</td>
<td>948</td>
<td>2e-04</td>
<td>30</td>
<td>27.47</td>
</tr>
<tr>
<td>VAS14_08875</td>
<td>Vibrio angustum</td>
<td>EAS63573.1</td>
<td>560</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>VAS14_08910</td>
<td>Vibrio angustum</td>
<td>EAS63580.1</td>
<td>732</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>V12G01_01435</td>
<td>Vibrio alginolyticus</td>
<td>EAS76629.1</td>
<td>718</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>V12G01_22308</td>
<td>Vibrio alginolyticus</td>
<td>EAS77796.1</td>
<td>307</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>VF1146</td>
<td>Aliivibrio fischeri</td>
<td>AW85641.1</td>
<td>789</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>VPA1177</td>
<td>Vibrio parahaemolyticus</td>
<td>BAC62520.1</td>
<td>430</td>
<td>0</td>
<td>100</td>
<td>76.57</td>
</tr>
</tbody>
</table>
Protein expression and immunoblotting

Single colonies of NiCo21(DE3) harbouring pMJ157 and pBAD33 (empty vector) were inoculated into 3 ml LB-Cm and cultured at 37 °C with shaking (200 r.p.m.) for 8 h. These were then diluted 1:100 into baffled flasks containing 25 ml LB-Cm supplemented with either d-(+)-glucose (BDH, now VWR; #101176K) or L-(+)-arabinose (Sigma-Aldrich; #A3256), both at 0.4% (w/v) final concentration. These cultures were grown for 18 h at 23 °C with shaking (200 r.p.m.). Cells were collected by centrifugation (3900 g, 5 min) and the supernatant was filter-sterilized (0.22 µm) and stored at −20 °C. Cell pellets were lysed in 3 ml BugBuster HT (Millipore; #70922-4) for 30 min at room temperature on a rotator. Debris was collected by centrifugation (3900 g, 5 min) and discarded. Lysates were stored at −20 °C.

Sixty microlitres of filtered supernatants and lysates was mixed 1:1 with 2× tris-glycine-SDS sample buffer (Invitrogen; #LC2676), boiled at 100 °C for 5 min, and 30 µl of each sample was used to load duplicate NuPAGE 4–12 % Bis-Tris standards were marked manually on the developed film. ECL film (GE Healthcare; #28906836). Coloured protein size standards were marked manually on the developed film. Fluorescence signal was allowed to decay overnight, with Clarity Western ECL substrate (Bio-Rad; #170-5060) for 5 min. Luminescence signal was allowed to decay overnight, and the blot was then imaged with Amersham Hyperfilm ECL film (GE Healthcare; #28906836). Coloured protein size standards were marked manually on the developed film.

Chitinase assay

Chitinase activity was assayed using fluorogenic substrates (Sigma-Aldrich; #CS1030). The kit was used according to the manufacturer’s instructions, with the following modifications. Ten microlitres of cell lysate or supernatant was used per assay well. Five microlitres of the supplied chitinase control enzyme was used per positive control reaction, rather than a 1:200 dilution of the control enzyme, to ensure that fluorescence was detectable. Assays were carried out in black Nunc flat-bottomed microtitre plates (Sigma-Aldrich; #P8741), and technical triplicates were included for each sample. Once mixed, reaction plates were incubated for 30 min (37 °C, static) before the addition of stop solution. Fluorescence was measured using a FLUOstar Omega plate reader (BMG Lab’Tech), set to excitation and emission wavelengths of 360 and 450 nm, respectively. A 1% gain was applied to the fluorescence measured by the reader. Blank fluorescence was subtracted from each sample reading prior to analysis.

RESULTS

Distribution of chitinase genes amongst V. cholerae

The key components of V. cholerae chitin catabolism summarized in Fig. 1 have been previously described [24, 45, 69]. The presence and absence of orthologues of each of the principal chitin-binding proteins and extracellular chitinases [45] known to be encoded by the V. cholerae 7PET reference strain N16961 (based on their N16961 locus identifiers) were identified in a pangenome calculated from 195 V. cholerae genomes, plus three Vibrio species genomes used as an outgroup (Table S1). Genes that were annotated as encoding putative chitinases, as well as those genes known to be present in N16961, were identified in the pangenome (Tables S2 and S3). A V. cholerae phylogenetic tree was calculated using an SNV-only alignment of 2520 core genes taken from the pangenome, and the distribution of these chitinase genes across the phylogeny is presented in Fig. 2.

chiA-2 is ubiquitous amongst diverse V. cholerae, but gbpA and chiA-1 are not

The first, and most striking, observation made from these data was that gbpA (VC_A0811) did not appear to be ubiquitous amongst all of the V. cholerae genomes included in this study. This was surprising because this gene had previously been reported to be ubiquitous amongst V. cholerae [70]. We found that gbpA was present in only 189 of 195 V. cholerae genomes (96.9%; Fig. 2, Table S2). We manually inspected the genome assembly for each isolate which lacked gbpA, to guard against this being an artefact of the computational approach taken (Fig. S1). Three genomic arrangements were observed at this locus – the presence of an intact VC_A0811 locus as found in gbpA’ genomes, a deletion of gbpA and two adjacent genes (VC_A0811-VC_A0813), and the replacement of these three genes with additional sequence in the genome of NCTC 30 (Fig. S1). In order to ensure that the VC_A0811-VC_A0813 genes were not present at a different position in the NCTC 30 genome, we mapped the Illumina short-reads for this
isolate to the N16961 reference sequence and inspected the mapping coverage across this region. This confirmed that the absence of the genes VC_A0811–VC_A0813 from NCTC 30 was not a result of a mis-assembly (Fig. S2). The two genes adjacent to gbpA, VC_A0812 and VC_A0813, encode LapX and Lap, respectively. Both genes are putatively regulated by the HapR master quorum-sensing regulator, and encode proteins that were detected in an hapA mutant [71]. Both Lap and LapX were found to be putative components of the Type 2 secretome in N16961 [72], and lap has been used as a polymorphic locus in multilocus enzyme electrophoresis (MLEE) schemes for classifying V. cholerae [73, 74]. We were unable to find published evidence linking these genes to GbpA activity or chitin adhesion more generally, although we note that lap and lapX are oriented in the same direction as gbpA, and we cannot exclude the possibility that these three genes are co-regulated or co-transcribed.

We found that VC_A0027 (encoding ChiA-2) was near-ubiquitous, initially being detected in 192/195 V. cholerae (Fig. 2; Table S2). Manual inspection of the assemblies for those three isolates which appeared to lack the gene confirmed that the majority of this gene was in fact present; assembly and resultant annotation errors were likely to be responsible for this result (data not shown). This suggests that VC_A0027 is core to V. cholerae (195/195 genomes). This is consistent with ChiA-2 being the most highly expressed chitinase enzyme in the species, and with the observation that deletion of this gene alone causes a significant growth defect on minimal media containing chitin as a sole carbon source [45].

However, although VC_1952 (ChiA-1) was present in all pandemic isolates (defined as those isolates which were members of the 7PET and Classical lineages), it was not ubiquitous across the species, and was only identified in 61.2% of the non-pandemic V. cholerae in this dataset (101/165; Fig. 2; Table S2). This observation was surprising, because both ChiA-1 and ChiA-2 have been shown to be necessary for V. cholerae N16961 to grow in media supplemented with colloidal chitin [24]. Keymer and colleagues previously observed, using microarray approaches, that some diverse environmental isolates of V. cholerae varied in terms of their VC_1952 genotype [75]. We propose that our data recapitulate this observation, albeit in silico. We manually examined the region surrounding the VC_1952 locus in a subset of the genome assemblies for isolates lacking this gene, and found both that the gene was absent in its entirety, and that this did not appear to affect the genes adjacent to chiA-1 (Figs S3 and S4). Moreover, the distribution of putative chitinases (Fig. 2) suggested that isolates lacking ChiA-1 may encode additional
chitinases. Since ChiA-1 is known to have a functional role in *V. cholerae* chitin metabolism, this led us to speculate that these additional putative chitinases, if functional, might be able to provide chitinase activity in the absence of ChiA-1.

**Identification and characterization of chiA-3**

Eleven gene clusters in the pangenome included genes with the annotation ‘chitinase’ or ‘putative chitinase’ (Table S3). Five of these were found only in one genome, of which four were found only in the non-*V. cholerae* outgroup. Of the remaining six genes, four are known to be present in N16961 (Tables S2 and S3). On further examination, the products of one of the two gene clusters, ‘endo I_2’, were not predicted in silico to contain a chitinase domain, although a putative chitin-binding domain was identified (Fig. S5; Table S3). The second gene identified was predicted to encode a protein containing a chitinase domain (Fig. 3a). The molecular weight (47.69 kDa) and domain composition of the protein were distinct from those of ChiA-2 and ChiA-1 (Fig. 3a), as was the genomic context and location of the gene, which was inserted between VC_A0620 and VC_A0621 in the smaller *V. cholerae* chromosome. This genomic position is conserved in *Vibrio* species which harbour *chiA-3* orthologues [76, 79]. The figure was generated using Easyfig [61] and blastn comparisons [58].

In order to determine whether *chiA-3* had been identified previously in other *Vibrio* species, the gene was compared to the nine genes listed by Hunt et al. as chitinases found in non-cholera vibrios [20] (Table 2). The most similar protein to ChiA-3 (76.57% aa identity) was that encoded by VPA1177 (*chiA*, accession no. BAC62520.1), found in *V. parahaemolyticus* strain RIMD 2210633 (Table 2) [76]. VPA1177 encodes a 430 aa protein (47.98 kDa) which previous genetic analyses have shown to make a minimal contribution to the ability of *V. parahaemolyticus* to degrade chitin; ChiA-2 (encoded by VPA0055, accession no. BAC61398.1) is the major protein responsible for this phenotype in *V. parahaemolyticus* [77]. Transcription of VPA1177 has been shown to be significantly reduced in the presence of chitin [77]; however, the VPA1177 protein has been shown to be expressed by *V. parahaemolyticus*, albeit at very low levels in culture supernatants [36].

A previous report had also identified a functional secreted chitinase from *Vibrio harveyi* of a similar molecular weight (47 kDa) to both VPA1177 and ChiA-3 [78]. The *V. harveyi* ATCC 33843 genome [79] contains a gene encoding a putative chitinase (blastp: 100% query coverage, 77.73% amino acid identity to ChiA-3; predicted molecular weight 48.0 kDa) in
a similar genomic context on chromosome 2 to that of \textit{chiA-3} in NCTC 30 (Fig. 3b). This is distinct from the location of the functionally characterized \textit{chiA} gene (\textit{LA59_20935}) which encodes an 850 aa ChiA chitinase precursor (accession no. Q9AMP1 [37, 80, 81]), and from other functionally characterized \textit{V. harveyi} \(\beta\)-N-acetylglucosaminidases [38]. This \textit{V. harveyi} protein is also 90.9\% identical to VPA1177. As well as their high amino acid identity, each of these proteins was predicted to contain similar domain compositions and configurations across the three species (Fig. S6). It is reasonable to infer that these enzymes are orthologues of ChiA-3.

Since VPA1177 has been shown to be transcribed [77] and to produce a translated protein in \textit{V. paraemoloyticus} [36], we sought to determine whether the product of \textit{chiA-3} from \textit{V. cholerae} had chitinase activity. We amplified the gene from the genome of NCTC 30, a non-pandemic lineage \textit{V. cholerae}, and cloned it directionally into pBAD33 such that expression of the gene was governed by the arabinose-inducible P\textsubscript{BAD} promoter and the translated product linked to a C-terminal 6xHis tag, similar to previous reports [40, 45] (denoted pMJD157, Fig. S7).

\textit{E. coli} harbouring pMJD157 produced a His-tagged protein of the expected molecular weight that was retained in the cell pellet when cultured with arabinose at 23°C (Fig. 4). We used a commercial fluorogenic assay for chitinase activity which relies on the hydrolysis of 4-methylumbelliferyl (4-MU) chitin analogues to detect chitinase activity. A similar assay has been used previously to assay chitinase activity in vibrios [78]. We found that samples from \textit{E. coli} cultures expressing 6xHis-tagged ChiA-3 demonstrated statistically significant activity on 4-MU-linked substrates (Fig. 5). These data were consistent with the His-tagged protein detected in Fig. 4) (ChiA-3-6xHis) having endochitinase and chitobiosidase activities, but lacking \(\beta\)-N-acetylglucosaminidase activity. These results are also fully consistent with previous studies of \textit{V. cholerae} chitinase enzyme activity when expressed in \textit{E. coli} [48].

**DISCUSSION**

In this study, we present three major observations: first, that \textit{gbpA} is not ubiquitous amongst \textit{V. cholerae}; second, we show that there is additional variability in the chitinase genes harboured by diverse \textit{V. cholerae}, which show phylogenetic signals in their distribution; third, we present functional evidence that one of these putatively novel genes encodes a protein with chitinase activity.

The fact that \textit{gbpA} is absent from these bacteria is interesting to consider. As discussed previously, GbpA is an important factor in both environmental and pathogenic colonization. The fact that \textit{gbpA} is absent from non-pandemic \textit{V. cholerae} that appear to be basal to the rest of the species (Fig. 2) suggests that its role in pathogenicity may be more complex than previously thought. It might be
that acquisition of gbpA by *V. cholerae* was an important step in its evolution as a human pathogen. Conversely, because several of the isolates in the lineage lacking gbpA are of clinical as well as environmental origin [8, 65, 84–86], including some which were isolated from cases of acute or ‘choleraic’ diarrhoea [65, 84, 85], it might be that gbpA may not be essential for pathogenic colonization. It remains to be seen whether the natural absence of gbpA affects the ability of such *V. cholerae* to colonize both the intestinal mucosa and chitinous surfaces. The roles played by other adhesins in these diverse *V. cholerae*, such as MSHA, should also be considered in the future.

Fig. 5. ChiA-3-6xHis displays chitobiosidase and endochitinase activities, but not β-N-acetylglucosaminidase activity. Lysates and supernatants included in Fig. 4(a, b) were assayed for chitinase enzyme activity using a fluorometric chitinase assay kit (see Methods for details). Lysed cells from *Escherichia coli* cultures harbouring pMJD157 and cultured in the presence of arabinose were the only samples which produced detectable and statistically significant signals on triacetylchitotriose and chitobiose substrates (a, b). No signal was detected in the presence of glucosaminide substrate (c). All plots are scaled equivalently. P=pellet; S=supernatant; EV=empty vector (pBAD33). Parametric t-tests were performed where indicated, testing for statistically significant differences between cell pellets of both induced and non-induced cultures harbouring each plasmid (effect of arabinose induction on chitinase activity), as well as between pellets from induced cultures (effect of plasmid insert on activity). ns=not significant; ***P<0.001; ****P<0.0001.
Although ChiA-3 orthologues have been examined in other vibrios, we believe that this is the first report of this gene in *V. cholerae*, and the first report that the *V. cholerae* chiA-3 gene encodes a functional chitinase. The fact that chiA-3 was found only in non-pandemic *V. cholerae* is also intriguing. It is not yet known whether non-pandemic *V. cholerae* harbouring chiA-3 can respire chitin as effectively, or more effectively, than N16961 or other laboratory strains. A limitation of our study is that the genome sequences available to us are derived from isolates that are biased towards clinical cases of disease. Thus, our study may underestimate the diversity within the species [87], as well as the numbers of chitinases present in the species. Future research should include sequencing the genomes of both clinical and non-clinical *V. cholerae*, to address this implicit bias in genomic datasets.

There are fundamental differences between *V. cholerae* from pandemic and non-pandemic lineages, in terms of both their ability to cause cholera epidemics and their basic biology. We still do not fully understand these differences, but in order to do so, we must study *V. cholerae* pathogenicity in conjunction with more fundamental biological processes. It is currently unclear whether variation in the complements of chitinases and chitin-binding proteins encoded by *V. cholerae* have physiological consequences for different lineages of the species. However, given the importance of these genes to pathogenicity [30, 40], environmental lifestyles [24] and natural competence [29], it is plausible that these differences reflect differences in the ecological niches occupied by different lineages of the species. Research in this area may provide further insights into the genetic and biochemical differences between *V. cholerae* lineages that cause dramatically different patterns of disease worldwide. Collectively, these findings underline the fact that, as we continue to study diverse *V. cholerae*, our understanding of the nuance and specifics of this species will improve and be refined.

**References**


---

**Five reasons to publish your next article with a Microbiology Society journal**

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as ‘excellent’ or ‘very good’.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.