Ecological diversification reveals routes of pathogen emergence in endemic *Vibrio vulnificus* populations

Mario López-Pérez[a,b,c#], Jane M. Jayakumar[a,b#], Trudy-Ann Grant[a,b], Asier Zaragoza-Solas[c], Pedro J. Cabello-Yeves[c], and Salvador Almagro-Moreno[a,b*].

[a]Burnett School of Biomedical Sciences, College of Medicine, University of Central Florida, Orlando, Florida, USA; [b]National Center for Integrated Coastal Research, University of Central Florida, Orlando, Florida, USA; [c]Evolutionary Genomics Group, División de Microbiología, Universidad Miguel Hernández, Alicante, Spain.

[#]These authors contributed equally to this work

*Address correspondence to: Salvador Almagro-Moreno, samoreno@ucf.edu

Running title: Ecosystem perspective on pathogen emergence

Keywords:

Pathogen emergence / *Vibrio vulnificus* / aquatic pathogen / virulence evolution / population genomics / ecological divergence

ABSTRACT

Pathogen emergence is a complex phenomenon that, despite its public health relevance, remains poorly understood. *Vibrio vulnificus*, an emergent human pathogen, can cause a deadly septicaemia with over 50% mortality rate. To date, the ecological drivers that lead to the emergence of clinical strains and the unique genetic traits that allow these clones to colonize the human host remain mostly unknown. We recently surveyed a large estuary in eastern Florida, where outbreaks of the disease frequently occur, and found endemic populations of the bacterium. We established two sampling sites and observed strong correlations between location and pathogenic potential. One site is significantly enriched with strains that belong to one phylogenomic cluster (C1) from which the majority of clinical strains belong to. Interestingly, strains isolated from this site exhibit phenotypic traits associated with clinical outcomes, whereas strains from the second site belong to a cluster that rarely causes disease in humans (C2). Analyses of C1 genomes indicate unique genetic markers in the form of clinical-associated alleles with potential role in virulence. Finally, metagenomic and physicochemical analyses of the sampling sites indicate that this marked cluster distribution and genetic traits are strongly associated with distinct biotic and abiotic factors (e.g. salinity, nutrients, or biodiversity), revealing how ecosystems generate selective pressures that facilitate the emergence of specific strains with pathogenic potential in a population. This knowledge can be applied to assess the risk of pathogen emergence from environmental sources, and integrated towards the development of novel strategies for the prevention of future outbreaks.

SIGNIFICANCE STATEMENT

Our study addresses one main question: What are the ecological and genetic factors that drive pathogen emergence? To date, systematic experimental approaches to address this phenomenon are very limited. Here, we implemented a holistic approach to dissect the ecological, genetic and evolutionary drivers that foster the selection of virulence traits and pathogenic clones within an environmental population using *Vibrio vulnificus*, an aquatic bacterium that can cause a deadly septicemia in humans, as a model system. Our results suggest how ecosystems may generate selective pressures that facilitate the emergence of specific strains with pathogenic potential in a natural population and can be applied towards predictive frameworks to assess the risk of pathogen emergence from environmental sources.

INTRODUCTION

The emergence of human pathogens is one of the most concerning public health topics of modern times (1–4). According to the World Health Organization, over 300 emerging infectious diseases have been reported in the 1940-2004 period, a trend that has continued steadily with recent outbreaks of Ebola in West Africa, Cholera in Yemen, and the global pandemic caused by COVID-19 (3–5). Even though classical molecular approaches have advanced our understanding of bacterial pathogenesis, to date, the genetic adaptations and ecological drivers that facilitate selected strains within a species to emerge as pathogens and successfully colonize the human host remain poorly understood. Given the magnitude and complexity of this urgent threat, it is critical to develop tractable organismal model systems and theoretical frameworks that allow us to dissect the molecular adaptations and environmental factors that lead to the emergence of such human pathogens.

Vibrio vulnificus, an emergent human pathogen, is one of the leading causes of non-Cholera, Vibrio-associated deaths globally (6). Despite being a natural inhabitant of estuarine, coastal, and brackish waters (7), this flesh-eating bacterium has gained particular notoriety as one of the fastest killing pathogens (8, 9). Humans are typically infected with *V. vulnificus* through ingestion of contaminated raw seafood or by direct exposure of open wounds to seawater (6). *V. vulnificus* infections often result in fulminant septicaemia with an alarming mortality rate exceeding 50% (6, 10–13). The bacterium is particularly lethal in some susceptible hosts, such as immunocompromised patients or those with alcohol-associated liver cirrhosis, diabetes mellitus, or hemochromatosis (14). The annual case counts of *V. vulnificus* infections have steadily increased over the past 20 years in the USA (15). An upsurge in its worldwide distribution over the past three decades, in correlation with climate change, has led to disease outbreaks in regions with no history of *V. vulnificus* infections (16–18). Furthermore, models predict this trend to continue resulting in a steady expansion of its geographical range and the subsequent increased risk of human infections (16, 19–21).

89 Based on a series of biochemical and phenotypic traits, *V. vulnificus* strains have
90 been historically classified into three Biotypes (BT): BT1, mostly associated with human
91 infections (22, 23), BT2, primarily pathogenic to eels (24, 25), and BT3, which is
92 geographically restricted to Israel and possesses hybrid characteristics from BT1 and
93 BT2 (26, 27). In contrast to *Vibrio cholerae*, where all strains capable of causing cholera
94 belong to a single clade, genomic comparisons of *V. vulnificus* reveal a more complex
95 pattern in the distribution of its clinical strains (28–30). Phylogenomic analyses indicate
96 that the population of *V. vulnificus* is composed of four distinct groups or clusters (Cluster
97 1-4), which largely overlap with the classical Biotype classification system (23, 26, 28,
98 31, 32). Our analyses indicate that the two largest clusters, C1 and C2, exhibit high
99 genomic divergence and appear to be speciating (28), with clinical strains from BT1
100 predominantly belonging to C1 (22, 23), whereas strains from C2 primarily associated
101 with BT2 (6, 24, 25). C3 is highly clonal and fully overlaps with BT3, and the rare C4
102 contains only four non-clonal strains and belongs to BT1 (28, 31). Interestingly, despite
103 patients showing conserved clinical symptoms, C1 clinical strains arise from different
104 clades within the cluster, suggesting independent emergence events of this deadly
105 pathogen (28, 31, 32). To date, the unique genetic traits that allow certain C1 strains to
106 cause severe septicemia remain mostly unknown, posing a daunting public health risk
107 as it hinders our ability to detect potentially pathogenic *V. vulnificus* (33).
108 Recently, using a combination of bioinformatic and phenotypic analyses that
109 surveyed more than one hundred strains of *V. vulnificus*, we determined that *V. vulnificus*
110 C1 appears to be associated with a unique ecological lifestyle or ecotype (28).
111 Nonetheless, to date, the ecological drivers that lead to the emergence of clinical *V.*
112 *vulnificus* C1 and their pathogenic traits remain poorly understood. In order to start
113 untangling the complex *in-situ* interactions between genotypes and the environment that
114 underlie the emergence of clinical strains, in this study we recently surveyed a large
115 estuary in eastern Florida, the Indian River Lagoon (IRL), where outbreaks of the disease
116 frequently occur (7, 34). We found endemic populations of *V. vulnificus* in the estuary

and established two sampling locations to study the environmental dynamics of this bacterium in several natural reservoirs such as water, sediment, oysters and cyanobacteria. Interestingly, the two sampling sites show major differences in the distribution of *V. vulnificus* clusters. One of them, Feller's house (Site A), appears to be significantly enriched with C1 strains whereas in the second sampling site, Shepard Park (Site B), we mostly recovered strains from C2. Genomic analyses of these strains indicate that, despite these major differences in distribution, high recombination rates as well as frequent exchange of mobile genetic elements and virulence factors between these *V. vulnificus* populations occur. Microdiversity analyses of these genomes revealed unique genomic markers among C1 strains in the form of clinical-associated alleles with potential direct role in virulence. The isolated *V. vulnificus* strains are resistant to numerous commonly used antibiotics irrespective of cluster or site of isolation, however, phenotypic analyses indicate that strains from Site A exhibit traits associated with clinical outcomes, including the ability to resist serum and catabolize sialic acid, unlike those from Site B. Finally, metagenomic and physicochemical analyses of the sampling sites indicate that this marked cluster distribution is strongly associated with distinct biotic and abiotic factors (e. g. salinity, nutrients or biodiversity) revealing how ecosystems might generate selective pressures that facilitate the emergence of specific strains in a population as with pathogenic potential.

RESULTS AND DISCUSSION

Gene marker, *thiF,* can detect *V. vulnificus* and distinguish between clusters. Before initiating our sampling protocol, we looked for specific markers to rapidly screen environmental samples on a large scale. Specifically, we needed reliable genetic markers that could a) detect specifically *V. vulnificus*, b) accurately characterize them based on their cluster, and c) discriminate between clonal and non-clonal strains. The hemolysin gene *vvhA*, typically used to detect *V. vulnificus*, although species specific, is limited in its potential to distinctly classify strains into clusters or discriminate between

non-clonal strains (35–37) (Fig. 1A and Supplementary Fig. 1A). Other approaches, such as MLST, although effective in characterizing strains, require the PCR amplification, assembly, and concatenation of several housekeeping genes (32, 37–40), which is resource intensive and impractical for the rapid screening of *V. vulnificus* in environmental samples. In order to identify potential markers that meet all the requirements above, we compared all available C1 and C2 genomes in public databases through pangenome analysis. The number of gene families shared was 978 (accounting for ~22% average number of genes in a *V. vulnificus* genome), which we consider the core genome. We performed individual phylogenetic trees for these gene clusters and identified, across both chromosomes, a total of 47 genes that clearly differentiated C1 and C2 clusters. We compared these genes, using representative strains from all clusters (Table S1), based on percentage of sequence identity, against CMCP6, a reference *V. vulnificus* strain. We singled out the genes that had the highest percentage identity with strains from C1 but the least identity with those from C2 and viceversa, and compared them against *vvhA*. We finally selected a total of six candidate genes (*yycF*, *pfeS*, *acuB, yqhD*, *uvrY* and *thiF*), three from each chromosome, as potential markers (Table S2). Although all six candidate marker genes clearly differentiate C1 *V. vulnificus* strains from C2 (Fig. S1A), the response regulator *uvrY*, and the sulfur carrier protein adenylyltransferase *thiF*, had the maximum resolution in distinguishing all four clusters (C1-C4) as well as individual strains within each cluster, which serves as a proxy for discrimination of clonal populations (Fig. 1A). Additionally, the relative distances of the four clusters in the phylogenetic tree of *thiF* most accurately corresponds to the evolutionary tree of *V. vulnificus* built using single nucleotide polymorphisms and average nucleotide identities of all known *V. vulnificus* strains (28). Upon testing the species specificity of *thiF* with *Vibrio parahaemolyticus* RMID2210633 or *V. cholerae* O395, *thiF* was found to be specific to *V. vulnificus* (Fig. S1D). Thus, *thiF* has the potential to a) detect *V. vulnificus* strains, b) separate them by clusters, and c) discriminate between clonal and non-clonal strains based on their whole genome.

Furthermore, the concatenation of all six genes had at least twice the resolution and discriminatory power to differentiate all four clusters than *vvhA* making it an accurate set of genes for MLST analyses of *V. vulnificus* strains (Fig. 1A).

Detection of *V. vulnificus* along the Indian River Lagoon (IRL). The Indian River Lagoon (IRL, Easter Florida, USA) is one of the most biodiverse estuaries spanning an expansive geographic range with contrasting environments in Florida, where outbreaks of the disease frequently occur (Fig. 1B) (7, 34, 41). We recently surveyed this large estuary and we stablish two sampling sites at environmentally distinctive locations along the IRL (Fig. 1B). We collected samples in three sampling events (15-November-2018; 24-July-2019 and 22-August-2019) including biotic reservoirs such as oysters and cyanobacteria. *V. vulnificus* was isolated by sequential plating of the enriched populations on Chromoagar *Vibrio* (CaV) and TCBS as described in the Materials and Methods section (42). From a total of 1,856 colonies screened, only 245 were identified as potential *V. vulnificus* isolates based on the chromogenic plating method. An overall higher proportion of *V. vulnificus* was detected at Site B (Fig. S1B). At Site A, the distribution of *V. vulnificus* was found to be highest in oysters (45.3%) and water (43.4%), in contrast to sediments, which contain on average only 11.32% (Fig. S1B). Furthermore, a higher proportion of *V. vulnificus* was observed during the summer at both sites (Fig. S1B; 96.3% at Site A, 91.7% at Site B), likely as a consequence of increased water temperatures (>20ºC). The 245 potential *V. vulnificus* isolates were further confirmed using the novel gene marker *thiF*. PCR amplification of the *thiF* gene yielded 141 confirmed *V. vulnificus* isolates. We sequenced these PCR products and constructed a phylogenetic tree to determine cluster affiliation. To minimize further examination of strains of clonal origin that might have proliferated during enrichment, we only analyzed one strain within a group if a) the *thiF* alignment looked identical within the group, b) the strains came from the same replicate and fraction, and c) they were isolated during the same sampling event. As a result, 87 out of the 141 confirmed *V. vulnificus* isolates were

selected for further analyses (39 isolates from Site A and 48 from Site B) (Fig. 1C). Strikingly, phylogenetic analysis using gene marker *thiF* showed that most isolates from Site A, belong to C1 (97.4%, 38/39), whereas the majority of isolates from Site B belong to C2 (87.5%. 42/48) (Fig. 1C). This clear ecological separation between the two clusters provides an ideal framework to examine evolutionary processes underlying the emergence of pathogenic traits within a population and a platform to understand how ecosystems generate pressures that facilitate the selection of strains with pathogenic potential. In order to address this, we first dissect the genomic determinants and population structure of these environmental *V. vulnificus* strains, assess their pathogenic potential, and finally link these results with environmental factors (abiotic and biotic) associated with their marked cluster distribution.

Genomic determinants of *V. vulnificus* emergence. ***a) Ecological preferences of V. vulnificus clusters*.** To investigate the genomic determinants that potentially drive the ecological niche preferences of the clusters, we selected several strains for genome sequencing to obtain a proportionate representation of each cluster, reservoir, fraction, host and date of isolation. This resulted in a total of 27 *V. vulnificus* isolates sequenced (Table S3), 13 from Site A (2 sediment, 6 oyster, and 5 water isolates, one of which belonged to C2) and 14 from Site B (4 sediment, 4 cyanobacteria, and 6 water isolates including two from C1). For a robust phylogenomic association we included 74 dereplicated *V. vulnificus* genomes (e. g. genomes >99% Average Nucleotide Identity; ANI) currently available in public databases. We used both phylogenomic trees and ANI-based clustering of both chromosomes separately to group the genomes into the previously defined clusters (C1 to C4; (28)) (Figs. S2A and S3). Based on these results we decided to use the ANI of chromosome I as a reference for taxonomic classification since coverage is high (>70%), even among the most divergent clusters (C1 and C2). Interestingly, we found the first evidence of mixing or transfer of chromosomes between clusters of *V. vulnificus*. For instance, while chromosome I from FORC_037, an

environmental strain isolated from soft-shell clam, had an ANI > 98% with members of C2 and *ca* 95% with C1, for chromosome II was the other way around (Figs. S2A and S2B).

Whole genome phylogeny confirmed the marked differences in the distribution of *V. vulnificus* clusters obtained with *thiF* gene, corroborating the enrichment of C1 strains in Site A (Fig. 2A), except for the strain IRLE0015 that together with NV22 clustered closely to BT3 strains from the Israel outbreak (Fig. 2A). As aforementioned, we selected one non-clonal strain from Site A that belong to C2 (IRLA0043), and two from Site B belonging to C1 (IRLE0056 and IRLE0004). These gave us the opportunity to investigate the presence of potential genomic determinants specifically associated with each site, that is, whether C1 and C2 strains from site A have a unique pool of genes that is absent in strains from site B irrespective of cluster. The common part of the pangenome of all C2 strains from Site B was subtracted from the genome of the IRLA0043 strain, the only one in this cluster isolated from Site A. More than 500 genes were specific to this strain, apart from the capsule glycosylation genes we found a second cluster of genes (*rtxB-rtxD-rtxE*) encoding a type I secretion system (T1SS) with a high similarity (99%) to several strains of *V. coralliilyticus*. Specifically, this system appears to be associated with excretion of an enterotoxin (Efa-1/LifA) (43). Within these specific genes we also found a second type VI secretion system (T6SS) (28) and an Integrative conjugative element (ICE). On the other hand, C1 strains from Site B (IRLE0004 and IRLE0056) had only 200 unique genes compared to C1 strains from Site A. Among the specific genes of IRLE0004, we found a gene cluster conferring the ability to utilize tetrathionate as an electron acceptor, a common sulfur compound present in most soils (44), interestingly, this strain has been isolated from sediment. The ability to utilize tetrathionate has been associated with virulence in *Salmonella enterica* by providing a growth advantage to the bacterium in the inflamed gut (45). The functional annotation associated with the specific part of IRLE0056 was limited to the use of rhamnose, several toxin-antitoxin systems and the gene encoding the HipA involved in dormancy (46). Although it highlights the

257 presence in the environment of some virulence factors that can be easily shared between
258 the two clusters, our analysis did not identify any specific genomic determinants that may
259 explain the differential distribution of these strains.

260 ***b) Ecologically meaningful populations of V. vulnificus***. Despite the marked
261 environmental preferences and genomic divergence between C1 and C2 clusters, our
262 recent *in silico* studies indicate frequent exchange of mobile genetic elements (28). Here,
263 we have the opportunity to study potential recombination in natural *V. vulnificus*
264 populations in an endemic area. Recombination is particularly worrisome as novel
265 practices such as aquaculture can lead to the emergence of hybrid strains, as evidenced
266 by a deadly outbreak in Israel caused by an entirely new cluster (C3) (27) and the
267 presence of a C3-like strain isolated in this study (IRLE0015) (Fig. 2A). To evaluate this
268 phenomenon, we used a novel approach for assessing recent recombination events that
269 enables the delineation of ecologically relevant populations, i.e groups with the potential
270 to exchange genetic material (47). Our analyses revealed the presence of 15 major
271 recombining populations. Some of these populations coincide with the cluster
272 classification indicative of high intra-cluster recombination e. g. C3 and C4 (Fig. 2A).
273 However, C2 is made up of 12 populations. Eleven of them formed by a single member
274 and therefore indicating that there are no recombination events that connect these
275 strains with the rest of the cluster (48). Interestingly, all members of C1 form a single
276 population (P15) with the majority of C2 representatives indicating that, despite
277 divergence (*ca*. 95% ANI), these clusters are connected by recent recombination events
278 (Fig. 2A).

279 The capsular polysaccharide (CPS) cluster is an essential virulence factor of *V.*
280 *vulnificus* (49). Our previous analyses suggest that recombination may be a major
281 evolutionary mechanism leading to the high diversity of the CPS cluster (28). Thus, we
282 investigated the genomic diversity of the CPS between both clusters in these natural
283 populations. Strain IRLA0152 (C1) isolated from the free-living fraction at Site A, had a
284 similar variant of the CPS found in an infected patient isolate (FDAARGOS_119) (Fig.

285 2B). One of the hypervariable parts of the CPS from the oyster isolate OH0023 was
286 identical to that found in the reference clinical strain CMCP6, highlighting the
287 environment as a reservoir of these essential virulence genes (Fig. 2B). Furthermore,
288 certain CPS clusters are distributed in the population irrespective of cluster of origin and
289 sampling location. Specifically, we found the same CPS in one C1 strain from Site B
290 (IRLE0056) and three C2 strains, one of them from Site A (IRLE0043) and two from Site
291 B (IRLE0062 and IRLE0057) (Fig. 2B). The only variation was a small insertion in
292 IRLE0043 due to several IS elements, which suggests that this may be another
293 mechanism that can introduce variability within the CPS cluster (Fig. 2B). Overall, our
294 results indicate that despite the genomic divergence and their marked ecological
295 differences, there is a wide recombination among the clusters in an endemic area such
296 as the IRL including the transfer of major virulence factors within their natural
297 environment.

298 ***c) Pangenome analyses reveal genetic drivers associated with virulence***
299 ***emergence.*** The majority of clinical *V. vulnificus* strains belong to C1, similarly to most
300 strains isolated from Site A. To date, the specific genomic determinants that allow some
301 C1 strains to successfully colonise human remain mostly unknown. In order to elucidate
302 genetic factors associated with the emergence of clinical *V. vulnificus* C1 from
303 environmental gene pools and to determine whether C1 strains from Site A encoded
304 clinical associated traits, we compared genomes from strains isolated in this study
305 against those from *bona fide* clinical C1 and non-pathogenic strains (50, 51). Specifically,
306 we selected genomes from four distinct groups, a) nine C1 strains isolated from Site A
307 and b) nine C2 strains from Site B together with c) nine C1 strains that are *bona fide*
308 clinical i.e. isolated from patients with septicaemia, as well as d) nine non-pathogenic
309 strains from C2, i.e. isolated from environmental sources and susceptible to the
310 bactericidal effect of serum and monocytes (50, 51). Microbial species diversity was
311 analysed via a Partitioned PanGenome Graph Of Linked Neighbours (PPanGGOLiN,
312 (52)). The estimated size of the “persistent genome” (gene families present in almost all

313 genomes) is similar for each individual group as well as for all the groups combined
314 together, *ca*. 3,700 gene families (*ca*. 52% of the total genes families per genome). This
315 is quite remarkable given the genomic divergence between groups (Fig. 3A). The
316 proportion of gene families that formed the “shell genome” (genes families present in 3-
317 7 genomes) was only 1% of the total for both C1 groups and 2% for C2 groups. The
318 remaining gene families present in low frequency (1-3 genomes) were classified as the
319 "cloud genome” (Fig. 3A). As predicted, the percentage of gene families assigned to
320 functional categories (SEED subsystems database) for each pangenome partition varied
321 significantly: from 64% assigned to the persistent genome, to *ca*. 20% for the cloud and
322 shell. The latter being typically associated with diverse environmental adaptations
323 including pathogenesis, which highlights the enormous genomic plasticity that remains
324 to be addressed for these organisms.

325 Next, we compared the functional classifications of the gene coding sequences
326 from the persistent genomes of the nine reference C1 clinical strains against the nine C1
327 strains analysed from site A. We found that both groups only differ in ~2% of the total
328 gene content of their persistent genome. Most of these differences were associated with
329 the presence of genes belonging to the “Sialic Acid Metabolism” classification in the
330 clinical C1 strains (Fig. 3B). This group of genes code for a complete tripartite ATP-
331 independent periplasmic transport system (TRAP) involved in the transport of sialic acid,
332 for the enzymes responsible for its catabolism (N-acetylneuraminate lyase, N-
333 acetylmannosamine kinase and N-acetylmannosamine-6-phosphate 2-epimerase) as
334 well as a sialic acid mutarotase (YjhT family) and sialic acid utilization regulator, RpiR
335 family (53). The ability to scavenge, decorate their surface and utilize sialic acid as a
336 carbon source is an important virulence factor for pathogenic and opportunistic bacteria
337 including *V. vulnificus* (54–57). Using the C1 clinical reference genome CMCP6 we found
338 that the complete cluster was located in a genomic island on chromosome II (Fig. 3B).
339 The same gene cluster can be found in other Vibrio species (*ca*. 70% BLASTN identity)

such as *V. cholerae* O1, *Vibrio mimicus* or *Vibrio anguillarum*, however, unlike *V. vulnificus*, in these species the cluster was flanked by insertion sequence elements.

Given the frequent horizontal gene transfer in *V. vulnificus* populations it is unlikely that presence/absence of genes or gene clusters is sufficient to explain the emergence of virulence traits that lead to clinical outcomes in this pathogen. Our previous investigations with *V. cholerae* suggest that allelic variations of core genes can be major drivers of virulence emergence (29). Thus, we evaluated the patterns of microdiversity of the persistent genome by estimating the ratio of non-synonymous (dN) to synonymous (dS) substitution rates in pairwise genome comparison. We found six genes within the C1 clinical strains which showed a strong positive selection compared to the C1 IRL strains which on average exhibited a strong purifying selection (Fig. 3C and Table S4). In addition, average dN/dS values for these genes within C2 groups, both in the environmental references and the ones isolated from the IRL also exhibited very low dN/dS values (Table S4). The genes encoding these clinical-associated alleles (CAAs) differ between clinical strains and are involved in virulence associated processes and host related nutrient metabolism (Table S4). For instance, one of these genes encodes the outer membrane porin regulator OmpR, which regulates virulence in *V. cholerae* via *aphB* (58, 59). Another, encoding the subunit EntD, forms part of the enterobactin-synthetase enzyme complex, an iron acquisition system essential for virulence in *Escherichia coli* (60) and was proposed to play a role in the late stages of enterobactin biosynthesis in *V. cholerae* (61). The endonuclease *vvn*, identified as a periplasmic nuclease in *V. vulnificus*, prevents uptake of foreign DNA (62), thus hindering introduction of plasmids by transformation. Riboflavin synthase, *ribE*, catalyses the final step in the biosynthesis of riboflavin or vitamin B2. Riboflavin is involved in a number of metabolic pathways e.g. iron bioavailability and acquisition (63) in many pathogens including *V. cholerae*. Pyridoxal phosphate, PdxA, the catalytically active form of vitamin B_6, is an important cofactor for many enzymatic pathways involving breakdown of amino acids (64) and the sulfur transfer complex TusBCD TusB component. On average these

genes had lower dN/dS values in the C1 IRL strains in comparison to clinical C1, however, given that clinical *V. vulnificus* are endemic to this area, it is possible that some individual C1 IRL strains encode CAAs. To determine this, we analysed their presence by identifying individual allelic variants that deviate from the average values (Fig. 3D). Interestingly, even though none of the alleles from C1 IRL stains were identical to those found in the clinical strains, each of them encoded at least one gene with a dN/dS above the average. Those ranged from strain OH0003 encoding one (*tus*B gene) to IRLA0186 that encodes four of them (*omp*R, *rib*E, *ent*D and *pdx*A) (Fig. 3D). Overall, our results demonstrate that a) clinical strains encode unique CAAs, and b) allelic variants of these genes circulate in natural populations.

Assessment of pathogenic potential of *V. vulnificus* strains. In order to evaluate the pathogenic potential of IRL environmental strains and their association with phylogeny and location, we phenotypically tested their a) antibiotic resistance profile, b) survival in the presence of human serum, and c) ability to use sialic acid as a sole carbon source. For these assays, we included *V. vulnificus* CMCP6 (clinical C1) and *V. vulnificus* SS108-A3A (environmental non-pathogenic C2) as *bona fide* reference strains. Furthermore, we constructed three isogenic mutant strains in the background of *V. vulnificus* CMCP6 where we deleted the genes encoding: a) the CPS transport protein Wza (Δ*wza*), which has been shown to play a role in serum survival and capsule production (65), b) N-acetylneuraminate lyase (Δ*nanA*), first enzyme in the catabolic pathway of sialic acid (54), and the c) sialic acid TRAP transporter large permease (Δ*siaM*), which is associated with sialic acid uptake and is also involved in serum resistance (66).

a) Antibiotic resistance. First, we examined the antibiotic resistance profile of the IRL strains to determine whether there were patterns associated with the differential distribution of the clusters, as both sites have vastly different exposure to manmade perturbances including antibiotics (67, 68). We tested several antibiotics recommended by the Centres for Disease Control and Prevention for the treatment of Vibrio *spp*. (69).

While *V. vulnificus* CMCP6 showed resistance or intermediate resistance to virtually all the antibiotics tested (Fig. 4A), Δ*wza*, Δ*nanA*, and Δ*siaM* showed increased sensitivity to several of them compared to the wild-type (Fig. 4A). The capsule typically confers resistance to antibiotics (70, 71), however, the mechanisms by which sialic acid catabolism and uptake are involved in antibiotic resistance remains to be elucidated. Most IRL strains are resistant to polymyxin B, gentamycin, sulfadiazine and imipenem, a β-lactam antibiotic. In contrast, virtually no IRL strain was resistant to chloramphenicol or oxytetracycline (Fig. 4A). Seven strains from Site B exhibited intermediate resistance to nalidixic acid and/or trimethoprim while only two of the isolates from Site A were resistant to these compounds. Strikingly, a C1 strain isolated from Site B (IRLE0004), showed varied resistance levels to all antibiotics tested with the exception of oxytetracycline. Interestingly, two C1 strains from Site A (IRLA0161 and IRLA0152) that belonged to the same clonal frame i.e ANI>99%, showed different antibiotic resistance patterns (Fig. 4A). Unlike IRLA0152, IRLA0161 is resistant to oxytetracycline, nalidixic acid and trimethoprim. Genome analysis showed the presence of a 172 Kb plasmid in this strain, in which we identified a coding gene for a trimethoprim-resistant dihydrofolate reductase, DfrA family. Although the genes directly responsible for the other two resistances were not identified, we found several genes related to efflux pumps encoded in the same plasmid. It appears, from our analysis, that selective pressures at Site B, the site with most anthropogenic exposure, favour the emergence of antibiotic resistance, particularly to the folate inhibitor, trimethoprim, and the quinolone, nalidixic acid (Fig. 4A). Furthermore, the presence of resistant plasmids and their ease of transmission between the two clusters (28), increases the likelihood that strains from C1 to acquire these genes through horizontal gene transfer.

b) Serum resistance. Some studies have previously reported the ability of clinical *V. vulnificus* strains to resist the bactericidal effect of serum, while most environmental strains tested being susceptible to it (50, 51). Given that serum resistance is an essential virulence trait for *V. vulnificus* pathogenesis, we analysed the

susceptibility of the IRL isolates to this primary host defence. As expected, the wild-type clinical C1 strain was resistant to serum, whereas the non-pathogenic C2 strain was sensitive to its bactericidal effect (3-4 log decreases in CFUs) (Fig. 4B). Only three out of twelve strains from Site A were sensitive to serum whereas in Site B we found the opposite pattern, with most of the strains (eight out of fourteen) being sensitive (Fig. 4B). These differences were strongly associated with cluster distribution and provided us with an opening to examine the possible genomic determinants that lead to serum resistance in *V. vulnificus*. We first compared the gene content between serum resistant C1 strains (OH0023 and IRLA0152) against sensitive ones (OH0012 and IRLA0153). Among those unique genes in the resistant strains we found several related to type I restriction-modification systems, capsule synthesis and those involved in sialic acid metabolism. Subsequently, we analysed the presence of the sialic acid cluster in the genomes of all IRL isolates in our study. We found that 12 out of 15 strains that were resistant to serum (8 Site A; 4 Site B) encoded the cluster, whereas only 1 out of 11 sensitive strains did (Fig. 4B). Given this clear association, we tested the serum resistance of Δ*wza* and the two sialic acid mutants, Δ*nanA* and Δ*siaM*. As expected, Δ*wza* was sensitive to serum. Interestingly, while Δ*siaM* exhibited a 2-log decrease in CFU compared to the wild-type, Δ*nanA* was not affected by the bactericidal effect of serum, the mechanism behind the difference in survival between these two mutants remains to be addressed.

c) Sialic acid catabolism. Sialic acid, besides playing an important role in host-pathogen interactions (54, 56) is critical for the interaction of several pathogenic Vibrios with some of their environmental reservoirs such as Cyanobacteria potentially linking different lifestyles of bacterial pathogens (72, 73). Both our pangenome and phenotypic analyses suggest that catabolism of this aminosugar appears to be an essential factor associated with clinical outcomes. In order to initially test our findings, we examined the ability of the IRL strains to utilize N-acetylneuraminic acid (NANA) as a sole carbon source. We tested their growth in M9 minimal media supplemented with NANA at two salinities reflective of the two sampling sites (1% and 3% NaCl; Table S5). Neither the

452 Δ*nanA* and Δ*siaM* mutants nor the IRL isolates that did not encode the sialic acid cluster
453 were able to grow in these media. All strains from Site A that possessed the sialic acid
454 cluster (eight of the twelve) exhibited similar growth patterns to the clinical reference
455 CMCP6 at both salinities. At Site B, only six of the fourteen isolates were able to grow,
456 all containing the sialic acid cluster (Fig. 4C).
457 Taken together, our genomic and phenotypic analyses of the IRL strains, and
458 their comparisons against clinical strains, showed differential potential for pathogen
459 emergence in these natural populations. For instance, strain IRLA0186 exhibits several
460 traits that indicate its strong capability for emergence as a clinical strain such as its ability
461 to resist serum, catabolize sialic acid, resistance to most of the antibiotics tested, as well
462 as encoding variations in four of the six CAAs. On the other hand, OH0008, isolated from
463 the same site IRLA0186 (ANI 98,3%) is sensitive to both serum and most of the
464 antibiotics we tested, but cannot grow on sialic acid and only encodes one allelic variation
465 similar to CAAs, suggesting limited likelihood of pathogenic outcomes.
466
467 **Environmental factors associated with cluster divergence.** Our analyses revealed
468 distinct genomic and phenotypic signatures associated with the emergence of clinical-
469 associated traits in environmental *V. vulnificus*. In order to uncover ecological drivers
470 leading to the selection of these traits and the skewed distribution of *V. vulnificus*
471 clusters, we investigated the abiotic and biotic parameters associated with each site.
472 First, we measured several abiotic factors from the aquatic samples collected during
473 strain isolation such as temperature, dissolved oxygen, pH, dissolved organic matter,
474 salinity, phosphorous, among others (Table S5). Next, water samples were sequentially
475 filtered through 20, 5, and 0.22 μm pore size filters. DNA was obtained from the 0.22μm
476 filter that contain the free-living microbial fraction to analyse the microbial community
477 structure (biotic factors) associated with each sampling site (Fig. 5A). We used a
478 Principle Coordinate Analysis (PCoA) to examine possible correlations between cluster
479 distribution and both abiotic (physicochemical parameters) and biotic factors (taxonomic

480 classification from 16S rRNA gene metagenomic fragments) (Fig. 5B). The community
481 structure from Site A is very similar to that found in marine environments where the main
482 taxa were Cyanobacteria, SAR11, Bacteriodetes, Oceanospirillales or *Ca*. Actinomarina
483 (Fig. 5A). In fact, salinity at this location was 29 ppm, slightly lower than seawater (35
484 ppm) (Table S5). The percentage of 16S rRNA reads associated with the genus Vibrio
485 accounted for a total of 1.8% of the total population (Fig. 5A). However, they are
486 undetectable at Site B, where the salinity was much lower than in Site A (5 to 18 ppm),
487 signatures of a brackish environment. We also found in Site B higher concentrations of
488 phosphates, nitrates and dissolved organic matter compared to Site A likely due to
489 runoffs from nearby Lake Okeechobee, which experiences influx of fertilizers from
490 nearby agricultural farms (Table S5). These variations in environmental factors likely
491 change the microbial community by predominantly low-salinity adapted microbes such
492 as the genera Polynucleobacter and Limnohabitans within the family Burkholderiales or
493 the Microtrichal and Frankial families within the order Actinobacteria (Fig. 5A). Microbial
494 diversity, measured as Shannon index, indicated that diversity was higher in Site A than
495 in Site B (Fig. 5C). These data suggest that C1 members prefer a more oligotrophic
496 marine-like environment with higher salinity and greater microbial diversity dominated by
497 cyanobacteria, whereas C2 members appear to be better adapted to nutrient-rich
498 brackish environments marked by the presence of several families of Actinobacteria (Fig.
499 5). Overall, our metagenomic and physicochemical analyses of the sampling sites
500 indicate that the marked cluster distribution and genetic traits are strongly associated
501 with distinct biotic and abiotic factors (e. g. salinity, nutrients or biodiversity) revealing
502 how ecosystems generate selective pressures that facilitate the emergence of specific
503 strains with pathogenic potential in a population.
504
505 **CONCLUSIONS**
506 Elucidating the factors associated with the emergence and spread of human pathogens
507 is critical in order to develop tools to predict potential sources of disease outbreaks and

to establish effective surveillance strategies. Pathogen emergence is a complex and multifactorial phenomenon that requires analytic methods and tools that can consider large and highly diverse data. Therefore, it is essential to develop tractable model systems that allow us to dissect the ecological, genetic and evolutionary drivers that foster the selection of virulence traits and pathogenic clones within an environmental population. In this study, we used *V. vulnificus*, an emerging coastal pathogen that causes fatal sepsis, as a model system to investigate the genetic and ecological forces leading to pathogen emergence. The high genome plasticity of *V. vulnificus* paired with the unexpected outcomes associated with manmade environmental changes make this bacterium a major threat to human health for which no effective vaccines or therapeutic strategies are available (16, 28, 74). Here, we implemented a holistic approach that combines fields such as genomics, metagenomics, ecology, molecular biology and bacterial pathogenesis to address this problem. Overall, we found a strong correlation between ecological factors (e.g. site of isolation, physicochemical parameters and community structure) and pathogenic potential, as exemplified by skewed cluster distribution, and genetic and phenotypic traits associated with clinical outcomes.

The layers of selection imposed by the different abiotic and biotic factors likely act as a major selective pressure driving the development of pathogenic features in *V. vulnificus* populations. From our analyses, there is a clear association between cluster distribution and abiotic (e.g. salinity or dissolved nutrients) and biotic factors (community structure, oysters or cyanobacteria). Give their relevance, investigating the association of *V. vulnificus* and the specific role of these and other abiotic factors and biotic reservoirs such as protists (e.g. amoeba) and other metazoans (e.g. fish and crustaceans) in cluster selection, will shed substantial light on the process of emergence of pathogenic traits in *V. vulnificus*.

Furthermore, each sampling site is exposed to different anthropogenic influences. For instance, Site A is located in a protected area with limited access in Cape Canaveral. Whereas Site B experiences nutrient over-enrichment due to urbanization

and agricultural expansion, as well as, other manmade contamination such as faecal waste discharges. Given the drastic differences in the anthropogenic exposure between the two locations, it is likely that they play a role in cluster selection and distribution. It would be of interest for future studies to address the role of these anthropogenic disturbances in the emergence of pathogenic Vibrios.

Overall, our results indicate how ecosystems may generate selective pressures that facilitate the emergence and selection of specific strains within a population with pathogenic potential. Our study closely aligns with the One Health initiative (75) by a) focusing on the connection between a disease agent and the environmental factors that lead to its emergence, and b) creating a combined approach to understand disease emergence from an integrated and tractable perspective. Our approach can serve to develop ecological and genetic markers for surveillance systems to predict sources of outbreaks or identify emergent human pathogens. Overall, we offer a general paradigm and methodology for studying and understanding disease emergence that can be naturally extended to other human pathogens.

MATERIALS AND METHODS

Strains and culture conditions. An extended version of the Material and Methods can be found as part of the Supplementary Material. Strains of *V. vulnificus* (Tables S1 and S3) were routinely cultured on Luria-Bertani (LB) agar plates supplemented with 2% NaCl (wt/vol; LB-2%), inoculated in LB-2% broth, and cultured for 16 hours aerobically at 37°C, unless otherwise specified. *V. vulnificus* strains CMCP6 and SS108-A3A were used as C1 clinical and C2 environmental controls, respectively, for all phenotypic assays. *E. coli* β2155, a diaminopimelic acid (DAP) auxotroph, was used for mutant construction and was cultured in LB supplemented with 0.3mM DAP (LB-DAP).

Sampling sites. Samples were collected at two environmentally distinctive locations along the IRL (Easter Florida, USA) in three sampling events. The first location, Fellers House Field Station (N28˚54'25.315"; W80˚49'15.017"; Northern IRL; **Site A**), is located

within the federally-protected Canaveral National Seashore. The second sampling site, Shepard Park, is located in Port St. Lucie (N27˚11'48.864"; W80˚15'33.172": Southern IRL; Site B), which due to urbanization and agricultural expansion, experiences nutrient over-enrichment leading to excessive macroalgal bloom (Fig. 1B) (76, 77).

Isolation of V. vulnificus from environmental sources. *Water samples*: *V. vulnificus* was isolated from water samples using a modified protocol from Huq et al (42). 500ml of each sample was filtered successively through 20 µm, 5 µm, and 0.2 µm membrane filters (Sterlitech) to separate planktonic and free-living fractions. The filters were suspended in Phosphate buffered saline, pH 7.5 (PBS), vortexed vigorously and cultured in alkaline peptone water (APW) overnight at 37°C. *Sediment samples***:** *V. vulnificus* was isolated from sediment using a modified protocol from Schuster et al (78). Samples were collected using a universal corer. Samples were suspended in PBS (1:1), homogenized and enriched in APW. *Oyster samples***:** Isolation of *V. vulnificus* from oysters was carried out by a protocol adopted and modified from the U.S. Food and Drug Administration's Bacteriological Analytical Manual for *Vibrio* (79). Briefly, oysters collected from Feller's House were washed to remove sediment or dirt. Each oyster was individually shucked, homogenized in 30 ml PBS using the SCILOGEX D160 Homogenizer (Connecticut, USA), and cultured in APW. *Cyanobacterial samples***:** Cyanobacteria collected from Shepard Park were pelleted, supernatant removed, and cultured in APW. All samples were collected in triplicate. Enriched cultures in APW from water, sediment, oyster, and cyanobacteria samples were serially diluted and plated on CHROMagar Vibrio (CaV; CHROMagar, Paris, France), a *Vibrio* spp. selective agar. Turquoise blue colonies were further screened on Thiosulfate Citrate Bile Salts Sucrose (TCBS; Sigma) agar plates on which *V. vulnificus* appear as green colonies. Colonies that appeared turquoise blue on CaV and green on TCBS were considered potential *V. vulnificus* isolates.

Verification of V. vulnificus isolates*.* Potential *V. vulnificus* IRL isolates were verified by PCR using primers for the *thiF* marker gene (Table S2). PCR products of isolates positive for *thiF* were sequenced (GENEWIZ, AT, GA) to determine cluster affiliation. A

number of diverse *V. vulnificus* isolates, from both clusters and from each of the environmental reservoirs, were selected for whole genome sequencing. *Genome Sequencing*: Libraries of whole genomes were prepared using the Nextera DNA Flex Library Prep Kit from Illumina, following the manufacturer's instructions, and sequenced using the Illumina iSeq100 Sequencing System. Sequenced genomes were analyzed using Illumina BaseSpace Sequence Hub. Reads obtained for each Biosample were assembled into contigs and scaffolds using the SPAdes Genome Assembler Version 3.9.0 and Velvet de novo Assembly Version 1.0.0.

***Assembly, gene prediction and annotation*.** Reads were trimmed using Trimmomatic v0.36 (80) and assembled de novo with SPAdes v3.11.1 (81). ORFs from the assembled contigs were predicted using Prodigal v2.6 (82). tRNA and rRNA genes were predicted using tRNAscan-SE v1.4 (83), ssu-align v0.1.1 (84) and meta-rna (85). Using DIAMOND (86) predicted protein were compared against the NCBI nr database, and against COG (87) and TIGFRAM (88) using HMMscan v3.1b2 (89) for taxonomic and functional annotation.

***Phylogenomic reconstructions*.** The assembled contigs were assigned a chromosome by comparison to this group of reference genomes using Blastn (90). Genes were predicted using Prodigal (82) and clustered using the software MMseqs (91). The resulting protein clusters that were present in all analyzed genomes were divided into two groups according to the chromosome they are encoded in, resulting in a group of 257 and 62 proteins for chromosomes 1 and 2, respectively. Protein clusters were then aligned with QuickProbs2 (92), trimmed with BGME (93) and concatenated. Finally, a phylogenetic tree was constructed using iqtree (94) with automatic model selection and 1,000 bootstrap replicates.

***Genomic pairwise comparisons*.** Reciprocal BLASTN and TBLASTXs searches between genomes were carried out, leading to the identification of regions of similarity, insertions, and rearrangements. Average nucleotide identity (ANI) and coverage between pairs of genomes were calculated using the PYANI software (95).

Pangenome and recombination analysis. To analyze the gene family prevalence across all genomes, we used the software PPanGGOLiN to divide the gene families into persistent/shell/cloud partitions (52). The genes constituted each partition were then annotated against the SEED subsystem database (96) using DIAMOND (86), keeping all matches with E < 0.001 and alignment length > 0.5 for both subject and query. Finally, dN/dS values for the different protein partitions were obtained using the Orthologr package in R (97). The PopCOGenT pipeline (47) was used to define the recombinant populations based on gene flow between the different sequenced genomes.

Mutant construction. In-frame deletions of genes of interest, *wza, nanA* and *siaM*, were constructed via homologous recombination (98) (Primer list can be found in Table S6). Briefly, two approx. 500 bp PCR fragments flanking the genes of interest were cloned into the *sacB*-counterselectable plasmid, pDS132, and electroporated into donor *E. coli* strain, β2155. The donor strains harbouring the knockout vectors were conjugated with wild-type *V. vulnificus* CMCP6 on LB-DAP and transconjugants were selected on LB-2% plates supplemented with chloramphenicol (Cm) (25 µg/ml). Cm^R exconjugant colonies were cultured in LB-2% without antibiotics, and serial dilutions were plated on LB-2% plates containing 10% (wt/vol) sucrose. Potential double-crossover deletion mutants were screened by PCR and putative deletions were confirmed by DNA sequencing.

Antibiotic resistance. *V. vulnificus* isolates were examined for susceptibilities to the antibiotics highlighted in Fig. 4 at the highest concentrations in the breakpoint concentration range recommended by Clinical and Laboratory Standards Institute in M45-A (99–102) (Supplementary Methods). Briefly, individual colonies of each strain were transferred sequentially using sterile toothpicks onto LB-2% plates supplemented with respective antibiotics and incubated at 37°C overnight. The diameter of the growth was measured and resistance was defined as growth of at least 2mm in the respective antibiotics. Strains exhibiting no growth were taken as sensitive, and any intermediate growth diameter was considered as intermediate resistance. Experiments were performed in three independent biological replicates.

Serum Resistance. *In vitro* serum survival assay was adapted from Bogard and Oliver (103). Briefly, overnight cells were sub-cultured in LB-2% to obtain log-phase cells at an OD of 0.15-0.25. Cells were then washed in PBS and inoculated at a 100-fold dilution into normal pooled human serum (Fisher Bioreagents, Fair Lawn, NJ, USA) and incubated at 37°C for 2 hours. Resistance to serum was assessed by comparing the CFU/ml before and after exposure to serum. Experiments were performed in three independent biological replicates.

Sialic acid catabolism. The ability to catabolize sialic acid by *V. vulnificus* isolates was assessed by growth in N-acetylneuraminic acid, the predominant form of sialic acid in human cells, as the sole carbon source (57). Briefly, overnight cultures of each strain were washed and resuspended in M9 minimal media, and a 100-fold dilution of cells were made in M9 minimal medium supplemented with N-acetylneuraminic acid (2 mg/ml) (Chem-Impex International, Wood Dale, IL). 200-µl aliquots of each sample were added per well to a 96-well microtiter plate and incubated at 37°C with shaking. Optical density at 595 nm (OD595) was measured every hour for 24 h using a Tecan Sunrise microplate reader (Tecan US, Durham, NC) and the results were evaluated using the Magellan plate reader software. Growth assays were performed in triplicate across three independent biological replicates.

Measurement of physicochemical parameters. Measurements of water temperature (ºC), salinity (g/L), dissolved oxygen (%), pH, pressure (mmHg), dissolved organic matter (QSU), chlorophyll-a (µg/L) and total algae (µg/L) were made during the isolations. The measurements were recorded using a YSI EX02 sonde deployed at the sites at the time of sampling that was calibrated within 24 hours prior to each sampling event. Water samples, collected in triplicates, were also examined for the concentration of phosphates (o-Phosphate–P, method 365.1), and nitrates (Nitrate–N, method 353.2; Ammonia–N, method 350.1) according to the standard protocols described by the USEPA (104, 105). Briefly, collected water samples filtered through a 0.2 µm membrane filter were acidified to a pH < 2 with double distilled H_2SO_4, and stored at 4°C until analysis. Samples were

analysed for nitrate + nitrite (NO3-), ammonium (NH4+), and ortho-phosphate (PO43-) on a SEAL AQ2 Automated Discrete Analyzer (Seal Analytical, Mequon, WI).

Metagenomic analysis. DNA extraction was performed from the 0.22 µm filter. Attached cells were disrupted using CTAB lysis buffer and glass beads followed by lysozyme treatment. The nucleic acids were then extracted using the phenol-chloroform extraction method (106). Metagenomes were sequenced using Illumina Hiseq-4000 (150 bp, paired-end read). To analyse the phylogenetic classification of the samples, candidate 16S rRNA gene sequences in the raw metagenomes were identified using USEARCH6 (107) (E-value < 10–5) against a database containing non-redundant 16S rRNA sequences downloaded from the RDP database (108). These sequences were then aligned to archaeal and bacterial 16S rRNA HMM models (109) using ssu-align to identify true sequences (84). Only hits to 16S rRNA sequences were then classified into a high-level taxon if the sequence identity was ≥80% and the alignment length ≥90 bp. Sequences failing these thresholds were discarded. Information on data availability can be found in the supplementary methods.

FIGURE LEGENDS

Fig. 1. Isolation of *Vibrio vulnificus* from Eastern Florida. A) Maximum likelihood phylogenetic tree of hemolysin gene, *vvhA*, Sulfur carrier protein adenylyltransferase, *thiF* and the concatenation of the six candidate genes (*yycF, pfeS, acuB, yqhD, uvrY* and *thiF*) for representative strains from all 4 cluster. Members of the same cluster (C1 to C4) are indicated with the same color. Trees are unrooted and drawn to scale. Branch lengths indicate number of substitutions per site. B) Map of Florida indicating the sampling sites: Feller's house and Shepard Park. C) Maximum likelihood phylogenetic tree of *V. vulnificus* isolates based on *thiF*. Branches containing members that belongs to the same cluster are indicated with the same color, green for C1 representatives and blue for C2. The names of the strains are colored in relation to the location from which they originate.

703 The colored circles represent where they were isolated from and the red stars represent
704 those strains that have been sequenced.
705 **Fig. 2**. **Phylogenomic and population structure of *V. vulnificus*.** A) Maximum
706 likelihood phylogenomic tree of *V. vulnificus* strains obtained in this study (highlighted in
707 red) together with all available reference genomes using core genome of chromosome
708 I. Branches containing members that belongs to the same cluster (C1 to C4) (ANI > 97%)
709 are indicated with the same color. The color chart of the circles of the plot indicates the
710 isolated source and the host of the corresponding strains. Gray box shows the 15
711 recombinant populations detected among all strains. The orange box highlights the
712 strains belonging to subpopulation 15. B) Schematic representation of the capsular
713 polysaccharide (CPS) genomic island. Color-coded arrows show locations of important
714 genomic features. Variable regions 1 and 2 are highlighted in blue and green,
715 respectively.
716 **Fig. 3**. **Pangenome analysis of *V. vulnificus* strains.** A) Pangenome analysis for
717 groups, i) nine C1 strains isolated from Site A (C1 IRL), ii) nine C2 strains from Site B
718 (C2 IRL) iii) nine reference C1 strains that are *bona fide* clinical (C1 Clinical) iv) nine
719 reference environmental strains from C2 (C2 environmental). The proportions of gene
720 families in the persistent, cloud and shell genome are highlighted in orange, green and
721 blue respectively. B) Schematic representation comparing the genomic island of the
722 gene cluster involve in sialic acid catabolism. C) Comparison of the ratio of
723 nonsynonymous to synonymous substitutions (dN/dS ratio) between reference clinical
724 strains and C1 strains isolated from Site A in the IRL D) Comparison of the dN/dS values
725 of each individual strain versus the rest in the C1 IRL group for genes encoding these
726 clinical-associated alleles (CAAs). Those with a value above the average have been
727 highlighted in red.
728 **Fig. 4. Assessment of pathogenic potential of *V. vulnificus* IRL isolates.** A) Patterns
729 of antibiotic resistance of 27 *V. vulnificus* isolates to commonly used 12 antibiotics. Red,
730 resistant; pink, intermediate resistance; white, sensitive. B) Serum resistance of *V.*

vulnificus exposed to normal pooled human serum for 2 hours and assessed for survival in terms of CFU/ml. Resistant strains, similar CFU/ml as input; sensitive strains, lower of CFU/ml than input; resistant and growth on serum, higher CFU/ml than input. C) Ability to catabolise sialic acid assessed by growth of *V. vulnificus* isolates in M9 minimal media supplemented with N-acetylneuraminic acid as the sole carbon source at salinities representing the two sampling locations. Growth was measured as a function of increased optical density (OD595) of the cultures overtime.

Fig. 5. Environmental factors associated with cluster divergence. A) Taxonomic classification based on 16S rRNA gene fragments (raw reads) of the different metagenomes obtained from seawater 0.22µm filter. Only those groups with abundance values larger than 1% in any of the metagenomes are shown. The size of the diameter of the circles indicates the percentage of the total reads for each taxon. B) Principle Coordinate Analysis (PCoA) between physicochemical parameters and abundance of the different taxon's based on 16S rRNA gene metagenomic fragments. C) Box-plots illustrating microbial community diversity measure using Shannon index.

ACKNOWLEDGEMENTS

We are thankful to the reviewers for their thoughtful comments and suggestions. We are grateful to Drs. Paul Gulig, E. Fidelma Boyd and Linda Walters for kindly providing *V. vulnificus* strains, plasmids and oyster samples. We thank Dr. Shibu Yooseph for critical reading of the manuscript. This article was funded with Startup funds from the Burnett School of Biomedical Sciences to SAM, a PhD fellowship from the Spanish Ministerio de Economía y Competitividad (BES-2017-079993) to AZS, and a Post-Doctoral Fellowship from Generalitat Valenciana (APOSTD/2019/009) to PJC-Y.

COMPETING INTERESTS

The authors declare no competing interests.

AUTHOR CONTRIBUTIONS

SAM conceived the study. JMJ and TAG collected samples, isolated *V. vulnificus* and analyzed physicochemical and ecological parameters. JMJ performed phenotypic assays. MLP, AZS and PJC-Y performed bioinformatic analyses. The manuscript was written by MLP, JMJ and SAM. All authors read and approved the final version.

REFERENCES

1. O. Faye, *et al.*, Molecular Evolution of Zika Virus during Its Emergence in the 20th Century. *PLoS Negl. Trop. Dis.* (2014) https:/doi.org/10.1371/journal.pntd.0002636.
2. E. M. H. Wellington, *et al.*, The role of the natural environment in the emergence of antibiotic resistance in Gram-negative bacteria. *Lancet Infect. Dis.* (2013) https:/doi.org/10.1016/S1473-3099(12)70317-1.
3. K. A. Alexander, *et al.*, What factors might have led to the emergence of ebola in West Africa? *PLoS Negl. Trop. Dis.* (2015) https:/doi.org/10.1371/journal.pntd.0003652.
4. S. Khan, *et al.*, The emergence of a novel coronavirus (SARS-CoV-2), their biology and therapeutic options. *J. Clin. Microbiol.* (2020).
5. F. Qadri, T. Islam, J. D. Clemens, Cholera in Yemen — An Old Foe Rearing Its Ugly Head. *N. Engl. J. Med.* (2017) https:/doi.org/10.1056/nejmp1712099.
6. K. E. Phillips, K. J. F. Satchell, Vibrio vulnificus: From Oyster Colonist to Human Pathogen. *PLoS Pathog.* **13** (2017).
7. S. P. Heng, *et al.*, Vibrio vulnificus: An environmental and clinical burden. *Front. Microbiol.* (2017) https:/doi.org/10.3389/fmicb.2017.00997.
8. M. K. Jones, J. D. Oliver, Vibrio vulnificus: Disease and pathogenesis. *Infect. Immun.* **77**, 1723–1733 (2009).
9. S. R. Rippey, Infectious diseases associated with molluscan shellfish

786 consumption. *Clin. Microbiol. Rev.* (1994) https:/doi.org/10.1128/CMR.7.4.419.
787 10. P. a Gulig, K. L. Bourdage, a M. Starks, Molecular Pathogenesis of Vibrio
788 vulnificus. *J. Microbiol.* **43 Spec No**, 118–131 (2005).
789 11. W. G. Hlady, K. C. Klontz, The epidemiology of Vibrio infections in Florida, 1981-
790 1993. *J. Infect. Dis.* (1996) https:/doi.org/10.1093/infdis/173.5.1176.
791 12. N. A. Daniels, Vibrio vulnificus oysters: Pearls and perils. *Clin. Infect. Dis.* (2011)
792 https:/doi.org/10.1093/cid/ciq251.
793 13. M. S. Strom, R. N. Paranjpye, Epidemiology and pathogenesis of Vibrio
794 vulnificus. *Microbes Infect.* (2000) https:/doi.org/10.1016/S1286-4579(00)00270-
795 7.
796 14. M. A. Horseman, S. Surani, A comprehensive review of Vibrio vulnificus: An
797 important cause of severe sepsis and skin and soft-tissue infection. *Int. J. Infect.*
798 *Dis.* **15** (2011).
799 15. C. Baker-Austin, J. D. Oliver, Vibrio vulnificus: new insights into a deadly
800 opportunistic pathogen. *Environ. Microbiol.* (2018) https:/doi.org/10.1111/1462-
801 2920.13955.
802 16. C. Baker-Austin, J. Trinanes, N. Gonzalez-Escalona, J. Martinez-Urtaza, Non-
803 Cholera Vibrios: The Microbial Barometer of Climate Change. *Trends Microbiol.*
804 **25**, 76–84 (2017).
805 17. C. W. Kaspar, M. L. Tamplin, Effects of temperature and salinity on the survival
806 of Vibrio vulnificus in seawater and shellfish. *Appl. Environ. Microbiol.* (1993)
807 https:/doi.org/10.1128/aem.59.8.2425-2429.1993.
808 18. S. Paz, N. Bisharat, E. Paz, O. Kidar, D. Cohen, Climate change and the
809 emergence of Vibrio vulnificus disease in Israel. *Environ. Res.* (2007)
810 https:/doi.org/10.1016/j.envres.2006.07.002.
811 19. L. Vezzulli, *et al.*, Climate influence on Vibrio and associated human diseases
812 during the past half-century in the coastal North Atlantic. *Proc. Natl. Acad. Sci.*
813 *U. S. A.* (2016) https:/doi.org/10.1073/pnas.1609157113.

20. R. Deeb, D. Tufford, G. I. Scott, J. G. Moore, K. Dow, Impact of Climate Change on Vibrio vulnificus Abundance and Exposure Risk. *Estuaries and Coasts* (2018) https:/doi.org/10.1007/s12237-018-0424-5.

21. E. K. Lipp, A. Huq, R. R. Colwell, Effects of global climate on infectious disease: The cholera model. *Clin. Microbiol. Rev.* (2002) https:/doi.org/10.1128/CMR.15.4.757-770.2002.

22. J. D. Oliver, The Biology of Vibrio vulnificus. *Microbiol. Spectr.* (2015) https:/doi.org/10.1128/microbiolspec.ve-0001-2014.

23. D. L. Tison, M. Nishibuchi, J. D. Greenwood, R. J. Seidler, Vibrio vulnificus biogroup 2: new biogroup pathogenic for eels. *Appl. Environ. Microbiol.* **44**, 640–646 (1982).

24. J. D. Oliver, Wound infections caused by Vibrio vulnificus and other marine bacteria. *Epidemiol. Infect.* **133**, 383–391 (2005).

25. C. Amaro, E. G. Biosca, Vibrio vulnificus biotype 2, pathogenic for eels, is also an opportunistic pathogen for humans. *Appl. Environ. Microbiol.* (1996) https:/doi.org/10.1128/aem.62.4.1454-1457.1996.

26. N. Bisharat, *et al.*, Clinical, epidemiological, and microbiological features of Vibrio vulnificus biogroup 3 causing outbreaks of wound infection and bacteraemia in Israel. *Lancet* (1999) https:/doi.org/10.1016/S0140-6736(99)02471-X.

27. R. Zaidenstein, *et al.*, Clinical characteristics and molecular subtyping of Vibrio vulnificus illnesses, Israel. *Emerg. Infect. Dis.* (2008) https:/doi.org/10.3201/eid1412.080499.

28. M. López-Pérez, *et al.*, Evolutionary model of cluster divergence of the emergent marine pathogen vibrio vulnificus: From genotype to ecotype. *MBio* **10** (2019).

29. B. J. Shapiro, I. Levade, G. Kovacikova, R. K. Taylor, S. Almagro-Moreno, Origins of pandemic Vibrio cholerae from environmental gene pools. *Nat. Microbiol.* **2** (2016).

30. S. N. Sakib, G. Reddi, S. Almagro-Moreno, Environmental role of pathogenic traits inVibrio cholerae. *J. Bacteriol.*, JB.00795-17 (2018).
31. F. J. Roig, *et al.*, Phylogeny of Vibrio vulnificus from the analysis of the core-genome: Implications for intra-species taxonomy. *Front. Microbiol.* **8** (2018).
32. N. Bisharat, *et al.*, Hybrid Vibrio vulnificus. *Emerg. Infect. Dis.* **11**, 30–35 (2005).
33. P. C. Thiaville, *et al.*, Genotype is correlated with but does not predict virulence of Vibrio vulnificus biotype 1 in subcutaneously inoculated, iron dextran-treated mice. *Infect. Immun.* (2011) https:/doi.org/10.1128/IAI.01031-10.
34. A. Newton, M. Kendall, D. J. Vugia, O. L. Henao, B. E. Mahon, Increasing rates of vibriosis in the United States, 1996-2010: Review of surveillance data from 2 systems. *Clin. Infect. Dis.* **54** (2012).
35. V. J. Harwood, J. P. Gandhi, A. C. Wright, Methods for isolation and confirmation of Vibrio vulnificus from oysters and environmental sources: A review. *J. Microbiol. Methods* (2004) https:/doi.org/10.1016/j.mimet.2004.08.001.
36. F. Han, B. Ge, Multiplex PCR assays for simultaneous detection and characterization of Vibrio vulnificus strains. *Lett. Appl. Microbiol.* (2010) https:/doi.org/10.1111/j.1472-765X.2010.02887.x.
37. E. Sanjuán, B. Fouz, J. D. Oliver, C. Amaro, Evaluation of genotypic and phenotypic methods to distinguish clinical from environmental vibrio vulnificus strains. *Appl. Environ. Microbiol.* (2009) https:/doi.org/10.1128/AEM.01594-08.
38. N. Bisharat, *et al.*, The evolution of genetic structure in the marine pathogen, Vibrio vulnificus. *Infect. Genet. Evol.* (2007) https:/doi.org/10.1016/j.meegid.2007.07.007.
39. H. J. Kim, J. C. Cho, Genotypic diversity and population structure of Vibrio vulnificus strains isolated in Taiwan and Korea as determined by multilocus sequence typing. *PLoS One* (2015) https:/doi.org/10.1371/journal.pone.0142657.
40. N. González-Escalona, B. Whitney, L. A. Jaykus, A. DePaola, Comparison of

direct genome restriction enzyme analysis and pulsed-field gel electrophoresis for typing of Vibrio vulnificus and their correspondence with multilocus sequence typing data. *Appl. Environ. Microbiol.* (2007) https:/doi.org/10.1128/AEM.00738-07.

41. D. J. Bradshaw, N. J. Dickens, J. H. Trefry, P. J. McCarthy, Defining the sediment prokaryotic communities of the Indian River Lagoon, FL, USA, an Estuary of National Significance. *PLoS One* (2020) https:/doi.org/10.1371/journal.pone.0236305.

42. A. Huq, *et al.*, Detection, isolation, and identification of Vibrio cholerae from the environment. *Curr. Protoc. Microbiol.* (2012) https:/doi.org/10.1002/9780471729259.mc06a05s.

43. J. M. A. Klapproth, The role of lymphostatin/EHEC factor for adherence-1 in the pathogenesis of gram negative infection. *Toxins (Basel).* (2010) https:/doi.org/10.3390/toxins2050954.

44. V. L. Barbosa-Jefferson, F. J. Zhao, S. P. McGrath, N. Magan, Thiosulphate and tetrathionate oxidation in arable soils. *Soil Biol. Biochem.* (1998) https:/doi.org/10.1016/S0038-0717(97)00177-6.

45. S. E. Winter, *et al.*, Gut inflammation provides a respiratory electron acceptor for Salmonella. *Nature* (2010) https:/doi.org/10.1038/nature09415.

46. S. B. Korch, T. M. Hill, Ectopic overexpression of wild-type and mutant hipA genes in Escherichia coli: Effects on macromolecular synthesis and persister formation. *J. Bacteriol.* (2006) https:/doi.org/10.1128/JB.01740-05.

47. P. Arevalo, D. VanInsberghe, J. Elsherbini, J. Gore, M. F. Polz, A Reverse Ecology Approach Based on a Biological Definition of Microbial Populations. *Cell* (2019) https:/doi.org/10.1016/j.cell.2019.06.033.

48. M. López-Pérez, J. M. Haro-Moreno, F. H. Coutinho, M. Martinez-Garcia, F. Rodriguez-Valera, The evolutionary success of the marine bacterium SAR11 analyzed through a metagenomic perspective. *mSystems* **5** (2020).

898 49. G. S. Pettis, A. S. Mukerji, Structure, function, and regulation of the essential
899 virulence factor capsular polysaccharide of vibrio vulnificus. *Int. J. Mol. Sci.*
900 (2020) https:/doi.org/10.3390/ijms21093259.
901 50. T. C. Williams, M. Ayrapetyan, H. Ryan, J. D. Oliver, Serum survival of Vibrio
902 vulnificus: Role of genotype, capsule, complement, clinical origin, and in Situ
903 incubation. *Pathogens* (2014) https:/doi.org/10.3390/pathogens3040822.
904 51. H. Y. Kim, M. Ayrapetyan, J. D. Oliver, Survival of Vibrio vulnificus genotypes in
905 male and female serum, and production of siderophores in human serum and
906 seawater. *Foodborne Pathog. Dis.* (2014) https:/doi.org/10.1089/fpd.2013.1581.
907 52. G. Gautreau, *et al.*, PPanGGOLiN: Depicting microbial diversity via a partitioned
908 pangenome graph. *PLoS Comput. Biol.* (2020)
909 https:/doi.org/10.1371/journal.pcbi.1007732.
910 53. S. Almagro-Moreno, E. F. Boyd, Sialic acid catabolism confers a competitive
911 advantage to pathogenic Vibrio cholerae in the mouse intestine. *Infect. Immun.*
912 **77**, 3807–3816 (2009).
913 54. S. Almagro-Moreno, E. F. Boyd, Insights into the evolution of sialic acid
914 catabolism among bacteria. *BMC Evol. Biol.* (2009) https:/doi.org/10.1186/1471-
915 2148-9-118.
916 55. V. Bouchet, *et al.*, Host-derived sialic acid is incorporated into Haemophilus
917 influenzae lipopolysaccharide and is a major virulence factor in experimental
918 otitis media. *Proc. Natl. Acad. Sci. U. S. A.* (2003)
919 https:/doi.org/10.1073/pnas.1432026100.
920 56. E. Severi, D. W. Hood, G. H. Thomas, Sialic acid utilization by bacterial
921 pathogens. *Microbiology* (2007) https:/doi.org/10.1099/mic.0.2007/009480-0.
922 57. J. B. Lubin, J. J. Kingston, N. Chowdhury, E. F. Boyd, Sialic acid catabolism and
923 transport gene clusters are lineage specific in Vibrio vulnificus. *Appl. Environ.*
924 *Microbiol.* (2012) https:/doi.org/10.1128/AEM.07395-11.
925 58. D. E. Kunkle, X. R. Bina, J. E. Bina, Vibrio cholerae ompr contributes to

virulence repression and fitness at alkaline ph. *Infect. Immun.* (2020) https:/doi.org/10.1128/IAI.00141-20.

59. D. E. Kunkle, T. F. Bina, X. R. Bina, J. E. Bina, Vibrio cholerae OmpR represses the ToxR regulon in response to membrane intercalating agents that are prevalent in the human gastrointestinal tract. *bioRxiv* (2019) https:/doi.org/10.1101/752626.

60. J. Liu, K. Duncan, C. T. Walsh, Nucleotide sequence of a cluster of Escherichia coli enterobactin biosynthesis genes: identification of entA and purification of its product 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase. *J. Bacteriol.* (1989) https:/doi.org/10.1128/jb.171.2.791-798.1989.

61. E. E. Wyckoff, J. A. Stoebner, K. E. Reed, S. M. Payne, Cloning of a Vibrio cholerae vibriobactin gene cluster: Identification of genes required for early steps in siderophore biosynthesis. *J. Bacteriol.* (1997) https:/doi.org/10.1128/jb.179.22.7055-7062.1997.

62. S. I. Wu, S. K. Lo, C. P. Shao, H. W. Tsai, L. I. Hor, Cloning and characterization of a periplasmic nuclease of Vibrio vulnificus and its role in preventing uptake of foreign DNA. *Appl. Environ. Microbiol.* (2001) https:/doi.org/10.1128/AEM.67.1.82-88.2001.

63. I. S. Cisternas, J. C. Salazar, V. A. García-Angulo, Overview on the bacterial iron-riboflavin metabolic axis. *Front. Microbiol.* (2018) https:/doi.org/10.3389/fimmu.2018.01478.

64. J. Sivaraman, *et al.*, Crystal Structure of Escherichia coli PdxA, an Enzyme Involved in the Pyridoxal Phosphate Biosynthesis Pathway. *J. Biol. Chem.* (2003) https:/doi.org/10.1074/jbc.M306344200.

65. A. Nakhamchik, C. Wilde, D. A. Rowe-Magnus, Identification of a wzy polymerase required for group IV capsular polysaccharide and lipopolysaccharide biosynthesis in Vibrio vulnificus. *Infect. Immun.* (2007) https:/doi.org/10.1128/IAI.00932-07.

66. E. Severi, *et al.*, Sialic acid transport in Haemophilus influenzae is essential for lipopolysaccharide sialylation and serum resistance and is dependent on a novel tripartite ATP-independent periplasmic transporter. *Mol. Microbiol.* (2005) https:/doi.org/10.1111/j.1365-2958.2005.04901.x.

67. B. P. Bougnom, L. J. V. Piddock, Wastewater for Urban Agriculture: A Significant Factor in Dissemination of Antibiotic Resistance. *Environ. Sci. Technol.* (2017) https:/doi.org/10.1021/acs.est.7b01852.

68. S. Bin Zaman, *et al.*, A Review on Antibiotic Resistance: Alarm Bells are Ringing. *Cureus* (2017) https:/doi.org/10.7759/cureus.1403.

69. K. S. Shaw, *et al.*, Antimicrobial susceptibility of Vibrio vulnificus and Vibrio parahaemolyticus recovered from recreational and commercial areas of Chesapeake Bay and Maryland Coastal Bays. *PLoS One* (2014) https:/doi.org/10.1371/journal.pone.0089616.

70. M. A. Campos, *et al.*, Capsule polysaccharide mediates bacterial resistance to antimicrobial peptides. *Infect. Immun.* (2004) https:/doi.org/10.1128/IAI.72.12.7107-7114.2004.

71. M. P. E. Slack, W. W. Nichols, Antibiotic penetration through bacterial capsules and exopolysaccharides. *J. Antimicrob. Chemother.* (1982) https:/doi.org/10.1093/jac/10.5.368.

72. G. Reddi, K. Pruss, K. L. Cottingham, R. K. Taylor, S. Almagro-Moreno, Catabolism of mucus components influences motility of vibrio cholerae in the presence of environmental reservoirs. *PLoS One* **13** (2018).

73. S. N. Sakib, G. Reddi, S. Almagro-Moreno, Environmental role of pathogenic traits in Vibrio cholerae. *J. Bacteriol.* (2018) https:/doi.org/10.1128/JB.00795-17.

74. C. Baker-Austin, *et al.*, Emerging Vibrio risk at high latitudes in response to ocean warming. *Nat. Clim. Chang.* **3**, 73–77 (2012).

75. J. S. Mackenzie, M. Jeggo, The one health approach-why is it so important? *Trop. Med. Infect. Dis.* (2019) https:/doi.org/10.3390/tropicalmed4020088.

76. G. C. Sigua, W. A. Tweedale, Watershed scale assessment of nitrogen and phosphorus loadings in the Indian River Lagoon basin, Florida. *J. Environ. Manage.* (2003) https:/doi.org/10.1016/S0301-4797(02)00220-7.

77. P. J. Barile, Widespread sewage pollution of the Indian River Lagoon system, Florida (USA) resolved by spatial analyses of macroalgal biogeochemistry. *Mar. Pollut. Bull.* (2018) https:/doi.org/10.1016/j.marpolbul.2018.01.046.

78. B. M. Schuster, *et al.*, Ecology and genetic structure of a northern temperate Vibrio cholerae population related to toxigenic isolates. *Appl. Environ. Microbiol.* (2011) https:/doi.org/10.1128/AEM.00378-11.

79. C. A. Kaysner, J. Angelo DePaola, Bacteriological Analytical Manual Chapter 9: Vibrio. *Adm. U.S. Food Drug* (2004).

80. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* (2014) https:/doi.org/10.1093/bioinformatics/btu170.

81. A. Bankevich, *et al.*, SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* (2012) https:/doi.org/10.1089/cmb.2012.0021.

82. D. Hyatt, *et al.*, Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).

83. T. M. Lowe, S. R. Eddy, TRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964 (1996).

84. E. P. Nawrocki, “Structural RNA Homology Search and Alignment using Covariance Models.” (2009).

85. Y. Huang, P. Gilna, W. Li, Identification of ribosomal RNA genes in metagenomic fragments. *Bioinformatics* **25**, 1338–1340 (2009).

86. B. Buchfink, C. Xie, D. H. Huson, Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).

87. R. L. Tatusov, *et al.*, The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* **29**, 22–8 (2001).

88. D. H. Haft, *et al.*, TIGRFAMs: a protein family resource for the functional identification of proteins. *Nucleic Acids Res.* **29**, 41–3 (2001).

89. S. R. Eddy, Accelerated profile HMM searches. *PLoS Comput. Biol.* (2011) https:/doi.org/10.1371/journal.pcbi.1002195.

90. S. F. Altschul, *et al.*, Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).

91. M. Mirdita, *et al.*, Uniclust databases of clustered and deeply annotated protein sequences and alignments. *Nucleic Acids Res.* (2017) https:/doi.org/10.1093/nar/gkw1081.

92. A. Gudyś, S. Deorowicz, QuickProbs 2: Towards rapid construction of high-quality alignments of large protein families. *Sci. Rep.* (2017) https:/doi.org/10.1038/srep41553.

93. A. Criscuolo, S. Gribaldo, BMGE (Block Mapping and Gathering with Entropy): A new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol. Biol.* (2010) https:/doi.org/10.1186/1471-2148-10-210.

94. L. T. Nguyen, H. A. Schmidt, A. Von Haeseler, B. Q. Minh, IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* (2015) https:/doi.org/10.1093/molbev/msu300.

95. L. Pritchard, R. H. Glover, S. Humphris, J. G. Elphinstone, I. K. Toth, Genomics and taxonomy in diagnostics for food security: Soft-rotting enterobacterial plant pathogens. *Anal. Methods* (2016) https:/doi.org/10.1039/c5ay02550h.

96. R. Overbeek, *et al.*, The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.* **33**, 5691–5702 (2005).

97. H. G. Drost, A. Gabel, I. Grosse, M. Quint, Evidence for active maintenance of phylotranscriptomic hourglass patterns in animal and plant embryogenesis. *Mol. Biol. Evol.* **32**, 1221–1231 (2015).

98. J. B. Lubin, *et al.*, Host-Like carbohydrates promote bloodstream survival of Vibrio vulnificus in vivo. *Infect. Immun.* **83**, 3126–3136 (2015).

99. C. Scarano, *et al.*, Antibiotic resistance of Vibrio species isolated from Sparus aurata reared in Italian mariculture. *New Microbiol.* (2014).

100. F. Han, R. D. Walker, M. E. Janes, W. Prinyawiwatkul, B. Ge, Antimicrobial susceptibilities of Vibrio parahaemolyticus and Vibrio vulnificus isolates from Louisiana Gulf and retail raw oysters. *Appl. Environ. Microbiol.* (2007) https:/doi.org/10.1128/AEM.01116-07.

101. C. Baker-Austin, *et al.*, Multi-site analysis reveals widespread antibiotic resistance in the marine pathogen Vibrio vulnificus. *Microb. Ecol.* (2009) https:/doi.org/10.1007/s00248-008-9413-8.

102. CLSI, *M45. Methods for Antimicrobial Dilution and Disk Susceptibility Testing of Infrequently Isolated or Fastidious Bacteria ; Proposed Guideline* (2015).

103. R. W. Bogard, J. D. Oliver, Role of iron in human serum resistance of the clinical and environmental Vibrio vulnificus genotypes. *Appl. Environ. Microbiol.* (2007) https:/doi.org/10.1128/AEM.01551-07.

104. Environmental Monitoring Systems Laboratory, W. McDaniel, Methods for the Determination of Metals in Environmental Samples. *Methods Determ. Met. Environ. Samples* (1996).

105. USEPA, Methods for the Determination of Metals in Environmental Sample, USEPA Office of Research and Development, 8/93. Cincinnati, OH, EPA 600/R-93/100s,. *1993* (2013).

106. J. Sambrook, W. Russel, D, Molecular Cloning, 3-Volume Set : A Laboratory Manual. *Cold Spring Harboc Lab. Press* (2000).

107. R. C. Edgar, Search and clustering orders of magnitude faster than BLAST.

Bioinformatics **26**, 2460–2461 (2010).

108. J. R. Cole, *et al.*, Ribosomal Database Project: Data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* (2014) https:/doi.org/10.1093/nar/gkt1244.

109. S. R. Eddy, Multiple alignment using hidden Markov models. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* (1995).