Coastal Microbiomes Reveal Associations between Pathogenic Vibrio Species, Environmental Factors, and Planktonic Communities

Rachel E. Diner
University of California, San Diego
Drishti Kaul
J Craig Venter Institute
Ariel Rabines
J. Craig Venter Institute
Hong Zheng
J. Craig Venter Institute
Joshua A Steele
Southern California Coastal Water Research Project
John F Griffith
Southern California Coastal Water Research Project
Andrew Ellis Allen (aallen@jcvi.org)
J Craig Venter Institute https://orcid.org/0000-0001-5911-6081

Research

Keywords: bacteria , emerging health, Vibrio interactions, eukaryotic and photosynthetic organism

Posted Date: November 4th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-100087/v1

Coastal microbiomes reveal associations between pathogenic *Vibrio* species, environmental factors, and planktonic communities

Running title: metabarcoding reveals vibrio-plankton associations

Rachel E. Diner[1,2], Drishti Kaul[2], Ariel Rabines[1,2], Hong Zheng[2], Joshua A. Steele[3], John F. Griffith[3], Andrew E. Allen[1,2*]

[1] Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92037, USA

[2] Microbial and Environmental Genomics Group, J. Craig Venter Institute, La Jolla, California 92037, USA

[3] Southern California Coastal Water Research Project, Costa Mesa, CA 92626, USA

* Correspondence: aallen@ucsd.edu

Author Emails: Rachel E. Diner: racheldiner@gmail.com, Drishti Kaul: dkaul@jcvi.org, Ariel Rabines: arabines@ucsd.edu, Hong Zheng: hzheng@jcvi.org, Joshua A. Steele: joshuas@sccwrp.org, John F. Griffith: johng@sccwrp.org, Andrew Allen: aallen@ucsd.edu

Abstract

Background

Many species of coastal *Vibrio* spp. bacteria can infect humans, representing an emerging health threat linked to increasing seawater temperatures. Vibrio interactions with the planktonic community impact coastal ecology and human infection potential. In particular, interactions with eukaryotic and photosynthetic organism may provide attachment substrate and critical nutrients (e.g. chitin, phytoplankton exudates) that facilitate the persistence, diversification, and spread of pathogenic *Vibrio* spp.. Vibrio interactions with these organisms in an environmental context are, however, poorly understood.

Results

After quantifying pathogenic *Vibrio* species, including *V. cholerae*, *V. parahaemolyticus*, and *V. vulnificus,* over one year at 5 sites, we found that all three species reached high abundances, particularly during Summer months, and exhibited species-specific temperature and salinity distributions. Using metabarcoding we established a detailed profile of the both prokaryotic and eukaryotic coastal microbial communities, finding that pathogenic species were frequently associated with specific ASVs of chitin-producing eukaryotes such as diatoms and copepods. Furthermore, environmental variables had a significant effect not only on pathogenic *Vibrio* species but entire microbial communities, suggesting in some cases shared environmental preferences. Several significant ASV-level associations were revealed, indicating that commonly used broad taxonomic classifications (e.g. based on microbial class or *Vibrio* as a genus) likely mask ecologically important interactions. Shotgun metagenomic analyses revealed diverse vibrio

communities that harbored additional potential vibrio pathogens, antibiotic resistance genes, and genes associated with virulence.

Conclusions

Taken together, this data shows that abundant pathogenic *Vibrio* species likely containing both antibiotic resistance and virulence-associated genes are associated with chitin producing organisms which could act as an attachment substrate, facilitating environmental persistence and horizontal gene transfer. Shared environmental conditions such as high temperatures were associated with both high levels of pathogenic vibrios and potential environmental reservoirs, which should be taken into consideration when modelling vibrio infection risk in the face of climate change and identifying biomarkers of pathogen species. Furthermore, ASV-level associations may be critical to understanding vibrio microbial ecology and should be taken into consideration while developing environmentally relevant laboratory model systems.

Background

Coastal bacterial *Vibrio* species can cause severe human infections, which are an emerging international health concern linked to rising global temperatures. *Vibrio cholerae*, the causative agent of the disease cholera, infects millions of people each year, killing thousands, and is typically spread through ingesting contaminated drinking water [1]. Two other species of major concern are *V. parahaemolyticus* and *V. vulnificus*, which can cause severe wound infections, septicemia, and gastroenteritis from ingesting vibrio-colonized seafood [2]. While many strains of these species are not harmful to humans, genes belonging to each species have been identified that are associated with increased virulence potential [3]. Several particularly dangerous pandemic strains have been

identified and innocuous strains may become virulent and/or antibiotic resistant via horizontal gene transfer as many infection-related genes are mobile [3, 4]. At least a dozen additional *Vibrio* species can infect humans or animals, extending the threat to aquaculture operations. Climate change may exacerbate the prevalence of these infections. Increasing air and water temperatures can facilitate increased metabolic growth capacity and temporal and geographic range expansion of *Vibrio* spp. pathogens [5–7]. Furthermore, *V. cholerae* epidemics have been linked to global temperature rise on decadal scales and are an important case study for understanding the link between the environment and human disease [8, 9].

Vibrio spp. interactions with the planktonic community have implications for both coastal ecology and human health. Coastal communities are highly productive environments; diverse and abundant populations of microbes and multicellular organisms are supported by primary productivity driven by ample nutrient availability. These communities are subject to frequent and extreme changes in environmental conditions, including fluctuations in temperature, salinity, and dissolved oxygen. *Vibrio* spp. attach to and form biofilms on particles and eukaryotic organisms, living and dead, [10–13], presumably to better acquire carbon and nutrients and avoid environmental stress. These “close quarters” incite competition and enable cooperation and horizontal gene transfer with co-occurring bacterial and eukaryotic species.

An important facet of these interactions involves chitin, an abundant polymer produced by many marine eukaryotes [14]. In addition to providing nutrients and an attachment substrate, chitin facilitates bacterial competition and horizontal gene transfer in *Vibrio* spp. [15, 16], which may spread virulence and antibiotic resistance genes among populations. Attachment also enables environmental persistence and dispersal; for example, *Vibrio* spp. attach to copepod exoskeletons by the thousands with high copepod abundances linked to cholerae epidemics. *Vibrio* spp. can also

attach to chitin-producing diatoms [17, 18], though the ecological relevance of these interactions is poorly understood.

Vibrio interactions with other organisms in their environment are likely species-specific. Pathogenic species possess unique functional traits and often occupy distinct environmental niches driven by temperature, salinity, and other biotic and abiotic factors (reviewed in [19]). Virulence mechanisms, while poorly understood, are also species dependent [20]. Despite this, total quantities of *Vibrio* spp. are frequently used to infer ecological associations and human health risks. Likewise, eukaryotes are often grouped into broad categories. For example, phytoplankton are often quantified and characterized based on bulk chlorophyll *a* concentrations or at broad taxonomic levels (e.g. diatoms, dinoflagellates). But physiological differences at lower taxonomic ranks may have functional consequences for interactions; for example, some diatom genera exude chitin while others may not [21, 22] and algae are known to host distinct bacterial communities [23]. Environmental drivers of these taxa also influence whether and how they might coexist with vibrios. Recent technological advances in *Vibrio* spp. quantification and microbial community characterization (i.e. next-generation sequencing) of both prokaryotes and eukaryotes now enable a deeper understanding of vibrio microbial ecology in the context of environmental drivers and biological interactions.

To address this important research gap we quantified the pathogenic *Vibrio* species *V. cholerae*, *V. parahaemolyticus*, and *V. vulnificus* for one year at five coastal sites in Southern California and used metabarcoding to characterize the co-occurring prokaryotic and eukaryotic communities. We then used shotgun metagenomic sequencing of vibrio bacterial isolates to identify genes of interest to human health (e.g. virulence and antibiotic resistance genes) and assess the diversity of potentially pathogenic *Vibrio* species. With this dataset we investigated the

influence of spatial, temporal, and environmental variation on the abundance and distribution of individual pathogenic *Vibrio* species and their co-occurring coastal microbiomes. Additionally, we examined the occurrence of important species- or amplicon sequence variant (ASV)- level associations, which might have been overlooked or mischaracterized by commonly used broad taxonomic groupings and whether or not current laboratory models for studying vibrio interactions with other organisms are ecologically relevant. Finally, we identified taxa and environmental conditions that are associated with *Vibrio* strains that possess genes of interest to human health (e.g. antibiotic resistance genes, virulence genes).

Results

Abundance, distribution, and environmental drivers of pathogenic Vibrio *species*

We quantified three pathogenic *Vibrio* species with high human health relevance: *Vibrio parahaemolyticus*, *V. vulnificus*, and *V. cholerae*. Digital droplet PCR (ddPCR) was performed to quantify single copy number genes specific to each species in known filtration volumes using previously designed qPCR primers (Additional File 1)[24–28]. A wide range of temperatures (13.2-33 °C), and salinities (2.6-42.4 ppt) were sampled (Figure 1E,F), with highly variable chlorophyll *a* concentrations (Figure 1G). Spearman's rank correlation analyses revealed salinity and chlorophyll *a* were positively associated with temperature (Figure 1H). Each of the three species was detected at each site during at least one sampling point (i.e. there was no site where any species went entirely undetected) and often co-occurred. (Figure 2A-C)

V. cholerae and *V. vulnificus* were associated with low salinities while *V. parahaemolyticus* was significantly associated with higher temperatures, though all target species were only abundant above 20 °C (Figure 2A-D). *V. parahaemolyticus* was detected in 80% of samples (Additional File

2) and was most abundant at warm temperatures and high salinities (Figure 2C), though only the association with temperature was significant ($p < 0.05$)(Figure 2D). We observed *V. parahaemolyticus* at extremely high salinities (> 40 ppt, Figure 2C), while *V. vulnificus* and *V. cholerae* were both significantly associated with low salinity but not temperature (Figure 2A,B). *V. vulnificus* was most abundant at moderate (relative to the observed range of our study) and *V. cholerae* at low salinity sites (Figure 2A,B). High numbers of *V. cholerae* (>280,000 copies/ 100 mL) were detected at Los Peñasquitos Lagoon (LPL) from March through May, corresponding with low salinity caused by lagoon closure and subsequent urban freshwater accumulation (Figure 1F, Figure 2A). All three species peaked between March and July at the LPL and San Diego River (SDR) sites. At the Tijuana River Estuary (TJ) sites, *V. parahaemolyticus* was predominantly detected, occurring between February and October and peaking at >33,000 copies/ 100 mL in September (Additional File 2).

Vibrio genes relevant to human health

We investigated genes related to human health using digital droplet PCR (ddPCR) and shotgun sequencing of vibrio-enriched communities isolated on CHROMagar Vibrio plates (CHROMagar, Additional File 3). We used ddPCR to quantify *Vibrio vulnificus* virulence-associated genes in all samples where this species was detected. Additionally, we conducted shotgun metagenomic sequencing for 23 samples collected during the months of February, March, May, July, and August to investigate additional genes associated with virulence and antibiotic resistance.

V. vulnificus was detected in 15/60 samples (Additional File 2), and we quantified the virulence-associated genes *pilF* and *vcgC* in these samples using ddPCR (Additional File 1, Figure

2E,F). *pilF* is a protein required for pilus-type IV assembly, and the particular polymorphism detected our study and in prior qPCR-based studies is strongly associated with *V. vulnificus* human pathogenicity [25]. The *vcgC* sequence quantified in our study was derived from a randomly amplified polymorphic DNA (RAPD) PCR amplicon associated with clinical isolates of *V. vulnificus*, which differs from the *vcgE* variant primarily found in environmental samples, and is commonly used to inform whether a *V. vulnificus* strain has pathogenicity potential [29]. Thus, detection of this gene in our study samples may be indicative that certain strains in the population may have the capacity to infect humans. Both were most abundant at the SDR sites, with *pilF* reaching >7,000 copies/ 100 mL (Figures 2E,F). When *V. vulnificus* was detected, 50% of samples also tested positive for one or both of the virulence-associated genes (Figure 2G). The ratio of these targets to *V. vulnificus* copies (potentially reflective of number of virulence-associated gene copies per *V. vulnificus* cell) was often below 1, though sometimes closer to 2 as in the LPL April sample (Figure 2H). We attempted to quantify the *V. parahaemolyticus* virulence-associated thermostable direct hemolysin gene (*tdh*) and TDH-related hemolyisn gene (*trh*) using standard PCR and ddPCR. These attempts were unsuccessful, however, we were able to collect data on the abundance and distribution of these genes using shotgun sequencing.

For the 23 samples with shotgun metagenomic sequencing data, we identified and compared genes of interest to human health in and among the vibrio isolate communities. First, we identified genes associated with virulence in the three target vibrio pathogens from the literature and quantified these in each sample based on mapping to a global co-assembly. We also utilized the Resistance Genome Identifier (RGI) against the Comprehensive Antibiotic Resistance Database [30] to identify and quantify antibiotic resistance genes in each sample.

Shotgun metagenomic analysis revealed that several additional genes known or suspected to be associated with virulence were present among the vibrio populations (Figure 3A). To additionally validate the ddPCR data, we confirmed that all 3 species-specific target genes (*toxR, ompW,* and *vvhA)* were present in the shotgun metagenomics dataset (Figure 3A). In *V. cholerae*, the *ctxA* gene [31], which is responsible for producing cholera toxin and causing the disease cholera, was not detected in the samples. The accessory cholera enterotoxin (*Ace*)[32] and zona occludens toxin (*zot*)[33] genes were also undetected, as was *tcpA* [34], which is required for *V. cholerae* host colonization. *rtxA* genes, which are thought to be associated with virulence but are not well understood [35], were detected in several samples, including those with low or undetected quantities of *V. cholerae* as was *vgrG*, which is required for T6SS-dependent cytotoxic effects of *V. cholerae* on eukaryotic cells [36]. The *V. parahaemolyticus* virulence-associated genes *trh* and *tdh*, which have high sequence homology and are combined in this analysis, were detected in many samples in the metagenomic data, with the highest normalized abundance in the LPL March sample (Figure 3A). The gene *tlh,* which is also proposed to play a role in virulence, was detected in all metagenomic samples with similar abundances to the presumed host marker gene *toxR*. The *V. vulnificus* gene *GbpA,* responsible for production of the *N*-Acetylglucosamine (GlcNAc) binding protein A, is expressed more highly in clinical *V ulnificus* isolates, and was found in our samples but at levels typically lower than the *vvhA* host marker gene, suggesting not all members of the population carry it. The vcgC/vcgE gene was also detected in the shotgun sequences, also in lower abundance than the host marker gene, though due to the very close identity of the two sequences the specific variants were combined in this analysis.

Several classes of antibiotic resistance genes were detected in the sequenced isolates. Among the most abundant across all samples were genes associated with resistance to glycopeptide

and aminoglycoside antibiotics, tetracycline, and proteins involved in resistance to both cephalosporin and penam antibiotics (Figure 3B). Many detected genes were associated with transport and regulation of multiple drugs. Among the samples with high concentrations of pathogenic vibrio bacteria, as quantified by ddPCR (Additional File 2, appearing in bold on the axis of Figure 3B), the LPL and SDR sites in may had relatively high abundances of several antibiotic classes compared to other samples, including carbapenem resistance at LPL and SDR2, and tetracycline and fluoroquinolone at SDR2. Log normalized abundances of antibiotic resistance genes had a similar distribution across sites, but more highly abundant antibiotic resistance genes (ARGs) at SDR2 and TJ2, despite the fact that TJ2 had fewer total samples (Figure 3C). Highly abundant ARGs were most prevalent during the months of February and May, though July may be underrepresented with an N=3 (Figure 3D), where TJ sites were not represented. ARG classes had a wide distribution across geographically separated vibrio populations (Additional File 4), though classes were absent during some months reflecting indicating changes in host prevalence or in community dynamics.

Vibrio *species diversity*

We characterized *Vibrio* species diversity using both shotgun metagenomic sequencing of vibrio-enriched communities isolated on CHROMagar Vibrio plates (CHROMagar, Additional File 3) and amplicon sequencing of the Heat Shock Protein 60 (HSP60, aka chaperonin protein 60 or cpn60). We used 23 samples for these analyses collected during the months of February, March, May, July, and August, and both analyses used DNA as a sequencing template. As some species may not be amenable to growth under our isolation conditions, it is possible that additional *Vibrio* species are present in these samples.

To investigate vibrio diversity using shotgun metagenomic sequencing, we assigned taxonomy to unassembled sequence reads derived from cultured bacteria using MetaPhlAn2, which relies on clade-specific marker genes identified from ~17,000 reference genomes [37]. Across all samples, we identified 22 species of *Vibrio* bacteria, including the 3 target pathogen species, and two vibrio phages. These included the opportunistic human pathogenic species *V. alginolyticus*, and species pathogenic to animals including *Vibrio anguillarum, V. ordalii, V. harveyi,* and *V. campbellii*, and *V. splendidus* (Figure 4A, Additional Table 1). Site-specific shotgun sequencing analyses were in agreement with the ddPCR detection of target pathogen species; for example, in the LPL sample from May all three pathogens were detected both by ddPCR and the shotgun sequencing analysis (Additional Table 1). All sequences were gammaproteobacterial, and primarily belonged to *Vibrio* species and close relatives such as *Grimontia hollisae,* also an opportunistic human pathogen. Other bacterial genera containing human pathogens were found in lower abundance (< 10% and in only a few samples) including *Pseudomonas* and *Aeromonas* which were predominantly observed at the Los Peñasquitos site. We also observed multiple vibrio viruses among the samples. For example, ~82% of sequences retrieved from the TJ2-Feb sample belonged to vibrio phage vB VpaM MAR (Additional Table 1), while the remaining sequences belonged primarily to *V. parahaemolyticus* and *V. EX25* (also known as *V. antiquarius* [38]) which may suggest one of these species is the phage host. The vibrio temperate phage VP882 was also observed, which was originally isolated from a pandemic *V. parahaeomolyticus* O3:K6 strain shown to lyse *V. parahaemolyticus*, *V. vulnificus*, and *V. cholerae* strains [39].

We also sequenced the HSP60 amplicon, which has been used to characterize vibrio community diversity in multiple types of datasets [40–42]. We used the protocol described in

252 Jesser et al. 2018, and included only forward read sequences for this analysis. Approximately 557
253 unique *Vibrio* ASVs were identified, though many were in low abundance and some could be
254 classified only at the genus level (Additional Table 2). Of the ASVs that represented at least 5%
255 abundance across samples, 46 ASVs belonging to the *Vibrio* genus were identified indicating a
256 high level of intraspecific diversity (Figure 4B). These included *V. parahaemolyticus* and *V.*
257 *cholerae*, but not *V. vulnificus* which was detected by ddPCR and shotgun sequencing but absent
258 in the HSP60 dataset. Several *V. antiquarius* ASVs were abundant across the majority of samples,
259 and were particularly abundant the SDR and TJ sites, though these ASVs were entirely missing
260 from the LPL May and August samples. Additional human and animal pathogens were detected in
261 the HSP60 dataset, including *V. furnissii, V. angularium,* and *V. metschnikovii*, but these were very
262 low abundance.
263
264 *16S and 18S community composition, diversity, and environmental drivers*
265 We characterized taxonomic composition of the active prokaryotic and eukaryotic
266 communities co-occurring with *Vibrio* species using metabarcoding of the 16S and 18S rRNA
267 gene region with RNA as a template. In the prokaryotic community, 16S rRNA gene sequencing
268 resulted in ~30,000 (ASVs) after removing eukaryotic, mitochondrial, and chloroplast sequences.
269 Bacteria in the genus *Vibrio* comprised 0.03-4.9% of the 16S community [Mean = 0.44%]. A few
270 major bacterial classes dominated community composition, including *Gammaproteobacteria*
271 (encompassing *Vibrio* spp.), *Bacteroidia*, and *Alphaproteobacteria* (Figure 5A). LPL and SDR
272 sites had sizeable populations of *Oxyphotobacteria* (i.e. cyanobacteria), while other prominent
273 classes included Campylobacteria and *Verrucomicrobia*. The high number of distinct ASVs
274 compared to fewer broadly distributed prokaryotic classes suggests high levels of diversity within

taxa. 18S rRNA gene sequencing revealed ~17,000 eukaryotic ASVs which included phytoplankton, heterotrophic protists, and small multicellular eukaryotes, such as copepods. These were very diverse at the class level, and at some sites more than half of the classes were comprised of taxa that were <5% relative abundance (Figure 5B). Diatoms were the most common eukaryotes, comprising ~28% of 18S reads (Figure 5B), and while common at the LPL and SDR sites, they were particularly abundant at the TJ sites, frequently representing >75% of 18S reads. Other abundant groups included unicellular *Spirotrichea* ciliates, photosynthetic *Cryptophyceae*, and chitin-producing zooplankton, such as copepods.

In addition to class-level taxonomic composition, which is not always indicative of functional traits, we identified the top 20 most abundant ASVs in the prokaryotic and eukaryotic communities (Figure 5C,D) and observed their associations with particular pathogenic *Vibrio* species (See below, *Relationships between pathogenic* Vibrio spp. *and planktonic community taxa*).

We next examined the environmental, spatial, and temporal drivers of prokaryotic and eukaryotic community diversity. To examine alpha diversity we calculated a number of alpha diversity metrics, focusing on the number of observed ASVs (an indicator of species richness) and Shannon Diversity (an indicator of both richness and evenness). We then statistically compared alpha diversity of the categorical groups site and month by performing Kruskal-Wallis tests and investigated the relationship between the continuous variables temperature, salinity, and chlorophyll *a* and alpha diversity by performing Spearman's rank correlations. In the prokaryotic community sites and months did not significantly differ in Shannon Diversity or observed ASVs (Additional File 5). Higher temperature and chlorophyll *a* were associated with a lower number of observed ASVs. In eukaryotic communities both observed ASVs and Shannon Diversity were

298 significantly different between months and sites (Additional File 6), and higher temperature and
299 chlorophyll *a* were associated with lower alpha diversity including observed ASVs and Shannon
300 Diversity.

301 We also examined beta diversity by calculating Bray-Curtis dissimilarity for both 16S and
302 18S communities and conducting a Principal Coordinate Analysis (PCoA) (Additional File 7). To
303 statistically evaluate relationships between metadata variables and community dissimilarity, we
304 conducted PERMANOVA tests targeting both categorical and continuous variables using the
305 VEGAN function Adonis to assess the predictive power of environmental variables on community
306 composition. For the categorical variables, we preceded PERMANOVA analyses with a
307 betadispersion permutation betadisp test and PCoA visualization to determine whether dispersion
308 of samples within each group was homogenous and to calculate distance to the centroid of each
309 group cluster (See Methods section for additional information). We also conducted pairwise-
310 PERMANOVA tests between groups to better characterize spatial and temporal patterns.

311 For both prokaryotic and eukaryotic communities, samples collected at nearby sites or
312 close in time were more similar to each other than to other communities. This finding is supported
313 by both the similar position of group centroids when group dispersions were plotted (Additional
314 File 8, Additional File 9) and adonis PERMANOVA tests based on either group or site, which
315 showed significant differences (Additional Table 3). Additionally, we plotted pairwise sample
316 dissimilarity values by site and month to quantitatively observe patterns in sample similarity
317 (Additional File 10,11). Pairwise PERMANOVA tests between sites showed significant
318 differences between locations (e.g. LPL vs SDR1, $p\text{-}value = 0.003$) but not between sites at the
319 same location (e.g. SDR1 vs. SDR2, $p\text{-}value = 0.604$), suggesting a spatial influence on community
320 composition (Additional Table 3). Site accounted for 17% and 16% of prokaryotic and eukaryotic

community variance, while month accounted for 39% and 30% of the prokaryotic and eukaryotic community variance, respectively (Additional Table 3). Additionally, random forest classifiers predicted the overall accuracy to be ~92% for month-based classification and ~83% for site-based classification of samples in prokaryotic communities, and 50% classification accuracy for both variables in eukaryotic communities (Additional File 12). PERMANOVA tests between months showed temporally closer months were not significantly different from each other (e.g. March vs April, *p-value* = 0.354, March vs. May, *p-value* = 0.203), and winter months in particular clustered together in the beta dispersion plots (Additional Files 8,9). In examining the effect of site on prokaryotic (16S) community diversity, the beta dispersion permutation test for homogeneity of multivariate dispersions was significant (Additional Table 3), indicating that the dispersions were not homogenous across groups. A TukeyHSD test revealed that this was due to a significant difference in dispersion between sites SDR2 and LPL while all other sites had similar dispersion (Additional Table 3).

Temperature, salinity, and chlorophyll *a* all had a significant influence on beta diversity in both prokaryotic and eukaryotic communities, though these factors did not explain a majority of the community variance (Additional Table 3). Adonis PERMANOVA tests indicated that these factors explained 12%, 8%, and 4% of the variance for temperature, salinity, and chlorophyll *a*, respectively, in the prokaryotic community, and 7%, 4%, and 4% in the eukaryotic community. Furthermore, the predictive accuracy for temperature (R=0.90, MSE=7.52) and salinity (R=0.85, MSE=23.85) was much better in the prokaryotic community than in eukaryotic communities (temperature: R=0.47, MSE=25.55; salinity: R=0.44, MSE=75.63), indicated by lower MSE and higher R values in the former (Additional File 12).

Relationships between pathogenic Vibrio spp. *and planktonic community taxa*

To investigate associations between pathogenic *Vibrio* spp. and planktonic community members we used Spearman rank correlations to compare quantities of target species with abundant and functionally relevant planktonic classes, genera, and ASVs, and conducted linear discriminant analysis effect size (LEfSe) analyses to identify additional taxa associated with high concentrations of pathogenic *Vibrio* spp. For 16S communities, *V. parahaemolyticus* and *V. vulnificus* were positively associated with *Verrucomicrobiae*, a class isolated from many sample types and thought to be nearly ubiquitous in the marine environment[43] (Additional File 13A). *V. vulnificus, V. cholerae,* and the *V. vulnificus* virulence-associated gene *pilF* were positively associated with cyanobacteria and negatively associated with *Campylobacteria*, a pattern mirroring the negative association between these three marker genes and salinity. Additionally, *V. cholerae* was negatively associated with *Bacteroidia* and *Kirimatiellae* (Additional File 13A). Individual ASVs, however, often exhibited associations masked at class level. For example, the *Gammaproteobacteria* and *Alphaproteobacteria* classes had no significant associations with any species or virulence genes (Additional File 13A). However, when examining associations with the top 20 most abundant ASVs, multiple gammaproteobacterial ASVs (including members of the *Glaciecola* and *Marinobacterium* genera unidentified at the species level) were associated with lower temperatures and lower concentrations of *V. parahaemolyticus*. In contrast, alphabacterial taxa including *Rhodobacteraceae* and *Salinihabitans* ASVs were associated with higher temperatures and higher *V. parahaemolyticus* concentrations (Figure 5C).

Eukaryotic taxa also exhibited ASV-level associations unapparent at higher taxonomic ranks. The most abundant class, consisting of diatoms (class Bacillariophyta), was positively associated with *V. parahaemolyticus* and temperature, salinity, and chlorophyll *a* (Additional File

11B), but individual diatom taxa exhibited genus and species-level associations. For example, two *Chaetoceros pumilum* ASVs, representing the most abundant diatom taxa (Figure 5D, Additional File 14A,B), and a *Thalassiosira pseudonana* ASV followed this class-level pattern of positive *V. parahaemolyticus* association, but an abundant *Skeletonema* ASV was negatively associated. In the case of *Thalassiosira,* at the genus level the only significant associations are with low temperature and salinity and high *V. vulnificus* concentrations (Additional File 14B). However, an abundant ASV *T. pseudonana* ASV was positively associated with all three quantified *Vibrio* species while a *T. weissflogii* ASV was associated with high salinity and low *V. cholerae* (Figure 5D). Additionally, chitin-producing copepod genera were positively linked to the *Vibrio* species found in lower salinity waters (*V. vulnificus* and *V. cholerae*) and the *V. vulnificus* virulence associated gene *pilF*. (Additional File 13 C,D). This association does not appear to be driven by the most abundant copepod genus, *Pseudodiaptomus*, which shows no correlations with target species or environmental variables at the genus (Additional File 14D) or ASV level (Figure 5D). Rather, it appears to be related to several less abundant copepod genera found in low salinity samples, including the genera *Canuella, Tigriopus, Sinocalanus,* and *Cyclops*, which are also linked to the pathogenic *Vibrio* spp. commonly found in lower salinity samples (Additional File 14C).

Target pathogenic *Vibrio* spp. were often co-abundant with particular organisms despite the absence of significant correlations across all samples. For example, *Pseudodiaptamus* was not significantly associated with any target *Vibrio* species but was common and a dominant part of the 18S arthropod community at high *V. vulnificus* and *V. cholerae* sites, particularly LPL during May (*V. vulnificus* and *V. cholerae*) and March (*V. cholerae*) (Additional File 14C). For diatoms, *Cyclotella* spp. were abundant at both SDR sites during April, and May for SDR2, and

390 *Chaetoceros* was abundant when *V. cholerae* concentrations were highest, at LPL March through
391 May. While *Chaetoceros* diatoms were positively associated with *V. parahaemolyticus* across all
392 samples, *Thalassiosira* and *Cyclotella* diatoms, which were not, were some of the most abundant
393 eukaryotes in the high-*V. parahaemolyticus* samples, for example, at TJ2 in September, TJ1 in
394 April, and SDR2 May (Additional File 13A). This suggests that despite a lack of correlation across
395 samples, certain diatom genera or ASVs, including *Thalassiosira* and *Chaetoceros* diatoms, could
396 potentially be important components of high-*Vibrio* communities.

397 Lastly, we performed a linear discriminant analysis effect size (LEfSe) analysis to identify
398 particular genera and ASVs, regardless of abundance, associated with different quantities of
399 pathogenic *Vibrio* spp. Bacterial genera including *Dovosia* and *Mycobacteria* were more abundant
400 in samples with high or very high levels of the lower salinity species *V. cholerae* and *V. vulnificus*
401 (Additional File 15) while different genera were associated with higher *V. parahaemolyticus*
402 samples. Several Cyanobacteria (class: *Oxyphotobacteria*) genera and ASVs were associated with
403 the pathogenic species; *Anabaena* was associated with low levels of *V. cholerae*, *Prochlorococcus*
404 with low levels of *V. parahaemolyticus*, and two genera, *Prochlorothrix* and *Pseudoanabaena*
405 were linked to very high levels of *V. vulnificus*. Additionally, two *Picochlorum* ASVs were
406 associated with medium-high levels of *V. cholerae*. For eukaryotic ASVs, some less common algae
407 (i.e. not among the most abundant 20 ASVs) were associated with high levels of pathogenic targets,
408 including poorly characterized Prymnesiophytes (f_*Prymnesiaceae*) and Chrysophytes
409 (*Chrysophyceae* clade F and D) associated with high *V. vulnificus* and *V. cholerae*. Some less
410 common diatom species were found to be differentially abundant with high and very high levels
411 of *V. vulnificus* (*Cyclotella striata, C. scaldensis, Cylindrotheca Closterium,* and *Skeletonema*

412 *subsalsum*), while a poorly characterized *Thalassiosira* sp. ASV was associated with low levels of
413 *V. parahaemolyticus*.
414
415 **3. Discussion**
416 *Pathogenic* Vibrios *exhibit species-specific environmental preferences*
417 We observed distinct environmental preferences among *V. cholerae, V. vulnificus*, and *V.*
418 *parahaemolyticus* related to salinity and temperature (Figure 2A-C). While these environmental
419 factors are known to drive *Vibrio* distribution [19] many studies focus on individual species or the
420 *Vibrio* genus as a whole, potentially overlooking species shifts in response to surrounding
421 environmental community changes. By screening all three species, we capture some of these
422 dynamics. We also present the first quantification and ecological analysis of pathogenic *Vibrio*
423 spp. in the Southern California coastal region, an area of emerging risk due to warm coastal
424 seawater temperatures, high residential and tourism recreational water use, and seafood
425 cultivation. *Vibrio* spp. infections in Southern California have increased in recent years [44],
426 particularly in San Diego County; the most recent year assessed, 2018, showed the highest number
427 of infections ever reported and an infection rate substantially higher than both the California and
428 US infection rates[45].
429 All three species were abundant above 20 °C, a temperature at which human *Vibrio*
430 infections become a serious concern [3, 46, 47], and salinity levels were linked to peak species
431 abundance (Figure 2). *V. cholerae* and *V. vulnificus* were associated with lower salinities, and *V.*
432 *parahaemolyticus* was abundant in high salinity environments and associated with warmer
433 temperatures. While *V. cholerae* has been reported in high salinity conditions, it is most common
434 in low salinities, hence its tendency to contaminate drinking water. Likewise, *V. vulnificus* grows

435 poorly at salinities higher than 25 ppt, preferring the range of 10-18 ppt [48, 49]. Both species
436 peaked during warm summer months, typically a month or two before the peak temperature, and
437 high abundances (relative to the samples in this study and infectious dose estimates, e.g. [50]) were
438 only found from March through July (Additional File 2). As temperature was associated with high
439 salinity, intermediate conditions where temperatures are warm, but salinity is low or moderate may
440 be ideal. *V. parahaemolyticus* abundance was significantly associated with high temperatures, but
441 not salinity, suggesting *V. parahaemolyticus* may be a more halotolerant species. This is supported
442 by a meta-analysis finding that in contrast to *V. cholerae*, *V. parahaemolyticus* was distributed
443 across a broader salinity range of 3-35 ppt, with a warmer, more narrow temperature range [49]
444 [17]. The abundant *V. parahaemolyticus* populations we observed at extremely high salinities (>40
445 ppt) (Figure 2A) were out of the reported range in the meta-analysis and for other prior studies we
446 examined, perhaps suggesting unique high-salinity adaptations. However, the fundamental
447 ecological niche of many *Vibrio* species, particularly in terms of salinity, is often larger than
448 realistic environmental conditions [51]. As the high-salinity populations were found at moderate
449 temperatures, salinity tolerance may allow *V. parahaemolyticus* to take advantage of fortuitous
450 warm temperatures, though other site-specific factors are likely to be involved. In general, our
451 study supports previously reported temperature and salinity preferences previously observed in
452 these species [50], and also confirms that increasing seawater temperatures due to global warming
453 may pose a risk to regional human health in the future.

454 The distribution patterns we observed over time also provide information about the
455 ecological dynamics of these species. All three species were detected at all sites, occasionally
456 simultaneously. This suggests either a continuous presence at all times, sometimes below
457 detectable concentrations, or a temporal residence in the sediments or a viable but non-culturable

458 (VBNC) [52] state until conditions become ideal for proliferation in the water column [53]. The
459 distribution patterns of these three species were also linked to site, with *V. cholerae* most abundant
460 at LPL, *V. vulnificus* most common at the SDR sites, and *V. parahaemolyticus* most abundant at
461 the TJ sites. It is unclear whether those sites happened to present an ideal ecological niche at a
462 given time, or if other factors such as biotic interactions limit concentrations of species that would
463 otherwise be abundant.
464
465 *Diverse Vibrio populations contain genes associated with virulence and antibiotic resistance*
466 We assessed the diversity and genetic composition of the *Vibrio* communities with the
467 primary aim of understanding whether any additional pathogenic species were present, and if
468 pathogenic *Vibrio* species possessed genes of interest to human health, including genes associated
469 with virulence and antibiotic resistance. The HSP60 and shotgun metagenomic sequencing of
470 isolated bacteria identified several more *Vibrio* species of ecological interest that are either
471 potentially pathogenic themselves or may be capable of transferring human health-associated
472 genes to pathogenic species. For example, *V. antiquarius* (formerly known as *Vibrio* sp. Ex25,
473 which is the annotation assigned in the shotgun sequencing), was a highly abundant member of
474 the *Vibrio* community that was found alongside relatively high abundances of *V. parahaemolyticus*
475 (Figure 4B) and in diatom-dominant eukaryotic communities. This species is closely related to *V.*
476 *parahaemolyticus* and *V. alginolyticus* and was originally isolated from deep-sea hydrothermal
477 vents [52]. It is predicted to possess both the functional potential to survive in extreme conditions
478 and factors potentially involved in human disease caused by coastal *Vibrio* spp. More recent
479 studies have identified isolates in heat-shocked oysters, confirming that it inhabits diverse
480 environments [54], though given its similarity to other *Vibrio* species its taxonomic designation is

still being explored [55]. In our study, this species co-occurred with highly abundant pathogenic *Vibrio* spp., and while the ecological role and pathogenicity potential of *V. antiquarius* is unknown, its close phylogenetic relationship to and co-occurrence with the pathogenic species at these sites suggests they may be interacting and potentially even horizontally sharing genes. Other known *Vibrio* pathogens identified include *V. alginolyticus*, a species of increasing concern for human health [56], *V. fluvialis*, *V. furnissi*, and *V. metschnikovi*. Other *Vibrio*s identified are known coral, fish, and shellfish pathogens. Additionally, finding *Vibrio* phage sequences along with those of potential hosts is useful for identifying good candidates for phage therapy which has been used to reduce presence of pathogenic *Vibrio* species in aquaculture [57, 58].

The presence of virulence-associated genes underscores the actual potential for *Vibrio* infection, particularly as these genes can be horizontally transferred among species in the community [58, 59]. The potentially pathogenic *V. vulnificus* targets we examined appeared to contain a high percentage of virulence-associated genes as measured by ddPCR (Figure 2G,H). Half of the samples that tested positive for *V. vulnificus* also tested positive for one or both of the virulence-associated genes tested. For example, the *vcgC* gene, which is a marker more common in clinical *V. vulnificus* strains than environmental counterparts, was detected in 20% of these samples. Along the North Carolina Coast Williams et al. found that 5.3% of the *V. vulnificus* examined possessed the *vcgC* gene [39]. The *pilF* gene, which based on human serum sensitivity is highly correlated with pathogenicity potential [40], was also detected in 45% of these samples. Additionally, we detected high concentrations of *V. cholerae* (>2800 cells/ mL) at the LPL site. While it is unclear whether these strains possess virulence-associated genetic markers such as the *ctxA* toxin-associated gene, *V. cholerae* can infect humans even without these virulence genes and

503 *Vibrio* communities lacking these virulence genes can acquire them rapidly via viral infection [41]
504 or other horizontal gene transfer events.

505 Shotgun sequencing revealed additional virulence-associated genes present in the *Vibrio*
506 populations, and a wide-range of antibiotic resistance genes across sampling sites, which could
507 also inform human health risks and treatment of infected individuals. The *V. parahaemolyticus*
508 genes *trh* and *tdh* are challenging to detect and quantify in part because of their highly variable
509 sequences [59]. The shotgun sequencing approach we applied was able to detect the presence of
510 these sequences following failed attempts using previously published PCR primers, and in future
511 studies could be utilized in designing region-specific primers for these targets. These sequences
512 were most abundant in the LPL March site (Figure 3A), which had detectable but not high levels
513 of *V. parahaemolyticus* relative to other sites (476 copies/100 mL, Additional File 2). This may
514 indicate that this population had a proportionally high level of these virulence-associated genes.
515 Furthermore, the most abundant eukaryotic community members at this sampling point were
516 chitin-producing diatoms (Figure 5B, Additional File 15), which could provide an environmental
517 cue for these *Vibrio*s to transfer these genes. Additionally, high abundance of *V. parahaemolyticus*
518 thermolabile hemolysin (*tlh*) genes across all samples may further suggest these strains have
519 pathogenicity potential. While genes implicated in *V. vulnificus* and *V. parahaemolyticus* virulence
520 were detected in our samples, critical genes involved with *V. cholerae* virulence (*ace, zot, ctxA*)
521 were not detected which may suggest that even though very high levels of *V. cholerae* were
522 observed, particularly during March through may at the LPL sites, these populations are not likely
523 to cause the disease cholerae, though they may still be capable of causing other forms of vibriosis.

524 We observed several antibiotic resistance gene classes present in all samples, with the most
525 abundant classes being associated with resistance to tetracycline, glycopeptides, cephalosporins

526 and penam, and aminoglycosides. Pathogenic species of *Vibrio* bacteria are known to harbor
527 multiple antibiotic resistance genes [60, 61] which, like virulence genes, can be transmitted
528 between strains and even species via horizontal gene transfer. In previous studies, isolates of *V.*
529 *vulnificus* have been shown to be resistant to 8 or more antibiotics [62]. The resistance profiles
530 were similar among virulent and non-virulent strains, which is in agreement with our study where
531 we found some samples to have notably higher concentrations of virulence markers than others
532 (Figure 2A-C,Figure 5A), but a more even distribution of antibiotic resistance genes across sites.
533 Two sites with high levels of all three pathogenic species, LPL May and SDR2 May, have
534 relatively high levels of many different antibiotic classes, suggesting that these strains may be
535 highly antibiotic resistant. Pairing this data with the abundance of virulence genes is a useful tool
536 for understanding what populations may be dangerous, and paired with the planktonic community
537 data, which other organisms may be serving as vectors or reservoirs for these strains in the
538 environment.

539 Sequencing of *Vibrio* bacterial isolates, while producing a robust set of sequences with
540 which to assess diversity and virulence potential, does present some limitations. Notably,
541 sequences are limited to cultivated organisms potentially resulting in underestimated diversity, and
542 relative abundances of species may be skewed if some outgrow others. Nonetheless, our results
543 reveal a highly diverse *Vibrio* community and the abundance of many genes potentially associated
544 with human health.

545

546 *Pathogenic* Vibrio *spp. are commonly associated with prokaryotic and eukaryotic community*
547 *members, including chitin producers*

548 Our study elucidates links between pathogenic *Vibrio* species, the environment, and the
549 planktonic community. In conducting these analyses, it is clear that taxonomic resolution plays an
550 important role in defining potential functionally relevant relationships, and in establishing an
551 ecologically relevant context for prior and future studies. Using *Vibrio spp.* abundance to assess
552 community interactions, ecological niche, and even health risk is a common practice [63–66] but
553 these findings have been shown to be conflicting or misleading, unsurprisingly since *Vibrio* spp.
554 occupy distinct ecological niches and possess unique physiological capabilities, including
555 virulence mechanisms and modes of infection [3, 19]. This applies to community associations as
556 well; demonstrating the potential importance of species-level associations, Turner et al. [68] found
557 that while total *Vibrio* spp. bacteria were negatively correlated with copepods in a particular size
558 fraction (63-200 µm), the pathogenic species *V. parahaemolyticus* and *V. vulnificus* were actually
559 positively associated with copepods.
560 Bacterial species interacting with *Vibrio* spp. may impact virulence and environmental
561 persistence through horizontal gene transfer, population dynamics via viral infection, and growth
562 through competition or cooperation. We observed that common bacterial classes were similarly
563 present and relatively abundant across most sites and months, however, particular genera exhibited
564 species and virulence gene-specific correlations. Individual ASVs of the most abundant classes
565 were either positively or negatively associated with pathogenic *Vibrio* spp. despite no clear
566 correlations at the class level (e.g. *Gammaproteobacteria* and *Alphaproteobacteria*) (Figure 5A,
567 Additional File 13). We observed associations between members of the class *Oxyphotobacteria*
568 (Cyanobacteria), which was positively associated with *V. vulnificus*, *V. cholerae*, and *pilF*
569 (Additional File 13). Jesser and Noble 2018 compared cyanobacterial pigments to *Vibrio* relative
570 abundance and found a negative association between Cyanobacteria and *V. vulnificus*, but a

positive association with *V. parahaemolyticus*. These contrasting findings may be the result of different methods, regional differences in *Vibrio* associations, and/or a general lack of consistency when examining interactions at high taxonomic ranks (i.e. the class level). In our dataset only one ASV, a *Prochlorothrix* sp., was among the 20 most relatively abundant cyanobacteria and appeared to drive this pattern (Figure 5C). Multiple less abundant ASVs had more variable associations, as revealed by LEfSe, which may guide future studies (Additional File 15). A prior laboratory study investigated the response of *Synechococcus* sp. WH8102 to co-culture with *V. parahaemolyticus*, finding significant transcriptional changes including evidence of possible phosphate stress and utilization of specific nitrogen sources [13, 69]. While we didn't observe this organismal pairing in our dataset, future transcriptomic studies may identify similarly important *Vibrio*-cyanobacterial interactions.

A primary focus of our study was assessing pathogenic *Vibrio* spp. interactions with eukaryotes in the community, including those that produce chitin and organisms crucial to ecosystem function such as primary producers and grazers which are frequently overlooked in environmental microbiome studies. Of particular interest are diatoms (class *Bacillariophyta*) and copepods (class *Arthropoda*) as both of these groups are capable of chitin production and have been shown to interact with pathogenic *Vibrio* spp. in laboratory studies. For most known *Vibrio* spp., chitin serves as a nutrient source and a substrate for biofilm formation and subsequent protection from environmental stressors and predation [15, 16, 70–73]. It also induces a well-studied suite of cellular interactions initiating bacterial competition via the Type VI secretion system (T6SS) and natural competence, which may be the mechanism for how non-virulent populations become virulent [18, 66, 74].

The most abundant eukaryotic organisms found in our samples (>28% of 18S sequences) were diatoms. Prior studies have suggested that this group of algae are frequently associated with high *Vibrio* spp. concentrations [76–78]. Individual diatom species can host distinct microbial communities [75–77], release unique dissolved organic matter substrates [78–80], have variable susceptibility to viral or bacterial infection [81, 82], and may or may not exude chitin. When investigating interactions with *Vibrio* spp., however, diatoms often analyzed as a single group or categorized with other algal species. In our study diatoms as a group were positively associated with *V. parahaemolyticus*, temperature, salinity, and chlorophyll *a* (Additional File 13). At the more resolved genus level, this effect appears to be driven by the most abundant diatom genus *Chaetoceros* (Additional File 14B). In particular, two ASVs most closely related to *Chaetoceros pumilus* comprised the majority of *Chaetoceros* diatoms (Figure 5D).

Though many diatom genera contain chitin synthesis genes or full pathways, and may potentially produce chitin as a component of the cell wall [16], only two have been shown to actually exude chitin: *Thalassiosira* and *Cyclotella* [83]. These chitin producing diatoms were highly abundant in our samples, with 3 ASVs among the most abundant 20 eukaryotic taxa (Figure 5D). Notably, they exhibited ASV-specific relationships with the pathogenic *Vibrio* spp.; *T. pseudonana* was positively linked to all of the target species while *T. weissflogii* was negatively correlated with *V. cholerae*. Additionally, a *Cyclotella striata* ASV was negatively associated with *V. vulnificus*, possibly due to their different environmental preferences (*C. striata* was associated with high salinity, and *V. vulnificus* with low salinity). Laboratory studies have observed that chitinase-producing bacteria can have an algicidal effect on *T. pseudonana* [10], which is a well-characterized and genetically tractable model organism, however the positive associations observed in our study do not suggest this relationship. A prior laboratory study by Frischkorn et

616 al. 2013 observed *V. parahaemolyticus* attaching to the chitin-producing diatom *T. weissflogii,*
617 suggesting an unexplored mechanism of environmental persistence [11]. While we observed no
618 significant relationship between *T. weissflogii* and *V. parahaemolyticus* in our samples, despite
619 both species being abundant/relatively abundant, it is likely that attachment might also be
620 occurring with the closely related species *T. pseudonana,* which was positively associated with *V.*
621 *parahaemolyticus*. Thus, *T. pseudonana* may actually be a more ecologically relevant model for
622 studying these interactions.

623 The interaction between pathogenic *Vibrio* spp. and planktonic copepods is an important,
624 well-studied coastal phenomenon with demonstrated human health implications. Huq et al. (1983)
625 found that *V. cholerae* 01 and non-01 serovars attached to living but not dead *Acartia tonsa*,
626 *Eurytemora affinis*, and *Scottolana* spp. copepods from natural samples [84]. Another laboratory
627 study investigating these same living copepod species found that *V. cholerae* preferentially
628 attached to *Acartia tonsa* copepods over *Eurytemora affinis*, and that individual *V. cholerae* strains
629 exhibited different attachment efficiencies [85, 86]. In contrast, an O1 *V. cholerae* serovar (strain
630 N16961) and two non-O1/O139 *V. cholerae* isolates, were found to preferentially attach to dead,
631 rather than living, *Tigriopus californicus* copepods, as well as dinoflagellates [85, 86]. It is unclear
632 whether this difference is due to experimental methodology, the copepod species, or the *Vibrio*
633 strains. Environmental studies accounting for copepod taxonomy are rare and inconclusive: one
634 found no association between *V. cholerae* and co-occurring *Diaptomus* and *Cyclops* genera
635 copepods [68], and while others have reported qualitative associations in field samples [87, 88]
636 relationships based on quantitative data and copepod specificity are rare and consequently poorly
637 defined.

638 We observed positive correlations between pathogenic *Vibrio* spp., particularly those found
639 in lower salinities, and several copepod genera (Additional File 14D). A copepod annotated as
640 *Pseudodiaptamus inopinus*, an invasive species originating in Asia [90], was not significantly
641 associated with any *Vibrio* species across all samples but was highly relatively abundant during
642 the months where the highest levels of *V. cholerae* and *V. vulnificus* were detected at LPL and the
643 SDR sites (Figure 5B, Additional File 14C), and where the highest *V. parahaemolyticus*
644 abundances were documented at the TJ sites. Other abundant copepods were the Harpacticoid
645 genera *Canuella* and *Tigriopus*, both positively associated with *V. vulnificus* and the virulence-
646 associated gene *pilF* (Additional File 14C,D). In laboratory studies, the type IV pilus (containing
647 the *pilF* subunit) has been shown to be involved in chitin attachment to *Vibrio* spp. [89–91].
648 *Tigriopus* was also positively associated with *V. cholerae*. Though we could not obtain species-
649 specific taxonomic resolution for *Tigriopus*, it is a well-established laboratory model genus with
650 gene-silencing capabilities and full or partially assembled genomes for several species [92]. Thus,
651 *Tigriopus* and *Canuella* spp. may be good candidate genera for future laboratory studies involving
652 ecologically relevant *Vibrio*-plankton interactions.
653
654 *Drivers and ecological relevance of* Vibrio-*plankton dynamics*
655 As pathogenic *Vibrio* species were found in specific environments, and for some species
656 during particular times of the year, we investigated how temporal, spatial, and environmental
657 factors also shaped the surrounding microbial community. While month and site explained the
658 highest amount of variability for both the prokaryotic and eukaryotic communities (combined,
659 ~55% and 45% respectively), temperature and salinity also played an important role in defining
660 community structure (Additional Table 3). This was particularly true for the prokaryotic

community in which temperature and salinity combined explained and additional ~19% of community variability. This suggests that in addition to *Vibrio* abundance, community composition may also be predictable based on these factors and could potentially serve as molecular fingerprints of "high-vibrio" communities depending on the drivers. Additionally, if the abundance of easily observable larger species can be predicted, such as phytoplankton and copepods, that are positively associated with pathogenic *Vibrio*s, this have practical applications in the development of bioindicators of high *Vibrio* abundance, which can be technologically challenging and expensive to assess. Future studies should investigate *Vibrio* interactions in the context of these communities and their environmental drivers.

Vibrio associations with individual planktonic taxa must also be viewed in the context of shared ecological preferences, which we observed frequently in positively associated *Vibrio*-plankton relationships. Positive associations between pathogenic *Vibrio* species and planktonic ASVs across a large number of diverse samples, of which we observed many in our study, may suggest either common environmental drivers or actual interactions (e.g. mutualism, commensalism). These two possibilities are not mutually exclusive and are challenging to tease apart, but this does not detract from the significance of the positive associations. For example, *V. vulnificus* and *V. cholerae* are associated with low salinities. They are positively associated with the diatom *Thalassiosira pseudonana*, which is also associated with lower salinities, suggesting a shared environmental preference. We cannot determine whether these species are associated because they are actually interacting or if they simply cooccur in the same conditions, but regardless, being mutually abundant in the same conditions likely increases that possibility whereas organisms not found together may be less likely to interact.

683 Negative associations may represent antagonistic interactions or differing environmental
684 niches, however a lack of statistically significant association between organisms that are both
685 abundant does not preclude potential interactions. A lack of a correlation does not mean that two
686 groups are not interacting. For example, environmental factors may drive abundance of a *Vibrio*
687 species, but grazing pressure drives diatom abundance or community dominance. This would mask
688 a correlation but not preclude interactions of these organisms in the environment. Furthermore,
689 associations that are not linked to shared environmental preferences may be particularly suggestive
690 of an ecological interaction. Following the above example, *T. pseudonana* is also positively
691 associated with *V. parahaemolyticus*, a species that is not associated with low salinities and is in
692 fact often found in very high salinities. This exemplifies a positive association that is not rooted in
693 shared environmental preferences, and thus may be indicative of an interaction even in the absence
694 of an ideal environmental niche. This is especially intriguing given that *V. parahaemolyticus* has
695 been shown to attach to chitin of a closely related diatom species (discussed above).
696
697 **Conclusions**
698 Since many *Vibrio* species in the marine environment can cause human disease, it is critical
699 to understand their environmental preferences, ecological interactions, and genetic potential in
700 order to protect human health. We observed that pathogenic *Vibrio* species exhibited unique
701 temperature and salinity preferences, and were part of a diverse *Vibrio* community that harbored
702 both antibiotic resistance genes and genes associated with virulence. Pathogenic species were
703 associated with specific ASVs of chitin-producing eukaryotes such as diatoms and copepods, and
704 these relationships may facilitate attachment, environmental persistence, and horizontal gene
705 transfer. Relationships between *Vibrio*s and both prokaryotic and eukaryotic community members

706 in general were often ASV-specific, suggesting associations based on higher-level taxonomic
707 classifications may mask important interactions and impede ecologically relevant laboratory
708 studies. Furthermore, we identified shared environmental conditions that correlate with high levels
709 of pathogenic *Vibrios* and potential environmental reservoirs, which should be taken into
710 consideration when modelling *Vibrio* infection risk and identifying biomarkers of pathogen
711 species.
712
713 **Methods**
714 *Environmental sampling and* Vibrio *isolate culturing*
715 Monthly sampling was conducted from December 2015 to November 2016 at 3 locations
716 in San Diego County: Los Peñasquitos Lagoon (LPL), the San Diego River (SDR), and the Tijuana
717 River Estuary (TJ) (Figure 1A-D). For intra-site comparisons, two different sites at SDR (SDR1
718 and SDR2) and TJ (TJ1 and TJ2) were sampled, totaling 5 sampling sites. Temperature and salinity
719 were measured between 12pm and 1pm using a YSI Pro 30 field instrument (YSI Inc.). Unfiltered
720 water samples were collected in 4 L opaque bottles and processed in lab beginning no more than
721 2 hours after collection. These samples were kept in a cool area at roughly room temperature rather
722 than at 4 °C to prevent a viable but non-culturable (VBNC) state in *Vibrio* bacteria [76].
723 Water samples were gently filtered and flash-frozen in the lab for downstream processing.
724 For chlorophyll *a* quantification, 10-100 mL samples were collected on GF/F filters (Whatman)
725 and stored at -20 °C. Samples were later extracted in 90% acetone overnight and measured on a
726 10AU fluorometer (Turner), followed by addition of HCl and re-measurement to account for the
727 chlorophyll *a* degradation product pheophytin[93]. For downstream nucleic acid extractions 50-

400 mL samples were filtered onto 0.4 µm polycarbonate filters (Whatman) and stored at -80 °C until processing.

Vibrio isolate communities were collected at 23 sampling points, representing all 5 sites and collected during the months of February, March, May, July, and August. We filtered 1-100mL of whole seawater (depending on concentration of particulate matter in samples) onto 0.45 µm sterile cellulose nitrate filter membranes, which were transferred to CHROMagar Vibrio (CHROMagar Microbiology) plates and incubated overnight at 37 °C. These communities (examples in Additional File 3) were resuspended in 1 mL of either LB broth (Amresco) or Zobell Marine Broth 2216 (HiMedia), depending on sampling salinity and frozen as 15% glycerol stocks at -80 °C. Half of each glycerol stock was pelleted and used for downstream DNA extraction.

These collection protocols resulted in two distinct sets of samples that were used for next-generation sequencing. Samples collected on filters at 60 time points (12 months and 5 sites) representing the >0.4 µm community were used for 16S and 18S amplicon sequencing and for ddPCR quantification of pathogenic *Vibrio* species. Additionally, *Vibrio* isolate community samples were collected at 23 time points (representing all 5 sites and collected during the months of February, March, May, July, and August). These isolate "communities" were also used for HSP60 amplicon sequencing, and also for shotgun metagenomic sequencing.

DNA and RNA extraction and cDNA synthesis

Nucleic acids were extracted from filter samples using the NucleoMag Plant kit (Macherey-Nagel) for genomic DNA (gDNA) and the NucleoMag RNA kit (Macherey-Nagel) for RNA. Initial sample lysis buffer resuspension and vortexing was completed manually, the remainder using an epMotion liquid handling system (Eppendorf). RNA was reverse-transcribed into cDNA

using the SuperScript III First-strand cDNA Synthesis System (Invitrogen). gDNA was quantified using the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen) and RNA using the Quant-iT RiboGreen RNA Assay Kit. Nucleic acid integrity was confirmed using an Agilent 2200 TapeStation (Agilent). Genomic DNA was extracted from *Vibrio* isolate pellets using a DNeasy Blood and Tissue Kit (Qiagen), with subsequent quantification and quality control as described above.

Vibrio *digital droplet and end-point PCR*

Select pathogenic *Vibrio* species and virulence genes were quantified using the QX200 digital droplet PCR (ddPCR) System (BioRad), following the manufacturer's protocols and recommended reagents. DNA was used as a template for the ddPCR as *Vibrio* species abundance was determined using genes with only a single copy per genome, and measuring expression of genes with varying copy numbers (i.e. using an RNA template) would invalidate the use of this data for quantification analysis. Previously published assays based on qPCR were optimized for ddPCR, including running temperature gradients for each target to establish optimum reaction temperature, and primers and probes were ordered from Integrated DNA Technologies. Results from technical replicates were merged for analysis, and more than 19,000 droplets were measured per sample. Target-specific gBlocks (Integrated DNA Technologies) were used as positive controls for all ddPCR and end-point PCR targets.

Single copy-number gene targets for the species *V. parahaemolyticus, V. vulnificus,* and *V. cholerae* were quantified and used to approximate cell number per 100 mL of sample (Additional File 1). We targeted *toxR* for *V. parahaemolyticus* [23], *vvhA* for *V. vulnificus* [24], and *ompW* for *V. cholerae* [25]. We also quantified the virulence-associated *V. vulnificus* genes *vcgC* [21] and

774 *pilF* [22]. *V. cholerae* was not quantified during the months of December and April at SDR1 due
775 to technical problems.
776
777 *Next-generation library preparation and sequencing*
778 Amplicon libraries were constructed and sequenced using either a cDNA or DNA template
779 derived from filtered samples. For the 16S and 18S amplicons, duplicate libraries (derived from
780 two separate filters) were constructed using cDNA in order to characterize biologically active
781 community members, for a total of 120 libraries sequenced per amplicon. For HSP60 analysis of
782 the *Vibrio* isolate communities single libraries were similarly constructed using a DNA template
783 for a total of 23 samples. Libraries were sequenced at either the Institute for Genomic Medicine
784 (IGM, University of California, San Diego) or at the UC Davis Genome Center
785 (https://dnatech.genomecenter.ucdavis.edu/), with 300-bp paired end sequencing for the 16S and
786 HSP60 amplicons (MiSeq reagent kit v3) and 150 bp paired end sequencing for the 18S amplicon
787 (MiSeq reagent kit v2).
788 The 16S rRNA gene small subunit (SSU-rRNA) V4-5 region was targeted to characterize
789 the prokaryotic bacterial and archaeal community using primers 515F-926R [94]. The V9 region
790 of the 18S rRNA gene was targeted for eukaryotic community composition using primers 1389F
791 and 1510R [40, 95, 96]. The universal region of heat shock protein 60 (HSP60), also known as
792 chaperonin 60 (cpn60), was amplified and sequenced as described in Jesser and Noble 2018 using
793 primers identified in previous studies [97]. Sequences were filtered for quality using bbduk[98,
794 99]. Due to the size of the amplicon (549-567), paired-end reads did not merge sufficiently for a
795 robust analysis, so we used only high-quality forward reads.

For shotgun metagenomics libraries, DNA extracted from *Vibrio* isolate communities was fragmented to 400bp on an E210 Sonicator (Covaris). Sequencing libraries were prepared using the NEBNext Ultra II DNA Library Prep Kit (New England Biotechnologies), combined into 2 equimolar concentration pools of 13 samples each and sequenced on an Illumina HiSeq4000 at the UC Davis Genome Center with 250-bp paired-end reads.

Bioinformatic and statistical analyses

Demultiplexed sequences were analyzed using the QIIME 2 (version 2019.4) [100] pipeline and additional analyses and visualizations were conducted using the R package phyloseq (version 1.26.1) [101] and the web-based tool MicrobiomeAnalyst [102]. Sequences were quality filtered, chimeric sequences were removed, and exact amplicon sequence variants (ASVs)[103] were defined using dada2 [104] with a maximum expected error threshold of 2.0 (default) for 16S and 18S rRNA gene amplicons and 5.0 for the HSP60 amplicon. For 16S and 18S amplicons, replicate samples were merged using the "qiime feature-table group" function. Taxonomy was assigned using Silva [105] version 132 for bacterial and archaeal 16S sequences, and PR2 (version 4.11) [96] for 18S sequences, and the cpn60 database [106] with taxonomic designations derived from NCBI for HSP60 sequences as described in Jesser and Noble 2018 [29]. Chloroplast, mitochondrial, and eukaryotic sequences were removed from 16S datasets prior to downstream analyses. Alpha and beta diversity metrics for community composition were calculated using phyloseq. Singleton ASVs were retained for alpha diversity analyses. Beta diversity-based analyses were conducted after filtering out taxa that were not observed at least 3 times in 20% of samples and transforming to an even sampling depth, and calculating Bray Curtis dissimilarity matrices. PERMANOVA and beta dispersion tests were conducted using the Adonis and

819 betadisper functions, respectively, in the R package Vegan (version 2.5.6) [107]. Supervised
820 learning in the form of a RandomForestClassifier estimator method was implemented using the
821 QIIME 2 (version 2019.4) plug-in "sample-classifier" to predict categorical metadata variables
822 (i.e. site, month) in response to community composition, while a RandomForestRegressor
823 estimator was employed to predict continuous metadata variables (i.e. temperature, salinity).
824 Default values were used for both algorithms, along with a random seed generator (--p-random-
825 state 123) to replicate results each run.

826 For taxonomic analysis, shotgun sequences we first quality filtered using trimmomatic
827 [108] and checked for quality using FastQC [108]. We then assigned taxonomy to raw reads using
828 MetaPhlAn2 [109] (default parameters) and visualized the output using GraphPhlAn [110] (using
829 export2graphlan.py converter: --most_abundant 100 --abundance_threshold 1 --least_biomarkers
830 10 --annotations 5,6 --external_annotations 7 --min_clade_size 1). In order to analyze and target
831 the virulence genes and antibiotic resistance genes, reads were first assembled into sample-
832 specific assemblies using clc-assembler (CLC-Assembly-Cell version 5.1.1.184548) with the
833 following parameters: "--wordsize 31 –paired fb es 0 700 --min-length 200". These were then
834 grouped and merged by site into a global assembly, utilizing CD-HIT (version 4.6)[111] to remove
835 subsequences/duplicates with a minimum alignment identity of 0.95 and minimum contig length
836 of 300bp for merging assemblies. Open reading frames (ORF) were called on all contigs using
837 FragGeneScan (version 1.31) and read counts for each ORF for all samples were obtained by
838 mapping all reads to predicted ORFs using clc_mapper (CLC Read Mapper - Version
839 5.1.1.184548). Subsequently, all mapped read counts were merged across samples for all ORFs.

840 In order to screen for the virulence genes representative sequences for these genes were
841 queried against the global assembly using Blast (version ncbi-blast+-2.4.0) [110] and ORFs with

blast hits having >80% percent identity and >75% query coverage were selected and mapped counts across samples for these ORFs were used for visualizing the results. Additionally, RGI was used against the CARD database [101] to predict the presence of antibiotic resistant genes based on homology and SNP models in site-specific assemblies. All resulting .json files were parsed to collate all hits across all genomes in a matrix-like format. All count data from each of these analyses were log-transformed, manipulated using "dplyr" (version 1.0.1) and "reshape2" (version 1.4.4) and visualized using "ggplot2" (version 3.3.2) to generate the plot in R (version 4.0.2).

The web-based platform MicrobiomeAnalyst [112] was used to conduct the LefSE analysis. Chloroplast, mitochondria, and eukaryotic ASVs were removed from the 16S dataset prior to importing. ASVs were kept for analysis if they had more than 2 total counts and a minimum of 4 reads in at least 10% of the samples. Data were scaled using total sum scaling. Copy numbers for each category were: low = <100 copies/ 100 mL, medium = 100-1000 copies/ 100 mL, medium-high = 100-10,000 copies/ 100 mL (this category was only applied for *V. cholerae* to account for too few representative samples per category), high = 1,000-10,000 copies/ 100 mL, and very high = >10,000 copies/ 100 mL. We employed an FDR-adjusted *p-value* cutoff of 0.05, and a Log LDA score of 2.

Spearman's rank correlation coefficients were calculated to explore relationships between environmental variables, *Vibrio* quantification data, and relative abundance of groups of interest in the amplicon sequencing. Correlations were visualized as correlograms using the corrplot package in R [113], or as network graphs using the R package qgraph [113].

Declarations

865 *Ethics approval and consent to participate:* Not Applicable

866 *Consent for publication:* Not Applicable

867 *Availability of data and material*: The sequencing datasets generated and/or analyzed during the

868 current study are available in the NCBI repository, (BioProject accession no. PRJNA593265;

869 BioSample accession nos. SAMN13474661-SAMN13474785, SAMN13475110-

870 SAMN13475236, and SAMN13475238-SAMN13475318.) Scripts for analyses are in Additional

871 File 16 and 17. The metadata file used for Qiime 2 and phyloseq analyses, which also includes

872 environmental metadata, is also included as Additional Table 4.

873

874 *Competing interests:* The authors declare that they have no competing interests

875 *Funding:* Funding was provided to R.E.D. by San Diego IRACDA (NIH/NIGMS IRACDA K12

876 GM068524), the National Science Foundation PRFB (P2011025), the UCSD Center for

877 Microbiome Innovation, and the ARCS Foundation. This study was supported, in part, by

878 National Science Foundation (NSF-OCE-1637632 and NSF-OCE-1756884), NOAA

879 (NA15OAR4320071), and Gordon and Betty Moore Foundation grants GBMF3828 to AEA.

880 *Authors' contributions:* RED, AEA, JAS, and JFG were responsible for the conception and

881 design of the study. RED, HZ, and AR were responsible for acquiring and analyzing study data,

882 RED and AEA were responsible for data interpretation.

883

884 **Acknowledgements:** We would like to acknowledge Dr. Rachel Noble and Dr. Brett Froelich for

885 project guidance and assistance in culturing, identifying, and screening *Vibrio* bacteria, and Dr.

886 Kelsey Jesser for assistance in analyzing HSP60 amplicon sequences. We would like to further

887 acknowledge Lucy Mao, Meredith Raith, and Amanda Thygerson for assistance with ddPCR and
888 sample collection.

889

890 **References**

891 1. Centers for Disease Control and Prevention. CDC: Cholera: *Vibrio cholerae* infection. 2019.
892 https://www.cdc.gov/cholera/index.html. Accessed 1 May 2019.

893 2. Centers for Disease Control and Prevention. *Vibrio* Species Causing Vibriosis. 2019.
894 https://www.cdc.gov/vibrio/index.html. Accessed 1 May 2019.

895 3. Baker-Austin C, Oliver JD, Alam M, Ali A, Waldor MK, Qadri F, Martinez-Urtaza J. *Vibrio*
896 spp. infections. Nat Rev Dis Primers. 2018; https://www.nature.com/articles/s41572-018-0005-8.

897 4. Le Roux F, Blokesch M. Eco-evolutionary dynamics linked to horizontal gene transfer in
898 vibrios. Annu Rev Microbiol. 2018; doi:10.1146/annurev-micro-090817.

899 5. Baker-Austin C, Trinanes J, Gonzalez-Escalona N, Martinez-Urtaza J. Non-cholera vibrios: the
900 microbial barometer of climate change. Trends Microbiol. 2017; doi:10.1016/j.tim.2016.09.008.

901 6. Vezzulli L, Grande C, Reid PC, Hélaouët P, Edwards M, Höfle MG, Brettar I, Colwell RR,
902 Pruzzo C. Climate influence on Vibrio and associated human diseases during the past half-century
903 in the coastal North Atlantic. PNAS USA. 2016; doi:10.1073/pnas.1609157113.

904 7. Levy S. Warming trend, how climate shapes vibrio ecology. Environ Health Perspect. 2015;
905 doi:10.1126/science.275.5301.729b.

906 8. Lipp EK, Huq A, Colwell RR. Effects of global climate on infectious disease: the cholera model.
907 Clin Microbiol Rev. 2002; doi: 10.1128/CMR.15.4.757-770.2002

908 9. Colwell RR. Global climate and infectious disease: the cholera paradigm. Science. 1996; doi:
909 10.1126/science.274.5295.2025

910 10. Huq A, Small EB, West PA, Huq MI. Ecological relationships between *Vibrio cholerae* and
911 planktonic crustacean copepods. Appl Environ Microbiol. 1983; 45(1):275-83.

912 11. Kaneko T, Colwell RR. Adsorption of *Vibrio parahaemolyticus* onto chitin and copepods.
913 Appl Microbiol. 1975;29:269–74.

914 12. Mueller RS, McDougald D, Cusumano D, Sodhi N, Kjelleberg S, Azam F, et al. Vibrio
915 cholerae strains possess multiple strategies for abiotic and biotic surface colonization. J Bacteriol.
916 2007; doi:10.1128/JB.01867-06.

13. Williams TC, Ayrapetyan M, Oliver JD. molecular and physical factors that influence attachment of *Vibrio vulnificus* to chitin. Appl Environ Microbiol. 2015;81:6158–65. doi:10.1128/AEM.00753-15.

14. Jeuniaux C, Voss-Foucart MF. Chitin biomass and production in the marine environment. Biochem Syst Ecol. 1991; doi:10.1016/0305-1978(91)90051-Z.

15. Borgeaud S, Metzger LC, Scrignari T. The type VI secretion system of *Vibrio cholerae* fosters horizontal gene transfer. Science. 2015; doi: 10.1126/science.1260064.

16. Meibom KL, Blokesch M, Dolganov NA, Wu CY, Schoolnik GK. Chitin induces natural competence in *Vibrio cholerae*. Science. 2005; doi:10.1126/science.1120096.

17. Frischkorn KR, Stojanovski A, Paranjpye R. *Vibrio parahaemolyticus* type IV pili mediate interactions with diatom-derived chitin and point to an unexplored mechanism of environmental persistence. Environ Microbiol. 2013; doi:10.1111/1462-2920.12093.

18. Turner JW, Malayil L, Guadagnoli D, Cole D, Lipp EK. Detection of *Vibrio parahaemolyticus*, *Vibrio vulnificus* and *Vibrio cholerae* with respect to seasonal fluctuations in temperature and plankton abundance. Environ Microbiol. 2014; doi:10.1111/1462-2920.12246.

19. Takemura AF, Chien DM, Polz MF. Associations and dynamics of vibrionaceae in the environment, from the genus to the population level. Front Microbiol. 2014; doi:10.3389/fmicb.2014.00038.

20. Oliver JD, Pruzzo C, Vezzulli L, Kaper JB. *Vibrio* Species. In: Doyle MP, Buchanan RL, editors. Food Microbiology. Washington, DC, USA: ASM Press; 2012. p. 401–39. doi:10.1128/9781555818463.ch16.

21. Durkin CA, Mock T, Armbrust EV. Chitin in diatoms and its association with the cell wall. Eukaryot Cell. 2009; doi:10.1128/EC.00079-09.

22. Brunner E, Richthammer P, Ehrlich H, Paasch S, Simon P, Ueberlein S, van Pée KH. Chitin-based organic networks: an integral part of cell wall biosilica in the diatom *Thalassiosira pseudonana*. Angew Chem Int Ed. 2009; doi:10.1002/anie.200905028.

23. Grossart H-P, Levold F, Allgaier M, Simon M, Brinkhoff T. Marine diatom species harbour distinct bacterial communities. Environ Microbiol. 2005; doi:10.1111/j.1462-2920.2005.00759.x.

24. Baker-Austin C, Gore A, Oliver JD, Rangdale R, McArthur JV, Lees DN. Rapid *in situ* detection of virulent *Vibrio vulnificus* strains in raw oyster matrices using real-time PCR. Environ Microbiol Rep. 2010; doi:10.1111/j.1758-2229.2009.00092.x.

25. Roig FJ, Sanjuán E, Llorens A, Amaro C. *pilF* Polymorphism-based PCR to distinguish *Vibrio vulnificus* strains potentially dangerous to public health. Appl Environ Microbiol. 2010; doi:10.1128/AEM.01042-09.

26. Taiwo M, Baker-Austin C, Powell A, Hodgson E, Natås OB, Walker DI. Comparison of *toxR* and *tlh* based PCR assays for *Vibrio parahaemolyticus*. Food Control. 2017; doi:10.1016/j.foodcont.2017.02.009.

27. Campbell MS, Wright AC. Real-Time PCR Analysis of *Vibrio vulnificus* from Oysters. Appl Environ Microbiol. 2003; doi:10.1128/AEM.69.12.7137.

28. Garrido-Maestu A, Chapela MJ, Peñaranda E, Vieites JM, Cabado AG. In-house validation of novel multiplex real-time PCR gene combination for the simultaneous detection of the main human pathogenic vibrios (*Vibrio cholerae*, *Vibrio parahaemolyticus*, and *Vibrio vulnificus*). Food Control. 2014; doi:10.1016/j.foodcont.2013.09.026.

29. Rosche TM, Yano Y, Oliver JD. A rapid and simple PCR analysis indicates there are two subgroups of *Vibrio vulnificus* which correlate with clinical or environmental isolation. Microbiol Immunol. 2005; doi:10.1111/j.1348-0421.2005.tb03731.x.

30. Alcock BP, Raphenya AR, Lau TT, Tsang KK, Bouchard M, Edalatmand A, Huynh W, Nguyen AL, Cheng AA, Liu S, Min SY. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res. 2020; doi:10.1093/nar/gkz935.

31. Waldor MK, Mekalanos JJ. Lysogenic conversion by a filamentous phage encoding cholera toxin. Science. 1996; doi:10.1126/science.272.5270.1910.

32. Trucksis M, Galen JE, Michalski J, Fasano A, Kaper JB. Accessory cholera enterotoxin (Ace), the third toxin of a *Vibrio cholerae* virulence cassette. PNAS USA. 1993; doi:10.1073/pnas.90.11.5267.

33. Fasano A, Baudry B, Pumplin DW, Wasserman SS, Tall BD, Ketley JM, Kaper JB. *Vibrio cholerae* produces a second enterotoxin, which affects intestinal tight junctions. PNAS USA. 1991; doi:10.1073/pnas.88.12.5242.

34. RK Taylor, RK, Miller,VL, Furlong, DB , Mekalanos, JJ. Use of phoA gene fusions to identify a pilus colonization factor coordinately regulated with cholera toxin. PNAS USA. 1987; doi:10.1073/pnas.84.9.2833.

35. Lin W, Fullner KJ, Clayton R, Sexton JA, Rogers MB, Calia KE, Calderwood SB, Fraser C, Mekalanos J. Identification of a *Vibrio cholerae* RTX toxin gene cluster that is tightly linked to the cholera toxin prophage. PNAS USA. 1999; doi:10.1073/pnas.96.3.1071.

36. Pukatzki S, Ma AT, Sturtevant D, Krastins B, Sarracino D, Nelson WC, Heidelberg JF, Mekalanos JJ. Identification of a conserved bacterial protein secretion system in *Vibrio cholerae* using the Dictyostelium host model system. PNAS USA. 2006; doi:10.1073/pnas.0510322103.

37. Truong DT, Franzosa EA, Tickle TL, Scholz M, Weingart G, Pasolli E, Tett A, Huttenhower C, Segata N. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nat Methods. 2015; doi:10.1038/nmeth.3589.

986 38. Hasan NA, Grim CJ, Lipp EK, Rivera IN, Chun J, Haley BJ, Taviani E, Choi SY, Hoq M,
987 Munk AC, Brettin TS. Deep-sea hydrothermal vent bacteria related to human pathogenic Vibrio
988 species. PNAS USA. 2015; doi:10.1073/pnas.1503928112.

989 39. Lan SF, Huang CH, Chang CH, Liao WC, Lin IH, Jian WN, Wu YG, Chen SY, Wong HC.
990 Characterization of a new plasmid-like prophage in a pandemic *Vibrio parahaemolyticus* O3:K6
991 strain. Appl Environ Microbiol. 2009; doi:10.1128/AEM.02483-08.

992 40. Jesser KJ, Noble RT. Vibrio ecology in the Neuse River Estuary, North Carolina, characterized
993 by next-generation amplicon sequencing of the gene encoding heat shock protein 60 (hsp60). Appl
994 Environ Microbiol. 2018; doi:10.1128/AEM.00333-18.

995 41. Hunt DE, David LA, Gevers D, Preheim SP, Alm EJ, Polz MF. Resource partitioning and
996 sympatric differentiation among closely related bacterioplankton. Science. 2008;
997 doi:10.1126/science.1157890.

998 42. King WL, Siboni N, Kahlke T, Green TJ, Labbate M, Seymour JR. A new high throughput
999 sequencing assay for characterizing the diversity of natural Vibrio communities and its application
1000 to a Pacific Oyster mortality event. Front Microbiol. 2019; doi:10.3389/fmicb.2019.02907.

1001 43. Freitas S, Hatosy S, Fuhrman JA, Huse SM, Welch DBM, Sogin ML, et al. Global distribution
1002 and diversity of marine Verrucomicrobia. ISME J. 2012; doi:10.1038/ismej.2012.3.

1003 44. County of Los Angeles Public Health. Acute Communicable Disease Control 2016 Annual
1004 Morbidity Report: VIBRIOSIS. 2016. http://publichealth.lacounty.gov/acd/Diseases/Vibrio.pdf.

1005 45. County of San Diego Health and Human Services Agency. San Diego County Annual
1006 Communicable Disease Report. 2018.
1007 https://www.sandiegocounty.gov/content/dam/sdc/hhsa/programs/phs/documents/SanDiegoCoun
1008 ty_AnnualCDReport_2016.pdf.

1009 46. Kaspar CW, Tamplin ML. Effects of temperature and salinity on the survival of *Vibrio*
1010 *vulnificus* in seawater and shellfish. Appl Environ Microbiol. 1993;59:2425–9.

1011 47. Oliver JD. Wound infections caused by *Vibrio vulnificus* and other marine bacteria. Epidemiol
1012 Infect. 2005; doi:10.1017/s0950268805003894.

1013 48. Oliver JD. The biology of *Vibrio vulnificus*. Microbiol Spectr. 2015;
1014 doi:10.1128/microbiolspec.VE-0001-2014.

1015 49. Kaspar CW, Tamplin ML. Effects of temperature and salinity on the survival of *Vibrio*
1016 *vulnificus* in seawater and shellfish. Appl Environ Microbiol. 1993; doi:10.1128/AEM.59.8.2425-
1017 2429.1993.

1018 50. Schmid-Hempel P, Frank SA. Pathogenesis, virulence, and infective dose. PLoS Pathog. 2007;
1019 doi:10.1371/journal.ppat.0030147.

1020 51. Materna AC, Friedman J, Bauer C, David C, Chen S, Huang IB, Gillens A, Clarke SA, Polz
1021 MF, Alm EJ. Shape and evolution of the fundamental niche in marine Vibrio. ISME J. 2012;
1022 doi:10.1038/ismej.2012.65.

1023 52. Colwell RR, Brayton PR, Grimes DJ, Roszak DB, Huq SA, Palmer LM. Viable but non-
1024 culturable *Vibrio cholerae* and related pathogens in the environment: Implications for release of
1025 genetically engineered microorganisms. Nat Biotechnol. 1985; doi:10.1038/nbt0985-817.

1026 53. O'Malley MA. The nineteenth century roots of "everything is everywhere." Nat Rev
1027 Microbiol. 2007; doi:10.1038/nrmicro1711.

1028 54. Green TJ, Siboni N, King WL, Labbate M, Seymour JR, Raftos D. Simulated marine heat wave
1029 alters abundance and structure of vibrio populations associated with the Pacific Oyster resulting in
1030 a mass mortality event. Microb Ecol. 2019; doi:10.1007/s00248-018-1242-9.

1031 55. Turner JW, Tallman JJ, Macias A, Pinnell LJ, Elledge NC, Nasr Azadani D, Nilsson WB,
1032 Paranjpye RN, Armbrust EV, Strom MS. Comparative genomic analysis of *Vibrio diabolicus* and
1033 six taxonomic synonyms: A first look at the distribution and diversity of the expanded species.
1034 Front Microbiol. 2018; doi:10.3389/fmicb.2018.01893.

1035 56. Jacobs Slifka KM, Newton AE, Mahon BE. *Vibrio alginolyticus* infections in the USA, 1988–
1036 2012. Epidemiology & Infection. 2017; doi:10.1017/S0950268817000140.

1037 57. Le TS, Southgate PC, O'Connor W, Vu SV, Kurtböke Dİ. application of bacteriophages to
1038 control *Vibrio alginolyticus* contamination in oyster (*Saccostrea glomerata*) Larvae. Antibiotics.
1039 2020; doi:10.3390/antibiotics9070415.

1040 58. Richards GP. Bacteriophage remediation of bacterial pathogens in aquaculture: a review of the
1041 technology. Bacteriophage. 2014; doi:10.4161/21597081.2014.975540.

1042 59. Nilsson WB, Turner JW. The thermostable direct hemolysin-related hemolysin (*trh*) gene of
1043 *Vibrio parahaemolyticus*: Sequence variation and implications for detection and function. J
1044 Microbiol Methods. 2016; doi:10.1016/j.mimet.2016.04.007.

1045 60. Elmahdi S, DaSilva LV, Parveen S. Antibiotic resistance of *Vibrio parahaemolyticus* and
1046 *Vibrio vulnificus* in various countries: a review. Food Microbiol. 2016;
1047 doi:10.1016/j.fm.2016.02.008.

1048 61. Das B, Verma J, Kumar P, Ghosh A, Ramamurthy T. Antibiotic resistance in *Vibrio cholerae*:
1049 understanding the ecology of resistance genes and mechanisms. Vaccine. 2019;
1050 doi:10.1016/j.vaccine.2019.06.031.

1051 62. Baker-Austin C, McArthur JV, Lindell AH, Wright MS, Tuckfield RC, Gooch J, Warner L,
1052 Oliver J, Stepanauskas R. Multi-site analysis reveals widespread antibiotic resistance in the marine
1053 pathogen *Vibrio vulnificus*. Microb Ecol. 2009; doi:10.1007/s00248-008-9413-8.

1054 63. Rehnstam-Holm AS, Godhe A, Härnström K, Raghunath P, Saravanan V, Collin B,
1055 Karunasagar I, Karunasagar I. Association between phytoplankton and *Vibrio* spp. along the

southwest coast of India: a mesocosm experiment. Aquat Microb Ecol. 2010; doi:10.3354/ame01360.

64. Williams TC, Froelich BA, Phippen B, Fowler P, Noble RT, Oliver JD. Different abundance and correlational patterns exist between total and presumed pathogenic *Vibrio vulnificus* and *V. parahaemolyticus* in shellfish and waters along the North Carolina coast. FEMS Microbiol Ecol. 2017; doi:10.1093/femsec/fix071.

65. Oberbeckmann S, Fuchs BM, Meiners M, Wichels A, Wiltshire KH, Gerdts G. seasonal dynamics and modeling of a vibrio community in coastal waters of the North Sea. Microb Ecol. 2012; doi:10.1007/s00248-011-9990-9.

66. Main CR, Salvitti LR, Whereat EB, Coyne KJ. Community-level and species-specific associations between phytoplankton and particle-associated *Vibrio* species in Delaware's inland bays. Appl Environ Microbiol. 2015; doi:10.1128/AEM.00580-15.

67. Turner JW, Good B, Cole D, Lipp EK. Plankton composition and environmental factors contribute to Vibrio seasonality. ISME J. 2009; doi:10.1038/ismej.2009.50.

68. Tai V, Paulsen IT, Phillippy K, Johnson DA, Palenik B. Whole-genome microarray analyses of Synechococcus-Vibrio interactions. Environ Microbiol. 2009; doi:10.1111/j.1462-2920.2009.01997.x.

69. Hunt DE, Gevers D, Vahora NM, Polz MF. Conservation of the chitin utilization pathway in the Vibrionaceae. Appl Environ Microbiol. 2008; doi:10.1128/AEM.01412-07.

70. Lutz C, Erken M, Noorian P, Sun S, McDougald D. Environmental reservoirs and mechanisms of persistence of *Vibrio cholerae*. Front Microbiol. 2013; doi:10.3389/fmicb.2013.00375.

71. Antonova ES, Hammer BK. Genetics of natural competence in *Vibrio cholerae* and other Vibrios. Microbiology spectrum. 2015; doi:10.1128/microbiolspec.VE-0010-2014.

72. Erken M, Lutz C, McDougald D. Interactions of *Vibrio* spp. with zooplankton. Microbiology Spectrum. 2015; doi:10.1128/microbiolspec.VE-0003-2014.

73. Sun Y, Bernardy EE, Hammer BK, Miyashiro T. Competence and natural transformation in vibrios. Mol Microbiol. 2013; doi:10.1111/mmi.12307.

74. Asplund ME, Rehnstam-Holm AS, Atnur V, Raghunath P, Saravanan V, Härnström K, Collin B, Karunasagar I, Godhe A. Water column dynamics of Vibrio in relation to phytoplankton community composition and environmental conditions in a tropical coastal area. Environ Microbiol. 2011; doi:10.1111/j.1462-2920.2011.02545.x.

75. Chen W, Wangersky PJ. Production of dissolved organic carbon in phytoplankton cultures as measured by high-temperature catalytic oxidation and ultraviolet photo-oxidation methods. J Plankton Res. 1996;18:1201–11. https://doi.org/10.1093/plankt/18.7.1201.

1090 76. Grossart H-P, Simon M. Interactions of planctonic algae and bacteria: effects on algal growth
1091 and organic matter dynamics. Aquat Microb Ecol. 2007; doi:10.3354/ame047163.

1092 77. Bruckner CG, Rehm C, Grossart H-P, Kroth PG. Growth and release of extracellular organic
1093 compounds by benthic diatoms depend on interactions with bacteria. Environ Microbiol. 2011;
1094 doi:10.1111/j.1462-2920.2010.02411.x.

1095 78. Su JQ, Yang XR, Zheng TL, Tian Y, Jiao NZ, Cai LZ, Hong HS. Isolation and characterization
1096 of a marine algicidal bacterium against the toxic dinoflagellate *Alexandrium tamarense*. Harmful
1097 Algae. 2007; doi:10.1016/j.hal.2007.04.004.

1098 79. Paul C, Pohnert G. Interactions of the algicidal bacterium *Kordia algicida* with diatoms:
1099 regulated protease excretion for specific algal lysis. PLoS One. 2011;
1100 doi:10.1371/journal.pone.0021032.

1101 80. Mayali X, Azam F. Algicidal bacteria in the sea and their impact on algal blooms. J Eukaryot
1102 Microbiol. 2004; doi: 10.1111/j.1550-7408.2004.tb00538.x.

1103 81. Hildebrand M, Lerch SJL, Shrestha RP. Understanding diatom cell wall silicification-moving
1104 forward. Front Mar Sci. 2018; doi:10.3389/fmars.2018.00125.

1105 82. Tesson B, Masse S, Laurent G, Maquet J, Livage J, Martin-Jézéquel V, Coradin T.
1106 Contribution of multi-nuclear solid state NMR to the characterization of the *Thalassiosira*
1107 *pseudonana* diatom cell wall. Anal Bioanal Chem. 2008; doi:10.1007/s00216-008-1908-0.

1108 83. Li Y, Lei X, Zhu H, Zhang H, Guan C, Chen Z, Zheng W, Fu L, Zheng T. Chitinase producing
1109 bacteria with direct algicidal activity on marine diatoms. Sci Rep. 2016; doi:10.1038/srep21984.

1110 84. Rawlings TK, Ruiz GM, Colwell RR. Association of *Vibrio cholerae* O1 El Tor and O139
1111 Bengal with the copepods *Acartia tonsa* and *Eurytemora affinis*. Appl Environ Microbiol. 2007;
1112 doi:10.1128/AEM.01238-07.

1113 85. Huq A, Sack RB, Nizam A, Longini IM, Nair GB, Ali A, Morris JG, Khan MH, Siddique AK,
1114 Yunus M, Albert MJ. Critical factors influencing the occurrence of *Vibrio cholerae* in the
1115 environment of Bangladesh. 2005; doi:10.1128/AEM.71.8.4645.

1116 86. Lizárraga-Partida ML, Mendez-Gómez E, Rivas-Montaño AM, Vargas-Hernández E, Portillo-
1117 López A, González-Ramírez AR, Huq A, Colwell RR. Association of *Vibrio cholerae* with
1118 plankton in coastal areas of Mexico. Environ Microbiol. 2009; doi:10.1111/j.1462-
1119 2920.2008.01753.x.

1120 87. Cordell JR, Morrison SM. The invasive Asian copepod *Pseudodiaptomus inopinus* in Oregon,
1121 Washington, and British Columbia estuaries. Estuaries. 1996; doi:10.2307/1352523.

1122 88. Cordell JR, Morgan CA, Simenstad CA. Occurrence Oftheasian calanoid copepod
1123 *Pseudodiaptomus Inopinus* in the zooplankton of the Columbia River Estuary. J Crustacean Biol.
1124 1992; doi: 10.2307/1549079.

89. Barreto FS, Schoville SD, Burton RS. Reverse genetics in the tide pool: knock-down of target gene expression via RNA interference in the copepod *Tigriopus californicus*. Mol Ecol Resour. 2015; doi:10.1111/1755-0998.12359.

90. Wang G, Xu J, Zeng C, Jia Q, Wu L, Li S. Pelagic microalgae as suitable diets for the benthic harpacticoid copepod *Tigriopus japonicus*. Hydrobiologia. 2015; doi:10.1007/s10750-015-2339-5.

91. Kang S, Ahn DH, Lee JH, Lee SG, Shin SC, Lee J, Min GS, Lee H, Kim HW, Kim S, Park H. The genome of the Antarctic-endemic copepod, *Tigriopus kingsejongensis*. Gigascience. 2017; doi:10.1093/gigascience/giw010.

92. Holm-Hansen O, Lorenzen CJ, Holmes RW, Strickland JDH. Fluorometric determination of chlorophyll. ICES J Mar Sci. 1965; doi:10.1093/icesjms/30.1.3.

93. Parada AE, Needham DM, Fuhrman JA. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environ Microbiol. 2016; doi:10.1111/1462-2920.13023.

94. Amaral-Zettler LA, McCliment EA, Ducklow HW, Huse SM. A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA Genes. PLoS One. 2009; doi:10.1371/journal.pone.0006372.

95. Hill JE, Town JR, Hemmingsen SM. Improved template representation in cpn 60 polymerase chain reaction (PCR) product libraries generated from complex templates by application of a specific mixture of PCR primers. Environ Microbiol. 2006; doi:10.1111/j.1462-2920.2005.00944.x.

96. Hill JE, Penny SL, Crowell KG, Goh SH, Hemmingsen SM. cpnDB: A chaperonin sequence database. Genome Res. 2004; doi:10.1101/gr.2649204.

97. BBDuk Guide. https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbduk-guide/. Accessed 23 Sep 2020.

98. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 2019; doi:10.1038/s41587-019-0209-9.

99. Estaki M, Jiang L, Bokulich NA, McDonald D, González A, Kosciolek T, Martino C, Zhu Q, Birmingham A, Vázquez-Baeza Y, Dillon MR. QIIME 2 Enables comprehensive end-to-end analysis of diverse microbiome data and comparative studies with publicly available data. Curr Protoc Bioinformatics. 2020; doi:10.1002/cpbi.100.

100. McMurdie PJ, Holmes S. Phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One. 2013; doi:10.1371/journal.pone.0061217.

101. Dhariwal A, Chong J, Habib S, King IL, Agellon LB, Xia J. MicrobiomeAnalyst: A web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. Nucleic Acids Res. 2017; doi:10.1093/nar/gkx295.

102. Callahan BJ, McMurdie PJ, Holmes SP. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J. 2017; doi:10.1038/ismej.2017.119.

103. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 2016; doi:10.1038/nmeth.3869.

104. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Res. 2013; doi:10.1093/nar/gks1219.

105. Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L, Boutte C, Burgaud G, De Vargas C, Decelle J, Del Campo J. The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. Nucleic Acids Res. 2013; doi:10.1093/nar/gks1160.

106. Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'hara RB, Simpson GL, Solymos P, Stevens MH, Wagner H. vegan: Community Ecology Package. 2019. https://CRAN.R-project.org/package=vegan.

107. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014; doi:10.1093/bioinformatics/btu170.

108. Babraham Bioinformatics. https://www.bioinformatics.babraham.ac.uk/index.html. Accessed 23 Sep 2020.

109. Asnicar F, Weingart G, Tickle TL, Huttenhower C, Segata N. Compact graphical representation of phylogenetic data and metadata with GraPhlAn. PeerJ. 2015; doi:10.7717/peerj.1029.

110. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006; doi:10.1093/bioinformatics/btl158.

111. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2018; doi:10.1093/nar/gkx1095.

112. Wei Taiyun; Simko V. R package: "corrplot": Visualization of a Correlation Matrix (version 0.84). https://github.com/taiyun/corrplot.

113. Epskamp S, Costantini G, Haslbeck J, Isvoranu A, Cramer AO, Waldorp LJ, Schmittmann VD, Borsboom D. Package 'qgraph'. https://cran.rproject.org/web/packages/qgraph/qgraph.pdf.

Additional Materials:

Additional Files

AdditionalMaterials_Dineretal2020Microbiome contains Additional File 1-14

(1) Additional File 1, Primers and probes used in this study.

(2) Additional File 2, *Vibrio* target gene copies measured for each sample and target.

(3) Additional File 3, Isolates of putative pathogenic Vibrio species on CHROMagar Vibrio agar plates.

(4) Additional File 4, Antibiotic resistance gene classes per site.

(5) Additional File 5, Prokaryotic community alpha diversity boxplots, statistical analyses, and correlations with pathogenic *Vibrio* species and environmental variables.

(6) Additional File 6, Eukaryotic community alpha diversity boxplots, statistical analyses, and correlations with pathogenic *Vibrio* species and environmental variables.

(7) Additional File 7, PCoA plots of 16S and 18S beta diversity based on Bray Curtis dissimilarity.

(8) Additional File 8, 16S beta dispersion PCoA and distance to centroid boxplots by month and site.

(9) Additional File 9, 18S beta dispersion PCoA and distance to centroid boxplots by month and site.

(10) Additional File 10, 16S Bray Curtis dissimilarities of pair-wise comparisons between all samples grouped by site and month.

(11) Additional File 11, 18S Bray Curtis dissimilarities of pair-wise comparisons between all samples grouped by site and month.

(12) Additional File 12, Random forest classification and regression results for 16S and 18S

amplicon data

(13) Additional File 13, Class-level associations between the top 10 most abundant prokaryotic and eukaryotic genera

(14) Additional File 14, Composition of diatom (Class: *Bacillariophyta*) and copepod (Class: *Arthropoda*) 18S communities and associations with *Vibrio* targets and environmental variables.

(15) Additional File 15, Linear discriminant effect size analysis (LEfSe) to identify ASVs differentially abundant between categorical levels of pathogenic *Vibrio* spp.

(16) Additional File 16. R scripts used for analyses.

(17) Additional File 17. Python scripts used for analyses.

Additional Tables

(1) Additional Table 1, AdditionalTable1_MetaPhlAn2_IsolateCultureTaxonomy

(2) Additional Table 2, AdditionalTable2_RelativeAbundanceofIsolateVibrioSpeciesHSP60

(3) Additional Table 3, AdditionalTable3_Adonis_betadispersion_stats

(4) Additional Table 4, Metadata file used for Qiime 2 and phyloseq analyses

Figure Legends:

Figure 1: Location of the sampling sites and environmental conditions at the time of sampling. Locations are mapped in the context of the San Diego region using Google Earth. Site abbreviations are as follows: (A) LPL = Los Peñasquitos Lagoon, (B) SDR = San Diego River (site 1 and 2), and (C)TJ = Tijuana River Estuary (site 1 and 2), with (D) depicting the regional context. Environmental conditions, including (E) temperature, (F) salinity, and (G) chlorophyll *a*,

a proxy for photosynthetic organism abundance, were measured monthly at each site (LPL = red, SDR1 = mustard, SDR2 = green, TJ1 = blue, and TJ2 = purple) for one year from December 2015-November 2016. Spearman rank correlations identified relationships between environmental variables, where values closer to 1 (dark blue) are positive correlations and values closer to -1 (dark red) are negative correlations, and * represents significant correlations (*p-value* < 0.05).

Figure 2. Number of single-genome copy genes (a proxy for cell numbers) per 100 mL detected by digital droplet PCR. (A) the *Vibrio parahaemolyticus (Vp)* species-specific gene target *toxR* (B) the *Vibrio vulnificus (Vv)* species-specific target *vvhA,* and (C) and the *Vibrio cholerae* species-specific target *ompW,* with marker size corresponding to copy number and color corresponding to site, plotted against temperature and salinity. (D) Spearman's rank correlation coefficients of associations between environmental the variables temperature, salinity, and chlorophyll *a*, and *Vibrio* species and virulence gene targets. Blue represents a strong positive correlation, while red represents a strong negative correlation, significant correlations (*p-value* < 0.05) are denoted with *. Number of copies detected per 100 mL by digital droplet PCR for the *Vibrio vulnificus* virulence-associated genes (E) *vcgC* and (F) *pilF,* plotted against temperature and salinity. (G) The percent of *V. vulnificus* samples where no virulence gene was detected, either *vcgC* or *pilF* were detected, or both were detected. (H) the ratio of the number of *pilF* and *vcgC* copies detected to total *V. vulnificus* determined by *vvhA* copy number.

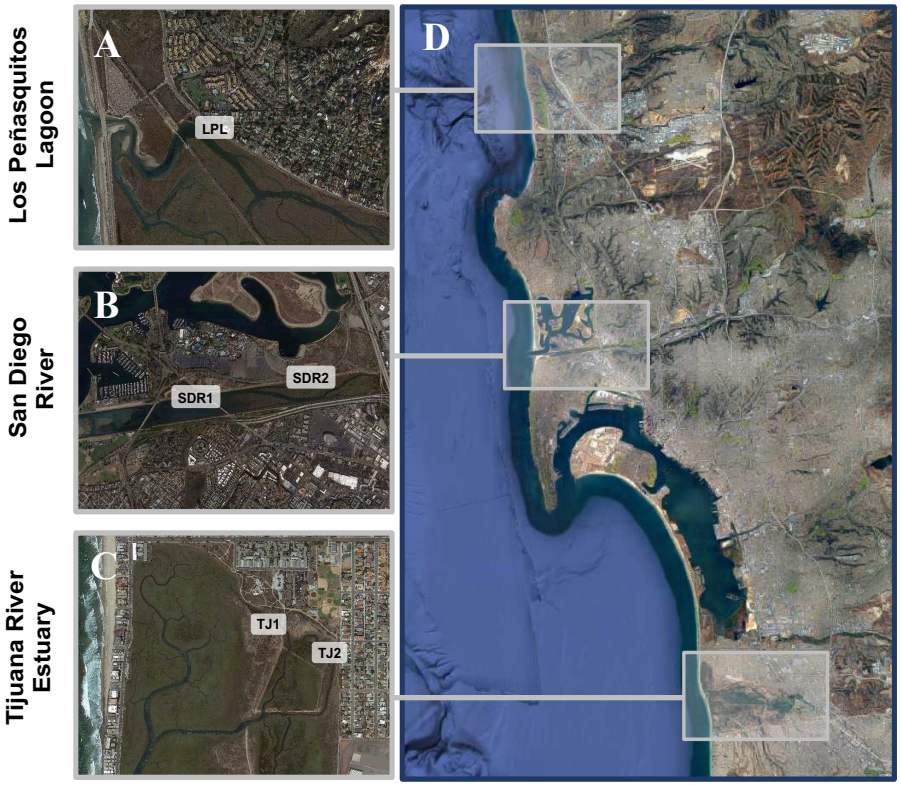
Figure 3. Virulence and antibiotic resistance gene profiles of *Vibrio* isolate communities based on shotgun metagenomic sequencing. Heatmap depicting (A) log normalized abundance of genes associated with virulence identified in the literature. On the Y axis, genes associated with

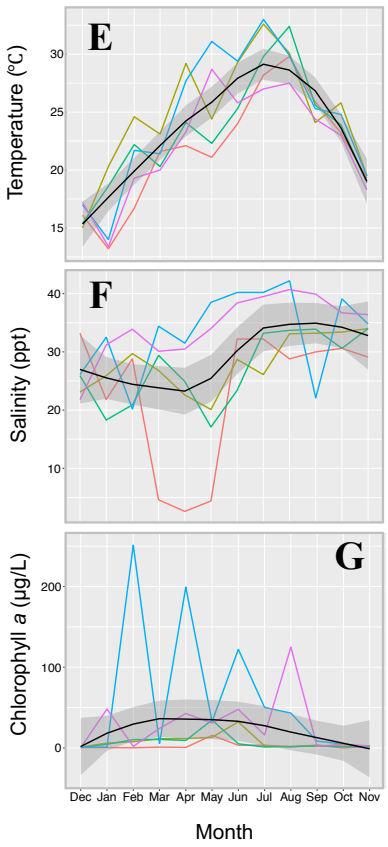
V.vulnificus are in dark blue, genes associated with *V. parahaemolyticus* in green, and *V. cholerae* in grey (B) Heatmap showing log normalized abundance of the top 20 most abundant antibiotic resistance gene (ARG) classes as identified in the CARD database across months and sites. For (A) and (B), months where high concentrations of one or more pathogenic *Vibrio* sp. were detected by ddPCR are indicated in red on the X axis. Overall abundance of ARGs are shown by (C) site and (D) month.

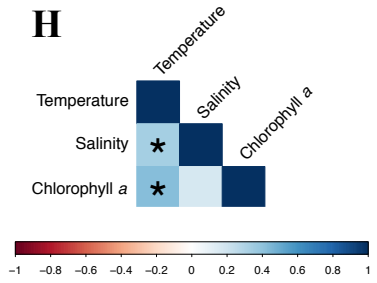
Figure 4. Diversity of isolated *Vibrio* species and related bacteria . (A) cladeogram depicting the taxonomy of species isolated from CHROMagar media identified by shotgun metagenomic sequencing. Taxonomic assignments were determined using MetaPhlAn2, which annotates and phylogenetically places unassembled sequence reads using clade-specific genetic markers, and (B) heatmap of the relative abundance of *Vibrio* spp. ASVs based on HSP60 amplicon sequencing and the associated abundance of quantified *Vibrio* target genes and environmental conditions (temperature and salinity). Hierarchical clustering is used to order samples based on Euclidean distance within sites so that samples within sites with similar abundance profiles are grouped near each other.

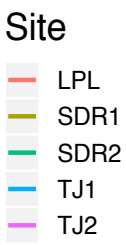
Figure 5. Taxonomic composition of prokaryotic (16S) and eukaryotic (18S) communities, and ASV-level associations between abundant prokaryotic and eukaryotic community members, *Vibrio* species and virulence gene targets, and environmental variables. Taxonomic composition of 16S (A) and 18S (B) communities by class. Cut-offs of <1% and <5% abundance represent taxa cumulative abundance across all samples. Months in red indicate that one or more of the pathogenic *Vibrio* species quantified by ddPCR were present at concentrations exceeding

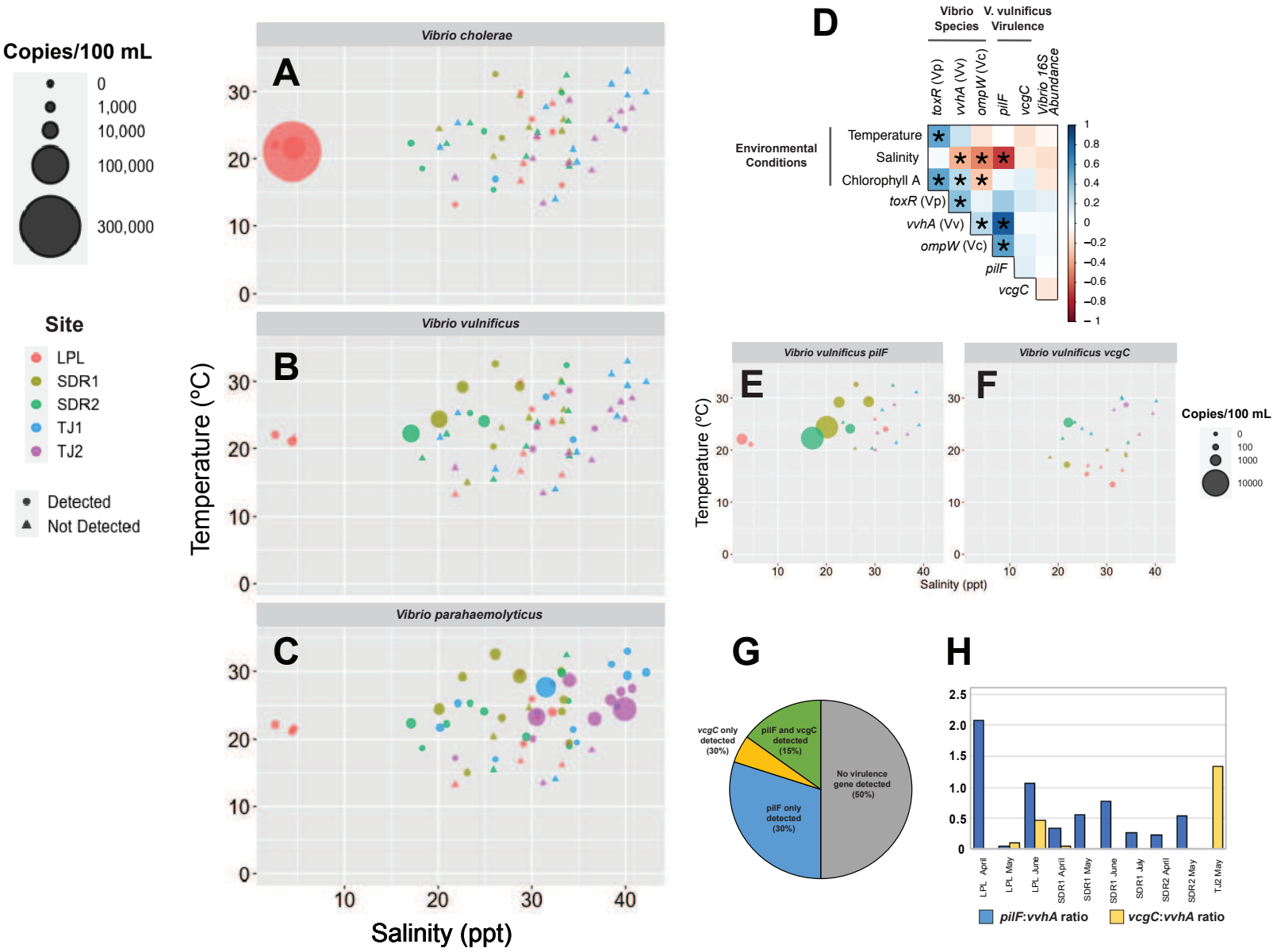
1000 copies/100 mL. Spearman's rank correlations between the top 20 most abundant (C) prokaryotic and (D) eukaryotic ASVs are organized by class, and only significant (*p-value* < 0.05) correlations are shown. Blue lines depict significant positive associations while red lines represent negative associations, and line thickness indicates the strength of the correlation. Marker genes used to quantify pathogenic *Vibrio* species are shown in grey circles, and correspond to the following species and virulence genes: *toxR: V. parahaemolyticus* species, *vvhA*: *V. vulnificus* species, *ompW*: *V. cholerae* species, *pilF: V. vulnificus* virulence-associated gene, *vcgC = V. vulnificus* virulence-associated gene.

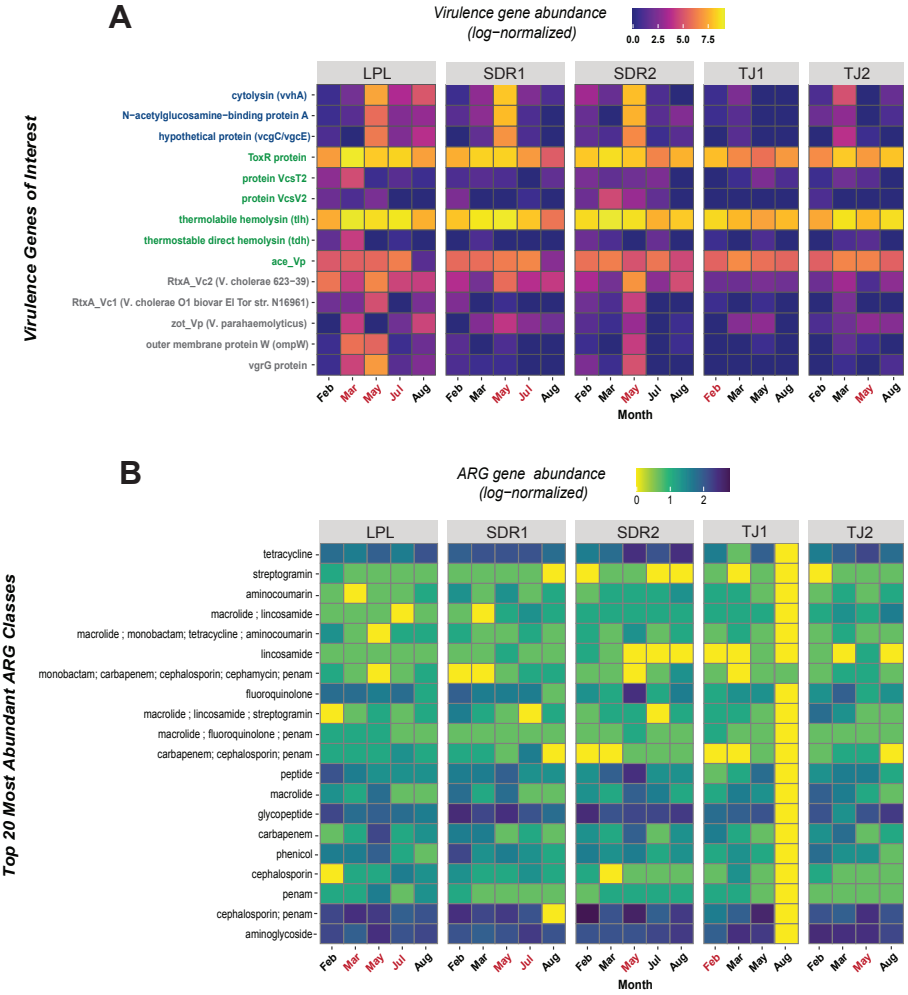

Los Peñasquitos Lagoon
San Diego River
Tijuana River Estuary
A
LPL
B
SDR1
SDR2
C
TJ1
TJ2
D

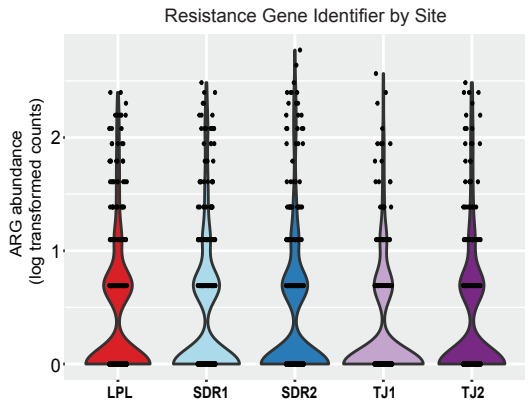

E
Temperature (°C)
30
25
20
15
F
Salinity (ppt)
40
30
20
10
G
Chlorophyll a (µg/L)
200
100
0
Dec Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov
Month

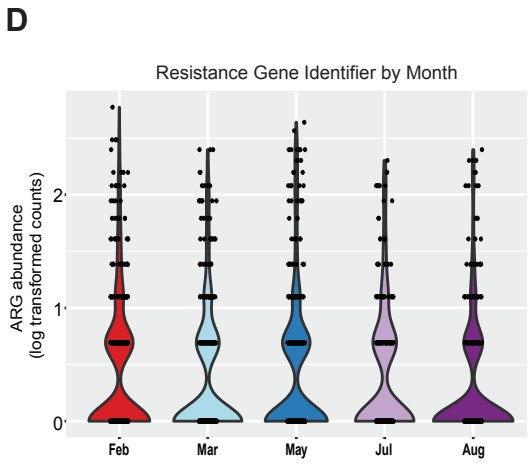

H
Temperature
Salinity
Chlorophyll a
*
*
−1 −0.8 −0.6 −0.4 −0.2 0 0.2 0.4 0.6 0.8 1
Spearman's Rank Correlation Coefficient


Site
LPL
SDR1
SDR2
TJ1
TJ2

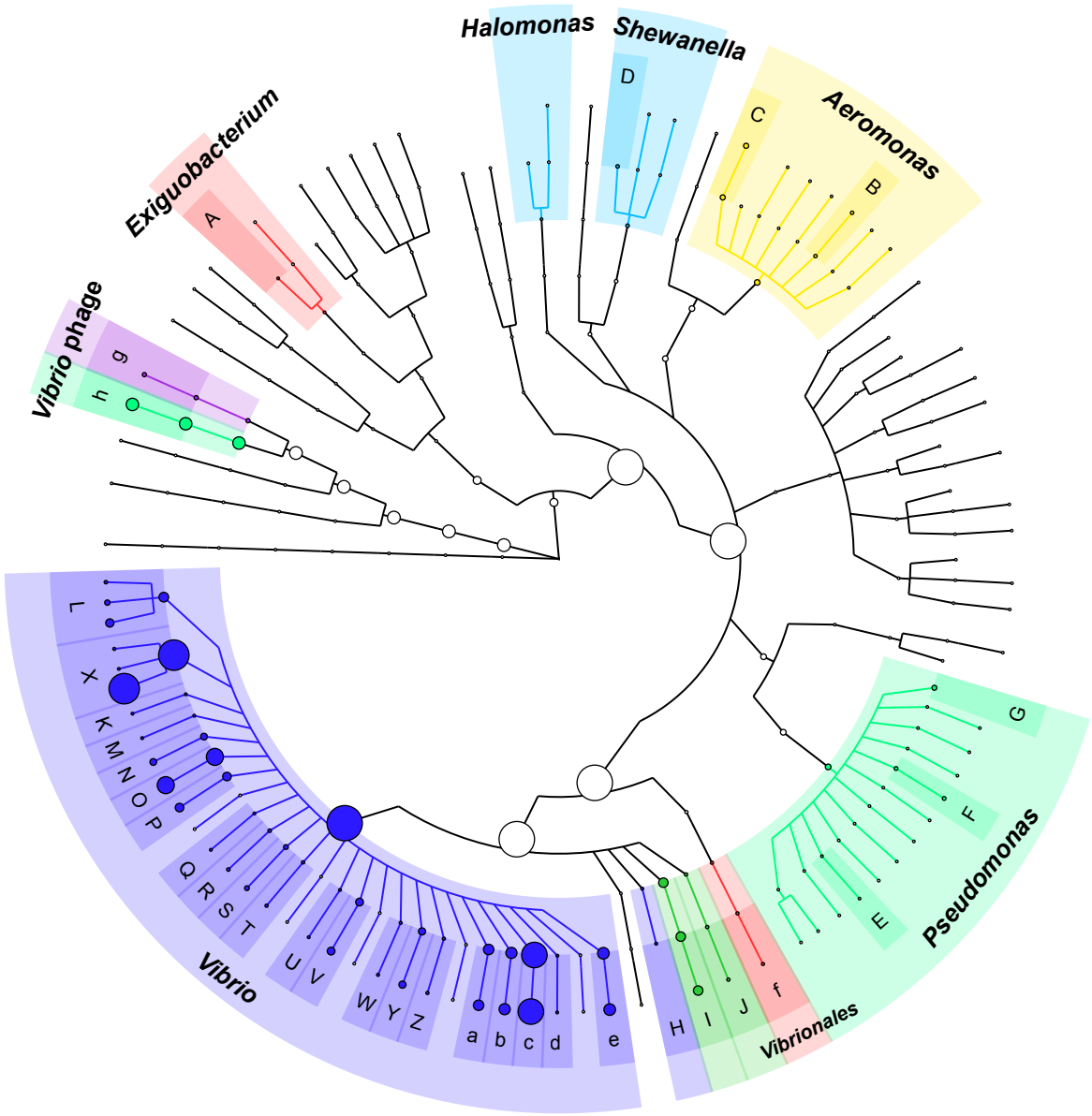

Copies/100 mL
0
1,000
10,000
100,000
300,000
Site
LPL
SDR1
SDR2
TJ1
TJ2
Detected
Not Detected
Vibrio cholerae
A
Vibrio vulnificus
B
Vibrio parahaemolyticus
C
Temperature (°C)
Salinity (ppt)
D
Vibrio Species
V. vulnificus Virulence
toxR (Vp)
vvhA (Vv)
ompW (Vc)
pilF
vcgC
Vibrio 16S Abundance
Environmental Conditions
Temperature
Salinity
Chlorophyll A
E
Vibrio vulnificus pilF
F
Vibrio vulnificus vcgC
Copies/100 mL
G
vcgC only detected (30%)
pilF and vcgC detected (15%)
No virulence gene detected (50%)
pilF only detected (30%)
H
LPL April
LPL May
LPL June
SDR1 April
SDR1 May
SDR1 June
SDR1 July
SDR2 April
SDR2 May
TJ2 May
pilF:vvhA ratio
vcgC:vvhA ratio

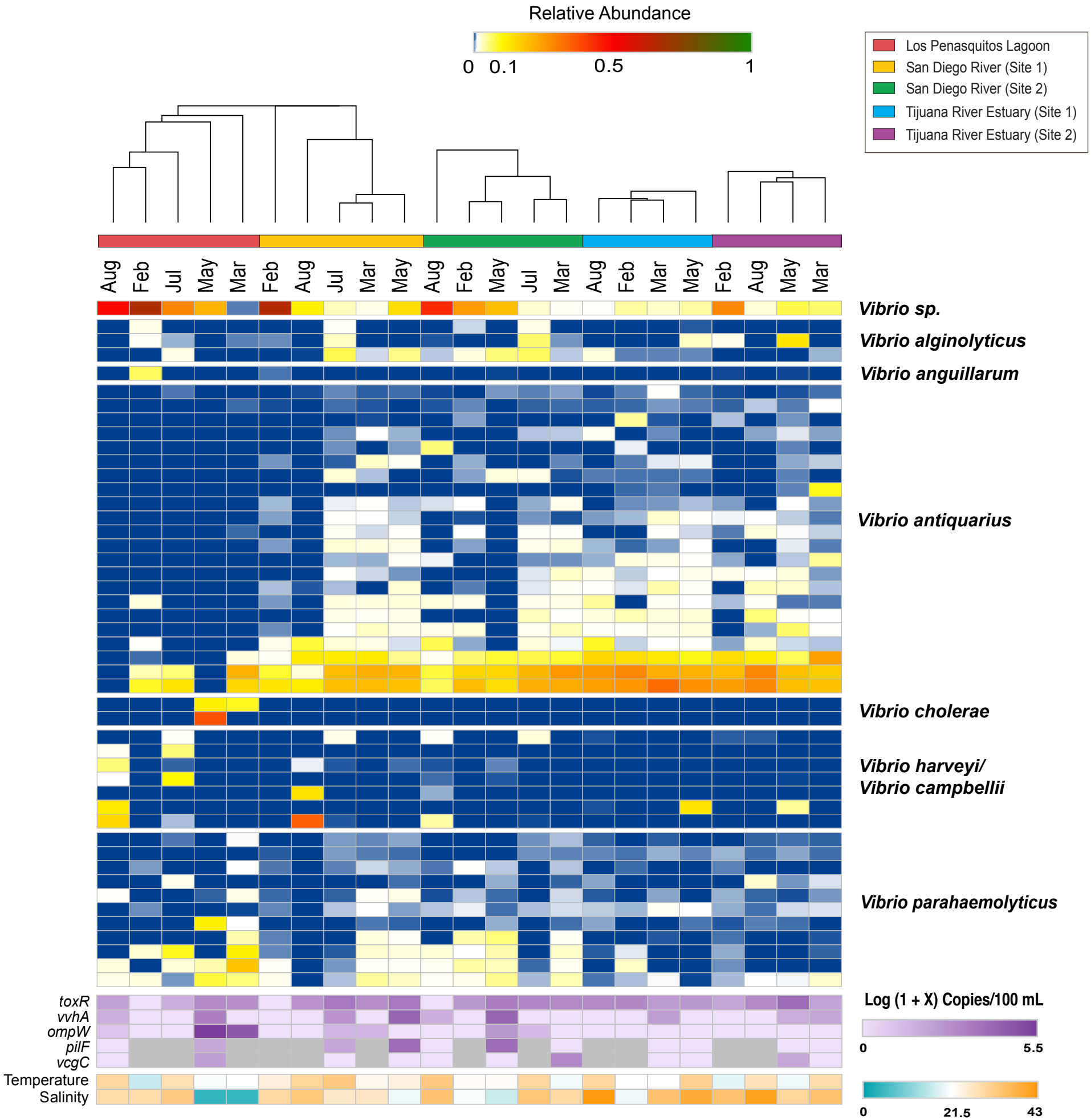

A
Virulence gene abundance
(log−normalized)
0.0 2.5 5.0 7.5
Virulence Genes of Interest
cytolysin (vvhA)
N−acetylglucosamine−binding protein A
hypothetical protein (vcgC/vgcE)
ToxR protein
protein VcsT2
protein VcsV2
thermolabile hemolysin (tlh)
thermostable direct hemolysin (tdh)
ace_Vp
RtxA_Vc2 (V. cholerae 623−39)
RtxA_Vc1 (V. cholerae O1 biovar El Tor str. N16961)
zot_Vp (V. parahaemolyticus)
outer membrane protein W (ompW)
vgrG protein
LPL
SDR1
SDR2
TJ1
TJ2
Feb Mar May Jul Aug
Month
B
ARG gene abundance
(log−normalized)
0 1 2
Top 20 Most Abundant ARG Classes
tetracycline
streptogramin
aminocoumarin
macrolide ; lincosamide
macrolide ; monobactam; tetracycline ; aminocoumarin
lincosamide
monobactam; carbapenem; cephalosporin; cephamycin; penam
fluoroquinolone
macrolide ; lincosamide ; streptogramin
macrolide ; fluoroquinolone ; penam
carbapenem; cephalosporin; penam
peptide
macrolide
glycopeptide
carbapenem
phenicol
cephalosporin
penam
cephalosporin; penam
aminoglycoside
LPL
SDR1
SDR2
TJ1
TJ2
Feb Mar May Jul Aug
Month

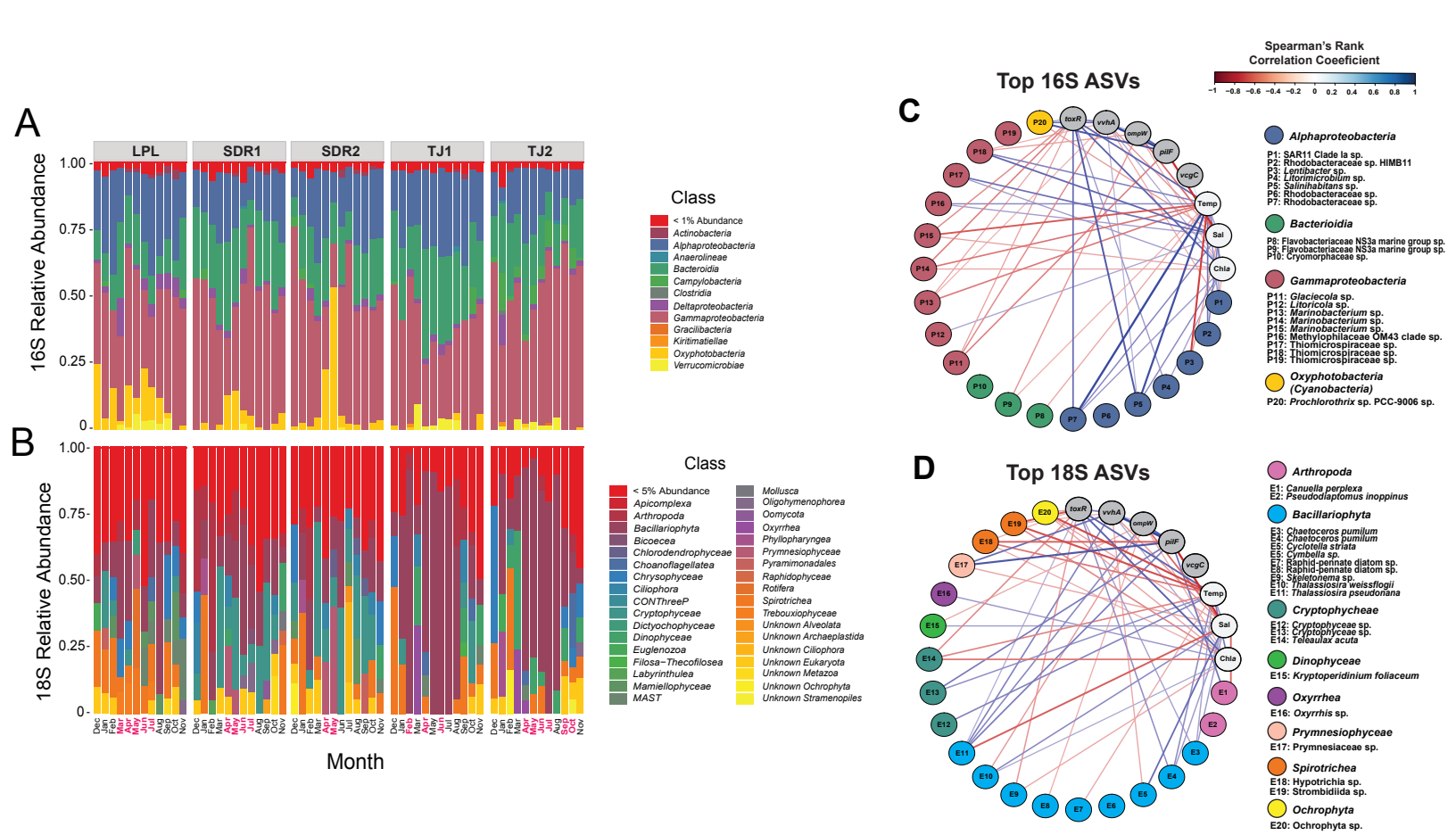

C
Resistance Gene Identifier by Site
ARG abundance
(log transformed counts)
0
1
2
LPL
SDR1
SDR2
TJ1
TJ2


D
Resistance Gene Identifier by Month
ARG abundance
(log transformed counts)
0
1
2
Feb
Mar
May
Jul
Aug

A: *Exiguobacterium* unclassified
B: Aeromonas hydrophila
C: Aeromonas veronii
D: *Shewanella* unclassified
E: Pseudomonas alcaligenes
F: Pseudomonas mendocina
G: *Pseudomonas* unclassified
H: *Enterovibrio* unclassified
I: Grimontia hollisae
J: Photobacterium damselae
K: Vibrio albensis
L: Vibrio alginolyticus
M: Vibrio anguillarum
N: Vibrio brasiliensis
O: Vibrio campbellii
P: Vibrio cholerae
Q: Vibrio cyclitrophicus
R: Vibrio furnissii
S: Vibrio harveyi
T: Vibrio kanaloae
U: Vibrio natriegens
V: Vibrio ordalii
W: Vibrio owensii
X: Vibrio parahaemolyticus
Y: Vibrio proteolyticus
Z: Vibrio rotiferianus
a: *Vibrio* sp 16
b: *Vibrio* sp 712i1
c: *Vibrio* sp Ex25
d: Vibrio splendidus
e: Vibrio vulnificus
f: Vibrionales bacterium SWAT 3
g: Vibrio phage VP882
h: Vibrio phage vB VpaM MAR




Relative Abundance
0
0.1
0.5
1
Los Penasquitos Lagoon
San Diego River (Site 1)
San Diego River (Site 2)
Tijuana River Estuary (Site 1)
Tijuana River Estuary (Site 2)
Aug Feb Jul May Mar Feb Aug Jul Mar May Aug Feb May Jul Mar Aug Feb Mar May Feb Aug May Mar
Vibrio sp.
Vibrio alginolyticus
Vibrio anguillarum
Vibrio antiquarius
Vibrio cholerae
Vibrio harveyi/
Vibrio campbellii
Vibrio parahaemolyticus
toxR
vvhA
ompW
pilF
vcgC
Temperature
Salinity
Log (1 + X) Copies/100 mL
0
5.5
0
21.5
43


A
LPL
SDR1
SDR2
TJ1
TJ2
16S Relative Abundance
1.00
0.75
0.50
0.25
0
Class
< 1% Abundance
Actinobacteria
Alphaproteobacteria
Anaerolineae
Bacteroidia
Campylobacteria
Clostridia
Deltaproteobacteria
Gammaproteobacteria
Gracilibacteria
Kiritimatiellae
Oxyphotobacteria
Verrucomicrobiae
B
18S Relative Abundance
1.00
0.75
0.50
0.25
0
Dec
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Month
Class
< 5% Abundance
Apicomplexa
Arthropoda
Bacillariophyta
Bicoecea
Chlorodendrophyceae
Choanoflagellatea
Chrysophyceae
Ciliophora
CONThreeP
Cryptophyceae
Dictyochophyceae
Dinophyceae
Euglenozoa
Filosa−Thecofilosea
Labyrinthulea
Mamiellophyceae
MAST
Mollusca
Oligohymenophorea
Oomycota
Oxyrrhea
Phyllopharyngea
Prymnesiophyceae
Pyramimonadales
Raphidophyceae
Rotifera
Spirotrichea
Trebouxiophyceae
Unknown Alveolata
Unknown Archaeplastida
Unknown Ciliophora
Unknown Eukaryota
Unknown Metazoa
Unknown Ochrophyta
Unknown Stramenopiles
C
Top 16S ASVs
Spearman's Rank
Correlation Coeefficient
−1
−0.8
−0.6
−0.4
−0.2
0
0.2
0.4
0.6
0.8
1
toxR
vvhA
ompW
pilF
vcgC
Temp
Sal
Chla
Alphaproteobacteria
P1: SAR11 Clade Ia sp.
P2: Rhodobacteraceae sp. HIMB11
P3: Lentibacter sp.
P4: Litorimicrobium sp.
P5: Salinihabitans sp.
P6: Rhodobacteraceae sp.
P7: Rhodobacteraceae sp.
Bacterioidia
P8: Flavobacteriaceae NS3a marine group sp.
P9: Flavobacteriaceae NS3a marine group sp.
P10: Cryomorphaceae sp.
Gammaproteobacteria
P11: Glaciecola sp.
P12: Litoricola sp.
P13: Marinobacterium sp.
P14: Marinobacterium sp.
P15: Marinobacterium sp.
P16: Methylophilaceae OM43 clade sp.
P17: Thiomicrospiraceae sp.
P18: Thiomicrospiraceae sp.
P19: Thiomicrospiraceae sp.
Oxyphotobacteria
(Cyanobacteria)
P20: Prochlorothrix sp. PCC-9006 sp.
D
Top 18S ASVs
Arthropoda
E1: Canuella perplexa
E2: Pseudodiaptomus inoppinus
Bacillariophyta
E3: Chaetoceros pumilum
E4: Chaetoceros pumilum
E5: Cyclotella striata
E6: Cymbella sp.
E7: Raphid-pennate diatom sp.
E8: Raphid-pennate diatom sp.
E9: Skeletonema sp.
E10: Thalassiosira weissflogii
E11: Thalassiosira pseudonana
Cryptophyceae
E12: Cryptophyceae sp.
E13: Cryptophyceae sp.
E14: Teleaulax acuta
Dinophyceae
E15: Kryptoperidinium foliaceum
Oxyrrhea
E16: Oxyrrhis sp.
Prymnesiophyceae
E17: Prymnesiaceae sp.
Spirotrichea
E18: Hypotrichia sp.
E19: Strombidiida sp.
Ochrophyta
E20: Ochrophyta sp.

Figures

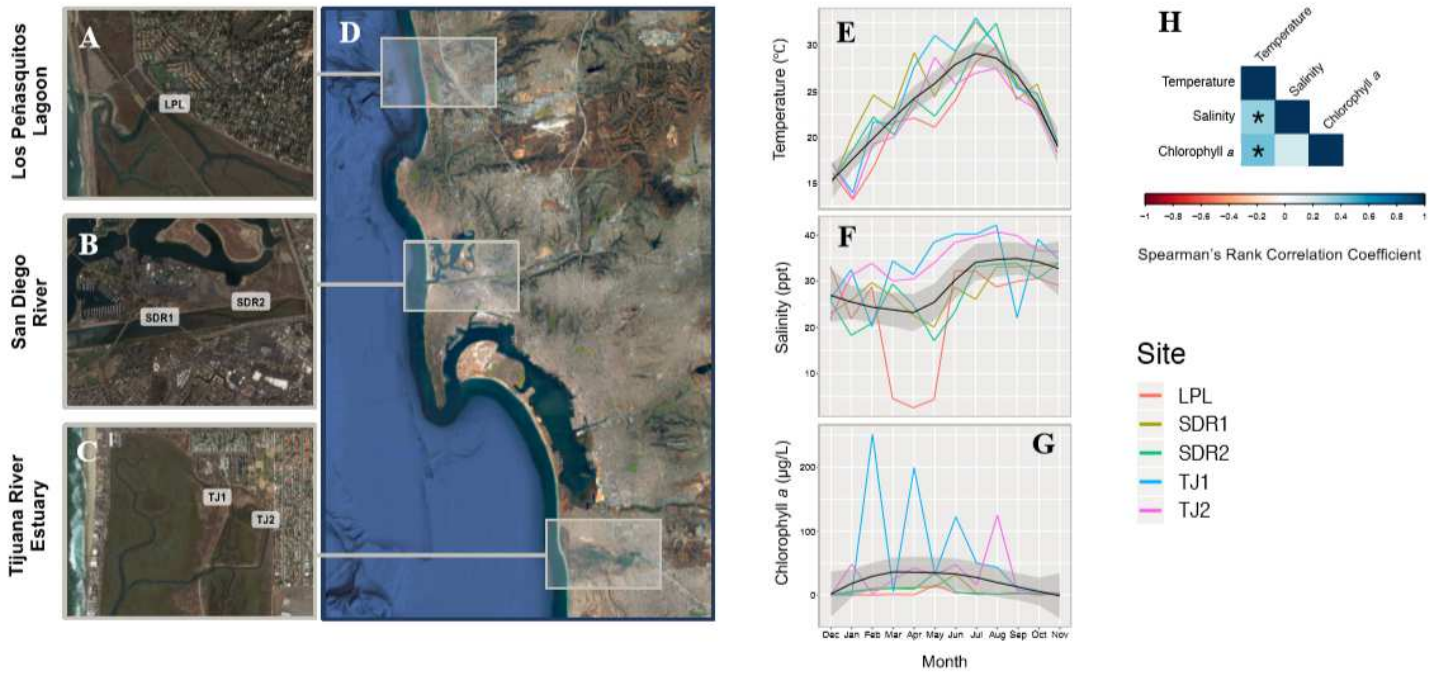


Figure 1

Location of the sampling sites and environmental conditions at the time of sampling. Locations are mapped in the context of the San Diego region using Google Earth. Site abbreviations are as follows: (A) LPL = Los Peñasquitos Lagoon, (B) SDR = San Diego River (site 1 and 2), and (C)TJ = Tijuana River Estuary (site 1 and 2), with (D) depicting the regional context. Environmental conditions, including (E) temperature, (F) salinity, and (G) chlorophyll a, a proxy for photosynthetic organism abundance, were measured monthly at each site (LPL = red, SDR1 = mustard, SDR2 = green, TJ1 = blue, and TJ2 = purple) for one year from December 2015-November 2016. Spearman rank correlations identified relationships between environmental variables, where values closer to 1 (dark blue) are positive correlations and values closer to -1 (dark red) are negative correlations, and * represents significant correlations (p-value < 0.05).

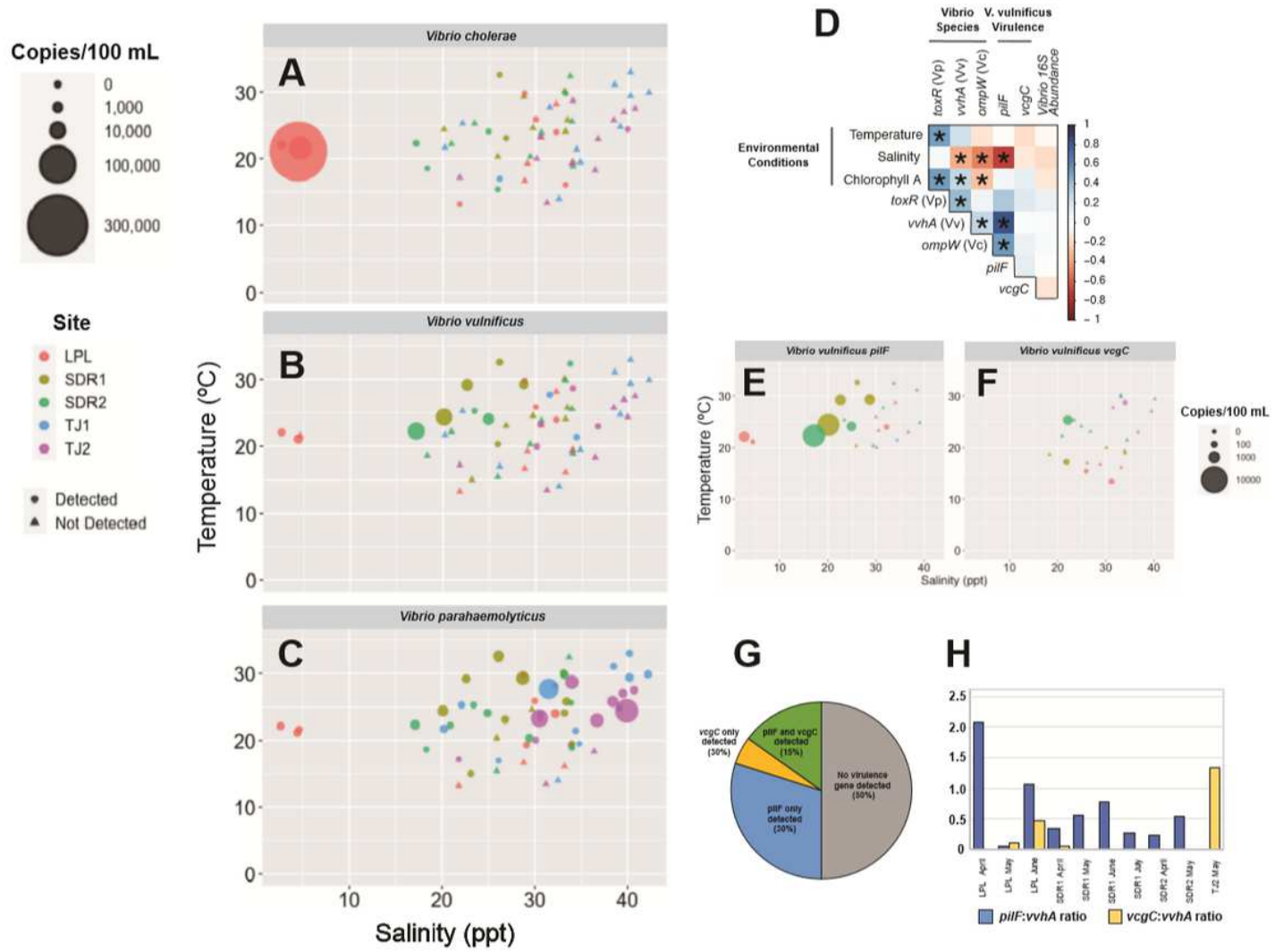


Figure 2

Number of single-genome copy genes (a proxy for cell numbers) per 100 mL detected by digital droplet PCR. (A) the Vibrio parahaemolyticus (Vp) species-specific gene target toxR (B) the Vibrio vulnificus (Vv) species-specific target vvhA, and (C) and the Vibrio cholerae species-specific target ompW, with marker size corresponding to copy number and color corresponding to site, plotted against temperature and salinity. (D) Spearman's rank correlation coefficients of associations between environmental the variables temperature, salinity, and chlorophyll a, and Vibrio species and virulence gene targets. Blue represents a strong positive correlation, while red represents a strong negative correlation, significant correlations (p-value < 0.05) are denoted with *. Number of copies detected per 100 mL by digital droplet PCR for the Vibrio vulnificus virulence-associated genes (E) vcgC and (F) pilF, plotted against temperature and salinity. (G) The percent of V. vulnificus samples where no virulence gene was detected, either vcgC or pilF were detected, or both were detected. (H) the ratio of the number of pilF and vcgC copies detected to total V. vulnificus determined by vvhA copy number.

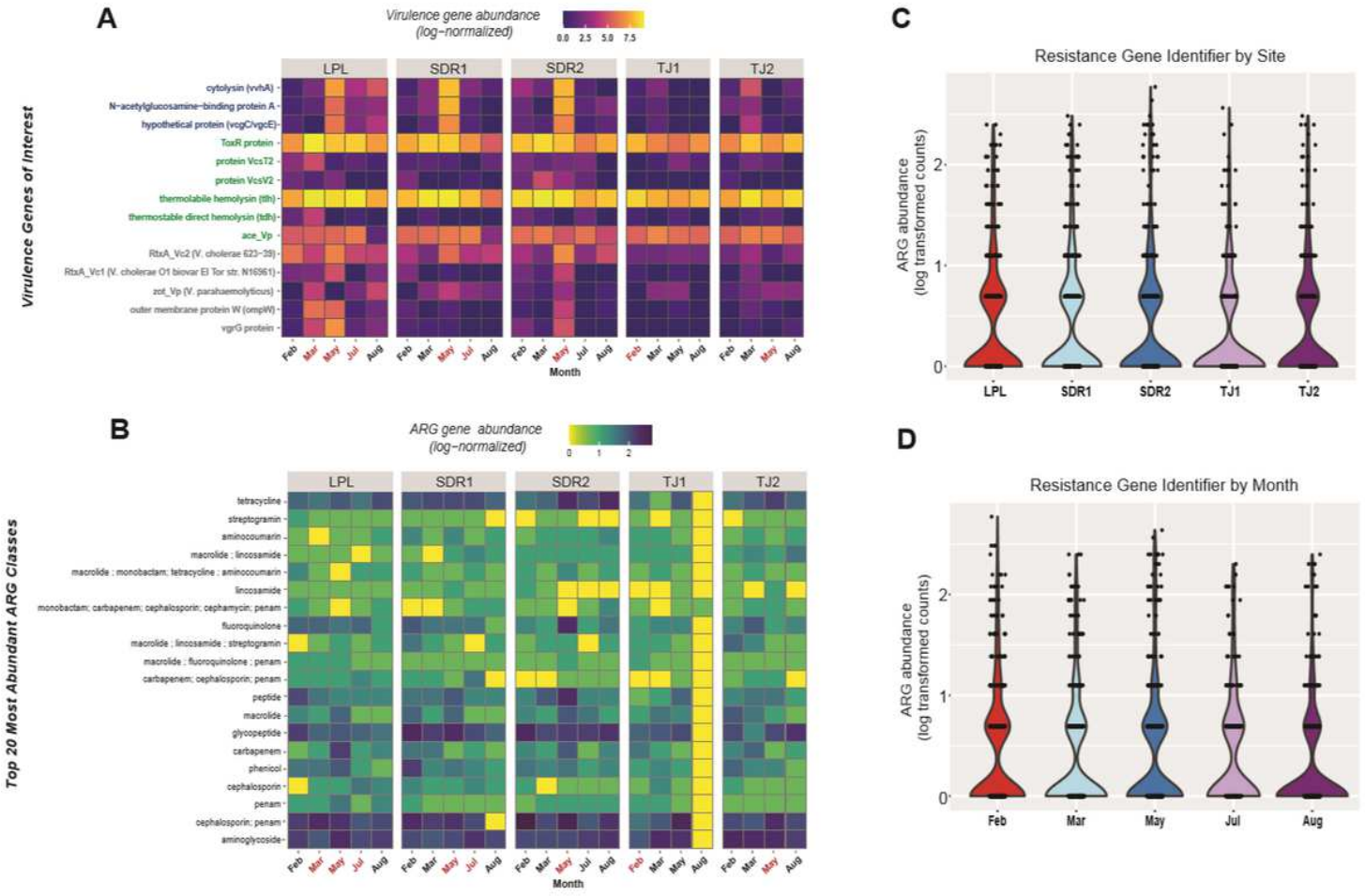


Figure 3

Virulence and antibiotic resistance gene profiles of Vibrio isolate communities based on shotgun metagenomic sequencing. Heatmap depicting (A) log normalized abundance of genes associated with virulence identified in the literature. On the Y axis, genes associated with V.vulnificus are in dark blue, genes associated with V. parahaemolyticus in green, and V. cholerae in grey (B) Heatmap showing log normalized abundance of the top 20 most abundant antibiotic resistance gene (ARG) classes as identified in the CARD database across months and sites. For (A) and (B), months where high concentrations of one or more pathogenic Vibrio sp. were detected by ddPCR are indicated in red on the X axis. Overall abundance of ARGs are shown by (C) site and (D) month.

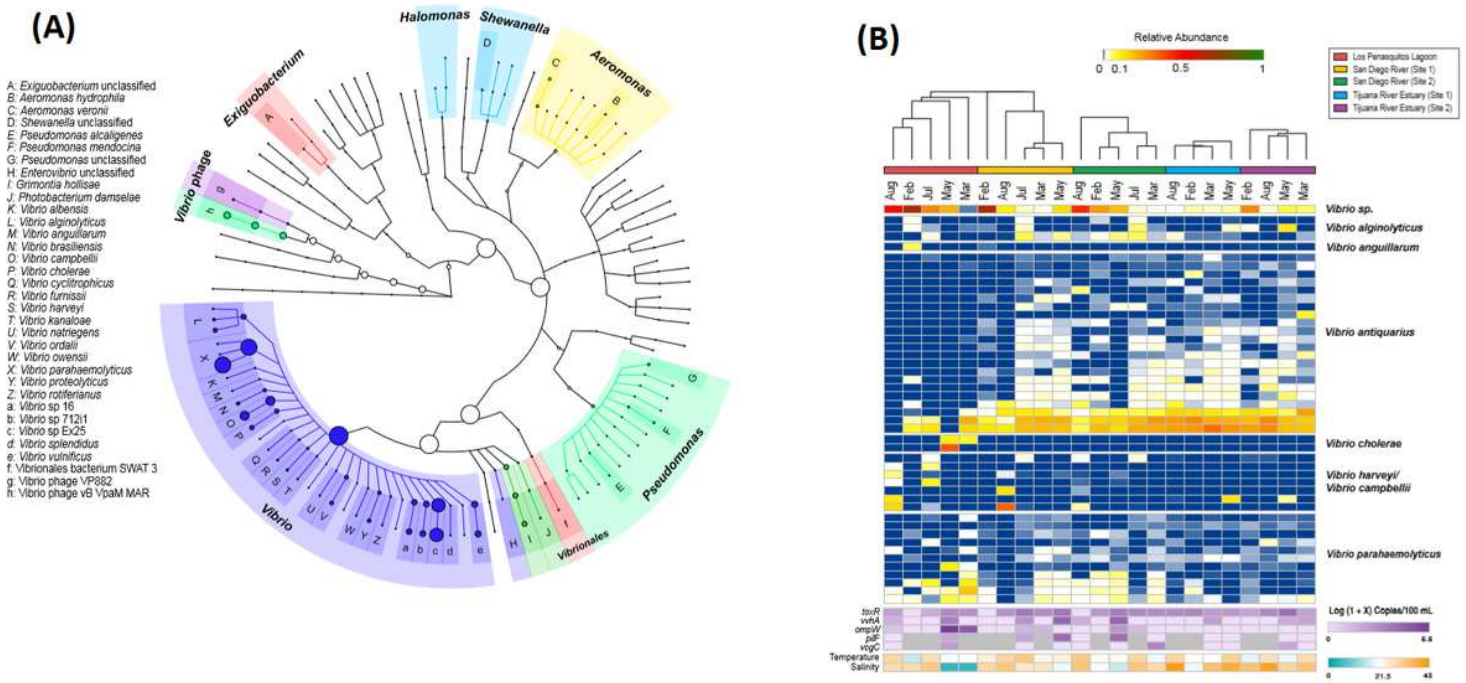


Figure 4

Diversity of isolated Vibrio species and related bacteria . (A) cladeogram depicting the taxonomy of species isolated from CHROMagar media identified by shotgun metagenomic sequencing. Taxonomic assignments were determined using MetaPhlAn2, which annotates and phylogenetically places unassembled sequence reads using clade-specific genetic markers, and (B) heatmap of the relative abundance of Vibrio spp. ASVs based on HSP60 amplicon sequencing and the associated abundance of quantified Vibrio target genes and environmental conditions (temperature and salinity). Hierarchical clustering is used to order samples based on Euclidean distance within sites so that samples within sites with similar abundance profiles are grouped near each other.

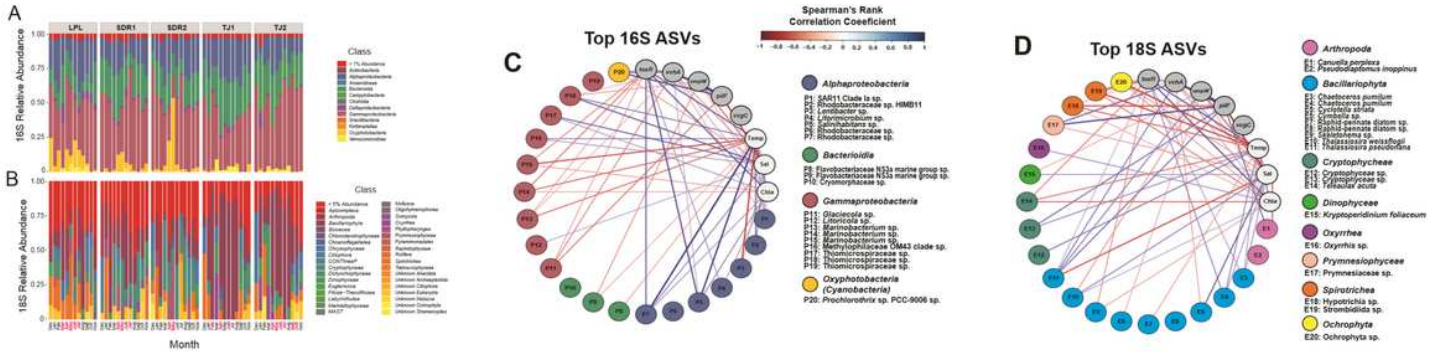


Figure 5

Taxonomic composition of prokaryotic (16S) and eukaryotic (18S) communities, and ASV-level associations between abundant prokaryotic and eukaryotic community members, Vibrio species and virulence gene targets, and environmental variables. Taxonomic composition of 16S (A) and 18S (B)

communities by class. Cut-offs of <1% and <5% abundance represent taxa cumulative abundance across all samples. Months in red indicate that one or more of the pathogenic Vibrio species quantified by ddPCR were present at concentrations exceeding 1000 copies/100 mL. Spearman's rank correlations between the top 20 most abundant (C) prokaryotic and (D) eukaryotic ASVs are organized by class, and only significant (p-value < 0.05) correlations are shown. Blue lines depict significant positive associations while red lines represent negative associations, and line thickness indicates the strength of the correlation. Marker genes used to quantify pathogenic Vibrio species are shown in grey circles, and correspond to the following species and virulence genes: toxR: V. parahaemolyticus species, vvhA: V. vulnificus species, ompW: V. cholerae species, pilF: V. vulnificus virulence-associated gene, vcgC = V. vulnificus virulence-associated gene.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AdditionalFile16Dineretal2020AmpliconAnalysisfigures.R
- AdditionalFile17Dineretal2020Qiime2analyses.ipynb
- AdditionalMaterialsDineretal2020Microbiome.docx
- AdditionalTable1MetaPhlAn2IsolateCultureTaxonomy.xls
- AdditionalTable2RelAbundIsolateVibrioHSP60.csv
- AdditionalTable3Adonisbetadispersionstats.xlsx
- AdditionalTable4Metadata.xlsx