AMERICAN SOCIETY FOR MICROBIOLOGY | Applied and Environmental Microbiology®

Environmental Determinants of *Vibrio parahaemolyticus* in the Chesapeake Bay

Benjamin J. K. Davis,[a] John M. Jacobs,[b] Meghan F. Davis,[a,c] Kellogg J. Schwab,[a] Angelo DePaola,[d] Frank C. Curriero[e]

Department of Environmental Health & Engineering, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, USA[a]; Cooperative Oxford Lab, National Centers for Coastal Ocean Science, National Ocean Service, National Oceanic and Atmospheric Administration, Oxford, Maryland, USA[b]; Department of Molecular and Comparative Pathobiology, Johns Hopkins School of Medicine, Johns Hopkins University, Baltimore, Maryland, USA[c]; Angelo DePaola Consulting, Coden, Alabama, USA[d]; Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, USA[e]

ABSTRACT *Vibrio parahaemolyticus* naturally occurs in brackish and marine waters and is one of the leading causes of seafood-borne illness. Previous work studying the ecology of *V. parahaemolyticus* has often been limited in geographic extent and lacked a full range of environmental measures. This study used a unique large data set of surface water samples in the Chesapeake Bay ($n = 1,385$) collected from 148 monitoring stations from 2007 to 2010. Water was analyzed for more than 20 environmental parameters, with additional meteorological and surrounding land use data. The *V. parahaemolyticus*-specific genetic markers thermolabile hemolysin (*tlh*), thermostable direct hemolysin (*tdh*), and *tdh*-related hemolysin (*trh*) were assayed using quantitative PCR (qPCR), and interval-censored regression models with nonlinear effects were estimated to account for limits of detection and quantitation. *tlh* was detected in 19.6% of water samples; *tdh* or *trh* markers were not detected. The results confirmed previously reported positive associations for *V. parahaemolyticus* abundance with temperature and turbidity and negative associations with high salinity (>10 to 23‰). Furthermore, the salinity relationship was determined to be a function of both low temperature and turbidity, with an increase of either nullifying the high salinity effect. Associations with dissolved oxygen and phosphate also appeared stronger when samples were taken near human developments. A renewed focus on the *V. parahaemolyticus* ecological paradigm is warranted to protect public health.

IMPORTANCE *Vibrio parahaemolyticus* is one of the leading causes of seafood-borne illness in the United States and across the globe. Exposure is often through consuming raw or undercooked shellfish. Given the natural presence of the bacterium in the marine environment, an improved understanding of its environmental determinants is necessary for future preventative measures. This analysis of environmental *Vibrio parahaemolyticus* is one of only a few that utilize a large data set measured over a wide geographic and temporal range. The analysis also includes a large number of environmental parameters for *Vibrio* modeling, many of which have previously only been tested sporadically, and some of which have not been considered before. The results of the analysis revealed previously unknown relationships between salinity, turbidity, and temperature that provide significant insight into the abundance and persistence of *V. parahaemolyticus* bacterium in the environment. This information will be essential for developing environmental forecast models for the bacterium.

KEYWORDS Chesapeake Bay, spatial, temporal, *Vibrio parahaemolyticus*, biostatistics, environmental microbiology, public health

*V*ibrio parahaemolyticus is a Gram-negative, halophilic, and facultative anaerobic bacterium that is autochthonous to estuarine environments. While only some strains of the bacterium are pathogenic, it is one of the most common agents of seafood-borne illnesses, causing an estimated 34,644 domestically acquired illnesses each year in the United States (1). Symptoms of infection are typically restricted to gastroenteritis, although life-threatening septicemia can also occur. The annual incidence rate of *V. parahaemolyticus* infections has increased, with the U.S. Centers for Disease Control and Prevention's Foodborne Disease Active Surveillance Network reporting an increase from 0.15 per 100,000 people in 1996 to 0.42 in 2010, and the Cholera and Other *Vibrio* Illness Surveillance system reporting an increase from 0.09 to 0.28 in the same time period (2). This trend is expected to continue as water temperatures rise globally due to general climate warming, particularly in high latitude bodies of water such as the Chesapeake Bay. This warming will likely expand the spatial-temporal extent of the bacterium in the environment as well as potentially increase the risk of infection (3–6).

Exposure to *V. parahaemolyticus* most frequently results from the consumption of raw or undercooked shellfish. Bivalve mollusks, through their filter-feeding behavior, are able to accumulate the bacterium before they are harvested. As *Vibrio* spp. cannot be eliminated from the environment, their increasing presence in coastal waters worldwide poses a serious public health concern for consumers and for the shellfish harvesting industry. There has therefore been substantial interest in understanding the biotic and abiotic determinants of *V. parahaemolyticus*, as such information would prove useful in forecasting its abundance in shellfish harvesting waters.

Many ecological studies of *V. parahaemolyticus* have reported associations with a number of environmental factors (7, 8). Water temperature has consistently been found to be an influential predictor, with greater environmental abundance and reported illnesses occurring during warmer months. Additional studies have proposed an association with chitin-producing phyto- and zooplankton (9, 10). Narrow ranges of salinity have been frequently observed in the literature, resulting in both negative and positive correlations with salinity (8). When measured across a sufficient observable range, studies have instead observed an "optimal" salinity (10 to 23‰) for the bacterium, such that abundance decreases as waters become too fresh or saline (11–14). However, these findings are incongruent with experimental studies that describe *V. parahaemolyticus* persisting at salinities even higher than those observed in marine waters (15, 16). Low dissolved oxygen (DO) and high turbidity have been sporadically identified as potential determinants, and a few studies have explored associations with nutrients, specifically, forms of nitrogen, phosphorus, and carbon (8).

While the established literature has provided a rich background for understanding many of the potential environmental determinants for *V. parahaemolyticus*, it is often difficult to identify consistent associations across studies, although the association between warmer waters and a greater presence and abundance is a notable exception. Previous environmental undertakings vary significantly in their spatial-temporal extent and resolution, as do the number of environmental parameters measured and their respective observable ranges (8). Furthermore, many studies only consider a small number of environmental measures or lack the sample size to test multiple associations in a parametric statistical model, making an identification of redundant environmental associations difficult.

A large data set of water samples in the Chesapeake Bay collected across multiple seasons and monitoring stations between 2007 and 2010 (17) provides a unique opportunity to further harmonize the ecological paradigm of *V. parahaemolyticus* in the water column. These samples were analyzed using a quantitative PCR (qPCR) method to detect the genetic markers thermolabile hemolysin (*tlh*), thermostable direct hemolysin (*tdh*), and *tdh*-related hemolysin (*trh*). As this method was not paired with a culture technique, abundance is reported as the genomic equivalents of CFU per milliliter (GE/ml). Samples were supplemented with a host of *in situ* water quality measurements, providing the opportunity to test the redundancy and robustness of many potential
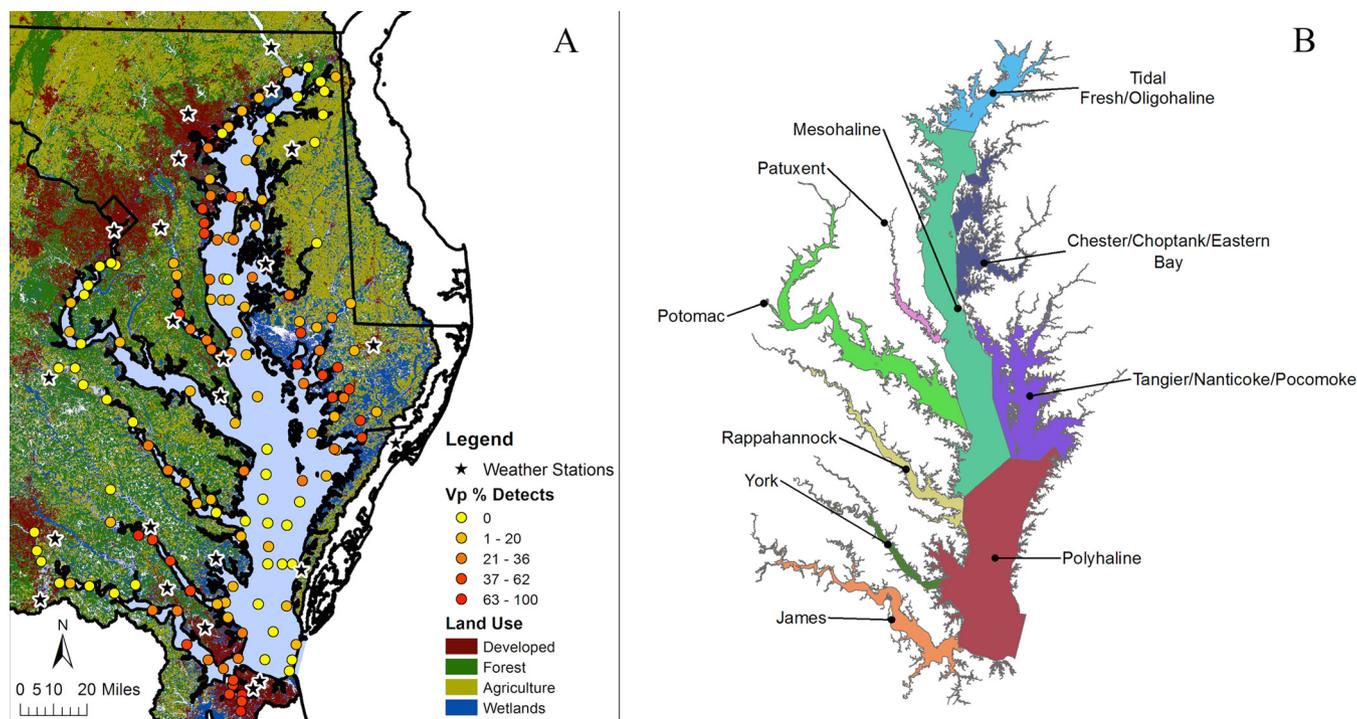
FIG 1 (A) NOAA water sampling sites with *V. parahaemolyticus* detection rates (Vp % Detects), weather stations, and watershed land use. (B) Ten regional categories used for spatial analysis. Overall salinity levels distinguished three main stem regions. Five regions on the western shore each encapsulated a major tributary, and two remaining regions aggregated smaller tributaries on the eastern shore.

environmental determinants. The current analysis also introduced a sophisticated form of parametric regression modeling that can simultaneously account for the limits of detection and quantitation from qPCR results. Linear B-splines were included in models to identify nonlinear associations while preserving easy interpretation of associations. This work also explored previously uninvestigated spatial and temporal interactions and the residual variation of environmental determinants on the abundance of *V. parahaemolyticus*. The analyses presented in this paper therefore offer a robust and original effort to evaluate a wide array of environmental determinants across a significant spatial and temporal range and to extend the literature on the ecology of *V. parahaemolyticus*.

RESULTS

The numbers of water samples taken across years were relatively similar, with 335 samples taken in 2007, 419 in 2008, 375 in 2009, and 394 in 2010. Similar variations in sample numbers were seen across seasons, with 479 samples taken in the spring (April), 551 in summer (July), and 493 in autumn (October). The numbers of samples across different regions of the Chesapeake Bay were more variable, with fewer samples taken in the narrow tributaries and more samples taken in the main stem of the bay (e.g., 75 samples taken in the York River versus 302 samples taken in the mesohaline region of the main stem) (Fig. 1). The numbers of samples collected at individual sampling stations ranged from 2 to 12, with an average of 10.3 samples per station.

For the extraction protocol, recovery estimates averaged (with standard deviation [SD]) 46.17% ± 7.55% (*n* = 40) of starting DNA over a 6-log range. High repeatability was found in both replicate water samples from the same station (R^2 = 0.92, *n* = 32) and replicate samples from the same source (R^2 = 0.94, *n* = 34). Freezing did not significantly affect the measured *V. parahaemolyticus* concentrations, with split samples demonstrating a 1:1 relationship (β = 1.10, R^2 = 0.84, *n* = 20). Standard curves of cycle threshold (C_T) values versus concentrations yielded an assay efficiency (with SD) of 88.60% ± 5.90% (*n* = 4), with a sensitivity of <1 GE/ml in a 200-ml water sample. No

TABLE 1 Descriptive characteristics of *V. parahaemolyticus* and environmental determinants, as well as their variation across sampling months

Characteristic	No. of samples	Overall[a]	Season[b] Spring (April)	Summer (July)	Autumn (October)
V. parahaemolyticus (% [*n*] detected)	1,523	19.6 (298)	2.7 (13)	28.3 (156)	26.2 (129)
V. parahaemolyticus (GE/ml for detected samples only)	298	2.6 (1.1–143.6)	0.8 (0.2–1.6)	3.0 (1.2–6.2)	2.6 (1.1–5.8)
Land/water					
Water temperature (°C)	1,523	19.1 (7.1–32.1)	13.8 (11.3–15.6)	27.3 (26.4–28.1)	17.9 (15.6–20.7)
Salinity (‰)	1,522	10.2 (0–27.1)	7.1 (0.1–12.7)	10.6 (1.4–14.9)	14.3 (4.1–18.3)
pH	1,523	7.9 (6.2–9.5)	7.9 (7.5–8.1)	7.8 (7.5–8.1)	7.9 (7.6–8.0)
Bathymetry (m)[c]	1,523	7.5 (1.5–35)	—[d]	—	—
%Water[c]	1,523	39.0 (1.5–100.0)	—	—	—
Abiotic water quality					
Total suspended solids (mg/liter)	1,517	9.0 (2.2–198.0)	9.0 (5.4–20.0)	9.4 (6.5–18.0)	8.4 (5.4–15.0)
Secchi depth (m)	1,497	0.8 (0.1–4.4)	0.8 (0.4–1.4)	0.8 (0.5–1.1)	1.0 (0.6–1.5)
Dissolved oxygen (mg/liter)	1,510	8.1 (3.1–17.7)	9.5 (8.6–10.6)	6.9 (6.0–7.6)	8.3 (7.3–9.0)
Dissolved organic nitrogen (mg/liter)	1,508	0.3 (0.1–1.6)	0.3 (0.2–0.4)	0.3 (0.2–0.4)	0.3 (0.2–0.4)
Dissolved organic phosphorus (µg/liter)	1,508	9.0 (0.0–87.5)	6.8 (4.3–10.0)	9.9 (7.0–14.0)	10.7 (7.0–15.1)
Ammonium (µg/liter)	1,513	19.8 (1.3–401.0)	28.0 (12.0–64.2)	14.0 (6.0–29.5)	17.7 (7.0–35.0)
Nitrate (µg/liter)	1,487	32.6 (0.0–3,143.0)	330.8 (66.8–700.2)	4.0 (1.6–31.5)	28.0 (4.23–167.6)
Nitrite (µg/liter)	1,486	4.0 (0.1–165.0)	6.7 (2.63–10.18)	2.0 (0.6–3.7)	7.85 (2.0–18.0)
Phosphate (µg/liter)	1,513	6.0 (0.4–204.0)	3.2 (2.2–11.3)	7 (3.7–19.0)	7.6 (3.1–22.1)
Dissolved inorganic N:P[e]	1,513	10.1 (0.2–1,610.0)	65.3 (20.9–170.0)	3.9 (1.7–9.8)	8.2 (4.2–17.5)
Particulate nitrogen (mg/liter)	1,522	0.2 (0.0–2.1)	0.2 (0.1–0.3)	0.3 (0.2–0.4)	0.2 (0.1–0.3)
Particulate phosphorus (µg/liter)	1,522	24.2 (4.2–494.6)	21.5 (12.5–41.8)	31.0 (22.0–50.8)	16.9 (12.1–32.3)
Biotic water quality					
Chlorophyll *a* (mg/liter)	1,471	10.0 (0.3–124.4)	8.9 (4.8–16.3)	12.0 (8.7–19.7)	7.9 (5.0–13.8)
Pheophytin (mg/liter)	1,498	1.6 (0–44.5)	1.4 (0.6–3.4)	2.1 (0.6–4.6)	1.3 (0.6–3.5)
Moderate-to-heavy rainfall (% [*n*])[f]					
Lag 0	1,520	2.9 (44)	3.1 (15)	3.5 (19)	2.0 (10)
Lag 1	1,519	3.4 (52)	1.9 (9)	6.6 (36)	1.4 (7)
Lag 3	1,518	3.5 (53)	2.1 (10)	4.4 (24)	3.9 (19)
Lag 7	1,521	1.8 (28)	1.3 (6)	0.5 (3)	3.9 (19)
Cum. 3	1,523	16.4 (250)	16.5 (79)	19.4 (107)	13.0 (64)
Cum. 7	1,523	3.5 (53)	1.9 (9)	3.1 (17)	5.5 (27)

[a]Values are medians and ranges, unless otherwise indicated.
[b]Values are medians and interquartile ranges, unless otherwise indicated.
[c]Does not vary seasonally.
[d]—, not applicable.
[e]Ratio of nitrogen to phosphorus.
[f]"Lag" indicates the exact number of days rainfall was measured before water samples were taken (e.g., "Lag 3" indicates rainfall was measured 3 days before a water sample was taken). "Cum." indicates the cumulative amount of lagged rainfall measurements (e.g., "Cum. 3" indicates that rainfall measurements were summed from 3 days prior up until and including the day of sampling).

inhibitors were observed in any environmental samples, based on the amplification of the internal control.

Rates of *V. parahaemolyticus* were relatively low in the water samples, with only 19.6% of samples containing any detectable concentration of *tlh* genetic material (Table 1). Among samples with detectable concentrations, the average was 2.6 GE/ml. Presence and abundance were both higher in summer and autumn, but with little variation between the two. The presence of genetic material was notably higher in 2007 and 2010 and low in 2008. The abundance was only substantially higher in 2010. Detection rates also varied significantly by area, with some of the highest rates of detection occurring in the mesohaline sections of the James and Rappahannock Rivers, as well as in many of the Tangier/Nanticoke/Pocomoke region's smaller tributaries (Fig. 1).

As expected, water temperature was the highest during the summer and lowest during the spring (Table 1). In contrast, salinity was lowest in the spring, likely due to annual freshwater inflow from snowmelt, and highest during the autumn. All nitrogen ion levels were much higher in the spring months, as were ratios of dissolved inorganic nitrogen and phosphorus (DINP). Chlorophyll *a* was highest in the summer, when algal

TABLE 2 Comparison of univariate (unadjusted) and multivariate (adjusted) analyses for the main effects of abundance of *V. parahaemolyticus* (interval-censored regression)[a]

Environmental determinant	RCGM	
	Univariate	Multivariate (95% CI)
Water characteristics		
Temperature (°C)[b]	**1.30**	**1.17 (1.09–1.27)**
Salinity (‰)		
0–10.2	**1.05**	**1.09 (1.07–1.1)**
10.2–27.1	**0.98**	1.01 (1.00–1.03)
pH	**0.82**	—[c]
%Water[b]	**0.97**	**0.95 (0.93–0.96)**
Abiotic water quality		
Secchi depth (m)	**0.82**	**0.8 (0.74–0.87)**
Dissolved oxygen (mg/liter)		
3–6	**0.49**	**0.63 (0.56–0.7)**
6–18	**0.93**	1.04 (0.99–1.08)
Dissolved organic nitrogen (mg/liter)	**1.79**	**1.29 (1.01–1.66)**
Dissolved organic phosphorus (μg/liter)[b]		
0–20	**1.22**	1.07 (1–1.14)
20–90	**0.92**	**0.89 (0.81–0.98)**
Ammonium (μg/liter)	1.00	—
Nitrate (μg/liter)	**0.99**	—
Nitrite (μg/liter)[b]		
0.0–0.91	1.26	1.74 (0.42–7.26)
0.91–170	**1.06**	**1.04 (1.01–1.06)**
Phosphate (μg/liter)[b]	**1.10**	**1.04 (1.02–1.07)**
Dissolved inorganic N:P	**0.99**	—
Biotic water quality		
Chlorophyll *a* (mg/liter)[b]	1.00	**0.96 (0.94–0.99)**
Pheophytin (mg/liter)		
0–5	**1.07**	**1.03 (1.01–1.06)**
5–45	**0.98**	1 (0.98–1.01)
Land use		
Forest (REF)[d]		
Developed	**1.26**	—
Agriculture	**0.87**	—
Wetlands	**1.12**	—
Moderate-to-heavy rainfall[e]		
Lag 0	1.17	—
Lag 1	**1.54**	**1.31 (1.11–1.55)**
Lag 7	**1.36**	**1.43 (1.15–1.78)**

[a]RCGM, relative change in geometric mean; N:P, ratio of nitrogen to phosphorus. Boldface font indicates $P < 0.10$ in univariate analysis and $P < 0.05$ in multivariate analysis. Linear combination tests were calculated for B-spline associations.
[b]Changes in units are expressed in increments of 10.
[c]—, not applicable.
[d]Forest was the reference level for the categorical variable of land use.
[e]Lag indicates the exact number of days rainfall was measured before water samples were taken (e.g., "Lag 3" indicates rainfall was measured 3 days before a water sample was taken).

blooms are likely to occur, and so it is not surprising to find that dissolved oxygen (DO) was lower during this season (Table 1). Many of the sampling stations were surrounded mostly by forest (46.8%) and by wetlands (25%). Daily wind speed averages were found to be positively associated with turbidity (as measured by Secchi disk depth; $P < 0.0001$).

After removing the samples with missing environmental data, 1,385 complete case samples were used for regression modeling. Almost all environmental determinants appeared to be significantly related to *V. parahaemolyticus* in the univariate models (Table 2). A clear positive association was maintained for water temperature across its observable range, with few detectable values observed under 15°C. Salinity displayed a nonlinear relationship, showing a strong positive association as it rose from fresh-

TABLE 3 Comparison of AIC across different interval-censored regression models for the abundance of *V. parahaemolyticus*[a]

Model No.	Model name	AIC
1	Null model (intercept only)	4,056
2	Main effects model (Table 2)	3,036
3	Model 2 + region	3,032
4	Model 2 + season	3,029
5	Model 2 + year	2,956
6	Model 2 + "year * season"	2,926
7	Model 6 + "season * salinity"	2,896
8	Model 6 + "WtempQ * season * salinity" + "TurbQ * salinity"	2,842
9	Model 8 + "year*DO" + "year*PO4"	2,786
10	Model 9 + "LU*DO" + "LU*PO4"	2,777

[a]AIC, Akaike information criterion; WtempQ, quartiles of water temperature; TurbQ, quartiles of turbidity (Secchi disk depth); LU, land use; DO, dissolved oxygen; PO4, phosphate.

water to 10.2‰ (relative change in geometric mean [RCGM] = 1.05), after which the abundance appeared to decrease (e.g., RCGM = 0.98). Dissolved organic nitrogen (DON), phosphate, and turbidity each had positive relationships with *V. parahaemolyticus* abundance, while nitrate, DINP, pH, and the percentage of pixels classified as water (%Water) had negative relationships. Land use had significant associations, with the proximity to developed lands indicating the greatest abundance rates and the proximity to agriculture being associated with the lowest concentrations overall. DO displayed a strong negative relationship with abundance up to 6 mg/liter, after which, a more gradual slope was observed. Additional nonlinear associations were observed for dissolved organic phosphorus, nitrite, and pheophytin. Moderate-to-heavy rainfalls 1 and 7 days prior were also significantly positively associated with the abundance of genetic material (Table 2).

Water temperature, %Water, DON, nitrite, and phosphate were not substantially affected by adjustments in the multivariate models (Table 2). Many environmental determinants, including pH, nitrate, DINP, and surrounding land use, were no longer statistically significant in multivariate models. In an effort to improve upon model simplicity, these variables were excluded from later multivariate models (Table 2).

The positive association across low levels of salinity was strengthened in the multivariate model (RCGM = 1.09; 95% confidence interval [CI], 1.07 to 1.10) (Table 2). However, there was no longer a statistically significant association with *V. parahaemolyticus* abundance at salinity concentrations above 10.2‰ (RCGM = 1.01; 95% CI, 0.99 to 1.03). Upon further investigation, the adjusted slopes were mostly influenced by the inclusion of the Secchi disk depth variable, which became even more negatively associated in the multivariate model (RCGM = 0.80; 95% CI, 0.74 to 0.87). For DO, the negative association with values less than 6 mg/liter was attenuated but still statistically significant; there was no longer a significant association above this threshold. Chlorophyll, despite having no association with abundance in the univariate analysis, was found to have a statistically significant negative association in the multivariate model (Table 2).

The inclusion of variables for season and year, but not region, substantially improved the model fit (Table 3, models 3 to 6). The inclusion of year also resulted in the DON association becoming nonsignificant ($P > 0.70$). Some spatial and temporal interactions were also observed in sensitivity analyses (Table 3, models 9 and 10). The most noteworthy interaction occurred between season and salinity (Fig. 2; Table 3, model 7). During the summer and above 10.2‰, the abundance of *V. parahaemolyticus* leveled off as was seen in the multivariate model. However, during autumn, the association with salinity displayed a more symmetric association, with the largest abundance of genetic material co-occurring at the 10.2‰ threshold. Further investigation revealed that this effect modification was highly dependent upon water temperature and turbidity. In warmer (>27°C) and in more turbid waters, there was no drop in abundance even at the highest salinities (Fig. 2; Table 3, model 8). Additional
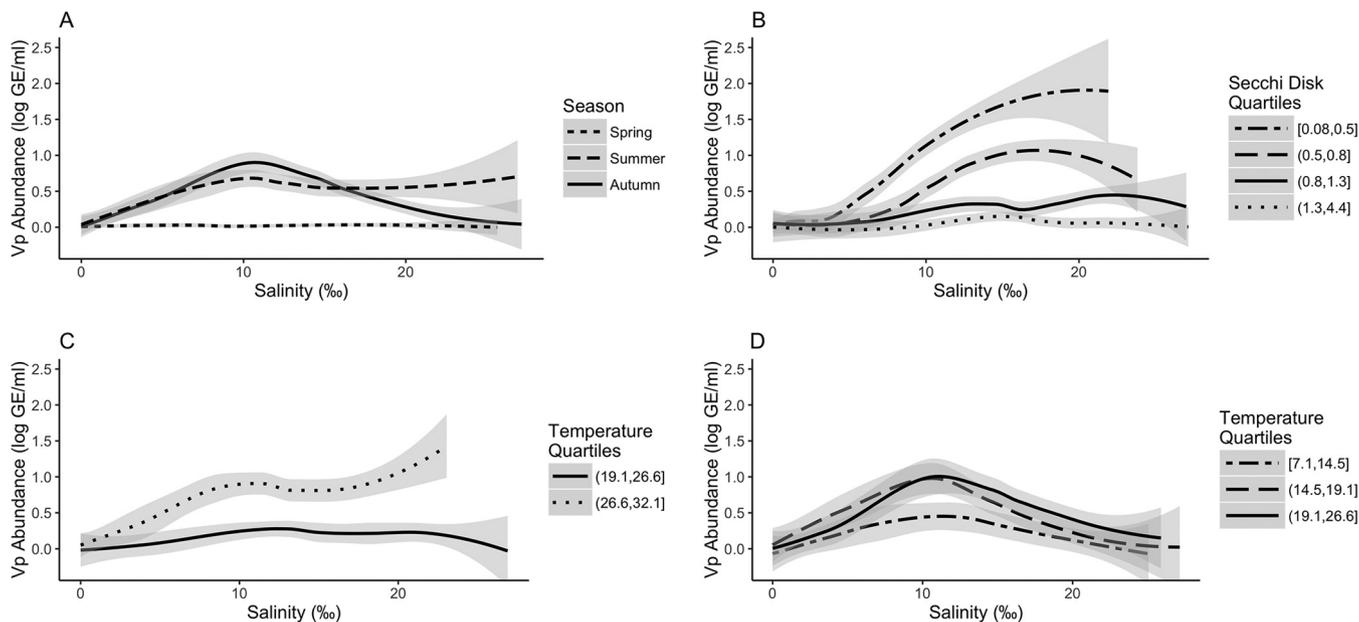
FIG 2 Association between *V. parahaemolyticus* abundance (log-transformed GE/ml) and salinity (‰) stratified by sampling season (A), Secchi disk quartiles (B), temperature quartiles in summer (C), and temperature quartiles in autumn (D). Note that panel C includes only the 3rd and 4th quartiles of temperature, while panel D includes the 1st through 3rd quartiles. Lines and 95% confidence bands were created using the local smoothing regression "LOESS."

interactions revealed that the associations of DO and phosphate were modified substantially by the year of sampling and by land use ($P < 0.05$ for all interaction terms) (Table 3, model 10). Specifically, both associations were most pronounced when monitoring stations were largely near human developments.

Prior to modeling, a multicollinearity analysis revealed many strong similarities across environmental determinants, notably for forms of nitrogen and phosphorus, for multiple rainfall measurements, and for air and water temperature (see Table S1 in the supplemental material). Representative variables were selected *a priori* for a regression analysis. An exploratory analysis revealed further redundancies between bathymetry and %Water (see Fig. S1). Given that the bathymetry variable was pseudocontinuous, %Water was used for model fitting. Similar redundancies were found for the Secchi depth and total suspended solids; the Secchi depth was used as it had fewer missing observations. Many rainfall measurements were also redundant, and so only variables for the day of sampling and 1 and 7 days prior were retained. Particulate nitrogen and phosphorus were excluded, as both were redundant given the variability of chlorophyll and pheophytin. No multicollinearity was identified in the final model.

Residuals showed a greater lack of fit (LOF) for summer and autumn. In addition to greater LOF in 2007 and 2010, it appears that *V. parahaemolyticus* abundance in 2010 may have been underestimated ($\bar{X} = 0.17$, standard error [SE] $= 0.04$). Model residuals also indicated there was an underestimation in the main stem of the Chesapeake Bay and overestimations in the Patuxent and Rappahannock regions (see Table S2). Finally, while spatial dependence was observed for abundance, semivariogram plots revealed no residual spatial variation in the multivariate models.

DISCUSSION

This *V. parahaemolyticus* ecological study examines a wide array of environmental determinants in a sizable data set with significant spatial and temporal extent. The findings have helped unify the ecological paradigm of *V. parahaemolyticus* in the water column by confirming many previously reported associations with environmental determinants while also identifying redundancies across such measures. The results from these analyses also revealed more complex associations that had not been

identified before. Given the Chesapeake Bay's unique hydrography, future work will be needed to further elucidate these findings.

The strong positive associations with water temperature and with turbidity are consistent across many previous studies (7, 8, 13). *V. parahaemolyticus* has been proposed to attach to chitin-producing biota such as zooplankton (9, 10). However, *V. parahaemolyticus* has also been found to attach to sediment and so may also be attached to particles suspended in the water column (13, 18). It seems likely that the higher turbidity observed in this analysis is from resuspended sediments where *V. parahaemolyticus* was already residing, given the measure's strong correlation with wind speed and a lack thereof with the seasonal variation of nutrients. The consistency of these results and the strength of the association with turbidity are of great interest and improve the overall understanding of *V. parahaemolyticus* abundance in the water column.

While the parabolic trend for salinity observed in the univariate analysis was expected, the impact of adjusting for turbidity at higher levels of salinity reveals a relationship that has not been identified before. Previous studies that have reported a nonlinear relationship between salinity and *V. parahaemolyticus* frequently used quadratic terms in their models, which force a parabolic relationship (11, 19, 20). The use of linear B-splines in the present analyses provided the flexibility to observe unrestricted nonlinear associations. Salts can cause a reduction in turbidity by aggregating suspended solids, which then deposit out of the water column (21). In addition, the more saline waters of the Chesapeake Bay are found closer to the mouth of the bay and so are often farther away from watershed inputs that can increase turbidity. The low densities in the main stem associated with higher salinities are also likely influenced by the proximity to shorelines, which was observed in the current analysis using the %Water variable. These areas may be preferred for relaying oysters to reduce *V. parahaemolyticus* densities relative to the more protected areas where oysters are normally cultured.

The interaction between water temperature and salinity, in addition to turbidity and salinity, provides great insight into the bacterium's persistence in the environment (Fig. 2). *V. parahaemolyticus* has been observed to be excellent at resisting osmotic stress in lab-based experimental settings and can persist at salinities up to 90‰ (15, 16). However, ecological studies often describe *V. parahaemolyticus* abundance as decreasing at higher levels of salinity (i.e., >10 to 23‰ [11–14]). Experimental settings often grow *V. parahaemolyticus* in medium that is nutrient rich and at temperatures much warmer than what is observed in estuarine waters (~37°C). Current results suggest that when water temperatures are generally cooler (<26°C) or less nutrient rich, *V. parahaemolyticus* may be less resistant to osmotic stress, and subsequently, high salinity can be a limiting factor for initial growth. Alternatively, fewer nutrients may simply not support higher densities. Regardless, the current findings clarify the incongruence of these two research settings. Future experimental studies can confirm these findings by subjecting *V. parahaemolyticus* to lower temperatures, and mesocosm studies could be used to confirm the salinity interaction with turbidity. It is difficult to fully separate out the effect modification of season and water temperature on salinity in the present results. Future environmental studies with greater temporal resolution are therefore also needed to better understand this interaction. Such studies can capture different ranges of water temperature across and between seasons.

The strong negative association observed for *V. parahaemolyticus* and dissolved oxygen runs counter to another study in the Chesapeake Bay that reported a positive association (22). However, the negative association in the present analyses was primarily observed at the lowest levels of DO (3 to 6 mg/liter), while the previous study only recorded a minimum DO of 5.3 mg/liter. Low DO in a water column implies deoxygenation, likely from rapid phytoplankton growth such as algal blooms, and so the observed negative relationship fits within the existing paradigm for *V. parahaemolyticus* ecology. However, this same paradigm runs contrary to the limited associations observed for both chlorophyll and pheophytin. It may be that these pigments have a

lagged association, such that the abundance only increases after an algal bloom has already subsided. Further analyses considering lagged measures of chlorophyll, pheophytin, and dissolved oxygen may further explain the underlying environmental relationships. The Chesapeake Bay is infamous for its widespread and intense algal blooms, especially during warmer months (23). Efforts are currently under way to limit nutrient runoff into the bay (24). Improving water clarity and preventing algal blooms in the bay may therefore also reduce the abundance of *V. parahaemolyticus* in its tidal waters. Further research is needed to investigate this hypothesis.

Many interesting and more complex relationships with tidal water nutrients, specifically, forms of nitrogen and phosphorus, were indicated in this study, several of which are congruous with previous findings (25–27). Most noteworthy is that in the present analysis, many of these potentially colinear measures could be adjusted for simultaneously along with other well-established *V. parahaemolyticus* environmental determinants. While nitrate has been found to have a negative association in at least one other study (26), the inclusion of additional environmental determinants in the present analysis may explain the lack of association in this analysis. The dynamics of ammonium, nitrate, and nitrite concentrations are complex, and so it can be difficult to interpret the multivariate associations of these spatially and temporally indexed variables. Overall, though, positive associations with nutrients indicate that the availability of nitrogen and phosphorus provides a hospitable environment for the bacterium.

The interactions observed for land use and both DO and phosphate indicate that runoff from nearby human development drives the association of *V. parahaemolyticus* with these environmental indicators. Impervious surfaces increase the rate of sediment runoff and can introduce untreated sanitary waste (28). The rapid introduction of inorganic material and microorganisms to the Chesapeake Bay may influence abundance in unique ways. Further research investigating the impact of human development on *V. parahaemolyticus* abundance is recommended.

The variation of *V. parahaemolyticus* that was explained by adding year and season to the model likely indicates that there are additional environmental determinants that were not accounted for in the present analysis. The summer and autumn of 2007 and 2010 were unexpectedly warm relative to long-term temperature averages. This likely contributed to the high abundance of the bacterium observed in these years. Incorporating such climatic anomalies as a covariate could reduce the model's residual temporal variation. The inclusion of phyto- and zooplankton counts, as well as measures of dissolved carbon in water, may further improve the model fit and reduce residual temporal variation. A previous study found that a complex, nonlinear temporal variable essentially explained all variability of the environmental determinants in their model (29). However, this term was likely representative of the natural temporal variation of the variables that became nonsignificant when the temporal regression terms were added. Unless such temporal terms can be shown to have systematic and predictable cycles, they may prove ineffective for the inference of *V. parahaemolyticus* ecology and for forecasting abundance in shellfish harvesting waters. A focus on easily measured environmental determinants may provide more flexibility when developing such models.

The extensive array of water quality measurements and wide observable ranges allowed the present analyses to identify redundancies in associations and to quantify each environmental determinant's relationship to *V. parahaemolyticus* independent of other parameters. The large sample size also provided the statistical power necessary to observe nonlinear relationships of many water quality variables simultaneously, as well as potential interactions across space and time. Other studies have often been unable to evaluate environmental associations with *V. parahaemolyticus* using parametric statistical relationships, due either to a limited number of water quality measurements or to too few samples collected to maintain statistical power.

This *V. parahaemolyticus* study uses interval-censored regression to incorporate all microbial outcome measures within a single model. This model structure will undoubtedly prove useful for modeling *V. parahaemolyticus* and other microbial species when

the detection rates are low but when there is still substantial variation in microbial abundance among the detected samples. This model structure is superior to simple imputation methods, which can substantially bias analytical results (30). The use of linear B-splines in the present analyses also provided flexibility to observe unrestricted nonlinear relationships without sacrificing the clear interpretation of associations that is provided by a generalized linear regression model.

It is important to note that the current microbial analyses were limited to qPCR analysis and were not paired with most probable number (MPN) culturing techniques. It is therefore unknown if the genetic material detected by the analysis was from living *V. parahaemolyticus*. As a result, genomic equivalents for the level of nucleic acid present in the examples were used to determine abundance. The overall similarities between the environmental determinant associations in the present analysis and those from studies that included a culture method suggest that water samples in this study likely included living *V. parahaemolyticus* bacteria.

Although *tdh* and *trh* genetic markers, which have been described as indicating virulence for *V. parahaemolyticus*, were measured for this study, neither was detected. However, this finding is similar to previous work that also identified low detection rates of these markers in Chesapeake Bay water samples (22). Previous studies have shown that the environmental relationships with *V. parahaemolyticus* strains containing these markers may be distinct from those that have only the *tlh* marker (13). An improved understanding of *V. parahaemolyticus* genetic markers is an area of active investigation, as many strains with *tdh-trh* have not been found to be pathogenic, while still other strains lacking these markers can be infectious (31). Regardless, the environmental associations described in this analysis should not necessarily be interpreted as representative of an abundance of pathogenic *V. parahaemolyticus* in the water column. The small detectable sample in the current study may also be contributing to a lack of detection of *tdh-trh*. Shellfish are known to contain relatively high levels of *V. parahaemolyticus*, and previous studies have detected *tdh-trh* markers in shellfish in the Chesapeake Bay (22). Therefore, future shellfish sampling in the Chesapeake Bay could yield higher detection rates of these pathogenic markers.

Bacterium abundance in oysters may be considered a more relevant measure for the public health concerns surrounding *V. parahaemolyticus* in the Chesapeake Bay given that shellfish consumption is the primary exposure route for infection. Unfortunately, a standard conversion between *V. parahaemolyticus* in the water column and in shellfish tissue has not been identified. However, *V. parahaemolyticus* variability across oysters can be high given the idiosyncrasies of shellfish physiology (e.g., age, sex, immune status, filter rate, etc.). In contrast, the variability across water samples is likely to be relatively more stable, potentially making the present analyses the preferred approach for understanding the relative differences in bacterium abundance across space and time and as a function of environmental measures in the bay. Therefore, an improved ecological understanding of *V. parahaemolyticus* in the water column, and future prediction models for shellfish harvesting waters, may provide a more suitable route for mitigating the public health burden of the bacterium compared to analogous efforts in shellfish. Predictions in the water column could also be directly beneficial in reducing direct exposure to *V. parahaemolyticus* in recreational waters. Regardless, future work in the Chesapeake Bay should additionally sample for *V. parahaemolyticus* abundance in shellfish to compare the prediction performance as well as to investigate abundance conversion models between sample types.

The inclusion of lagged variables in the current analysis is supported by significant associations of rainfall at 1- and 7-day lags and the nonsignificant association for same-day rainfall in the multivariate model. A full time-series analysis of precipitation could further improve the understanding of the impacts of rainfall on *V. parahaemolyticus*. Such work should also consider watershed river discharge following an array of weather events, including thunderstorms and droughts. Additional investigations could further address nuances of land use and the spatial-temporal impact of runoff by coupling unique weather events with a digital elevation model.

Lagged associations may also exist for other environmental determinants. The Chesapeake Bay is one of the most researched bodies of water in the world, with many of the water quality measurements used in this analysis being consistently sampled over space and time. The availability of this larger sampling scheme is well suited for more complex spatial-temporal analyses that incorporate lags of the environmental measures used in this study. A better model fit through inclusion of environmental lags would be of great interest for predicting or forecasting *V. parahaemolyticus* abundance in water and shellfish.

MATERIALS AND METHODS

Study area. The Chesapeake Bay is the largest estuary in the United States, with a surface area of 11,601 km² and a volume of over 68 trillion liters (32). Seawater from the Atlantic Ocean enters through the mouth of the bay in the south, while freshwater enters from six major rivers and over 100 other smaller rivers and streams. The temperatures of the Chesapeake Bay range widely over space and time, dipping below 4°C in the winter and reaching higher than 28°C in the summer. Salinity in the bay also varies substantially; fresh tidal waters can be found at the heads of rivers, while salinity concentrations of up to 30‰ are common near the mouth of the bay (33). The Chesapeake Watershed extends into six states and has an area of approximately 166,000 square kilometers. This watershed is home to approximately 18 million people, contains multiple metropolitan areas, and is known for its agricultural activities. Approximately 500 million pounds of seafood are harvested from the bay each year (32).

Data collection. Water sampling methods were described previously in reference 17. Briefly, surface water samples (0.5-m depth) were collected at 148 sampling stations across the Chesapeake Bay by the Maryland Department of Natural Resources and the Virginia Department of Environmental Quality's respective water quality monitoring programs according to standard Chesapeake Bay Program protocols (34) (Fig. 1). Samples were taken during the months of April (spring), July (summer), and October (autumn) from 2007 to 2010, as well as in January (winter) of 2007 ($n = 1,592$). Winter data were removed from the present analysis due to the lack of comparisons across years. Therefore, only 1,523 surface samples were considered for this analysis.

Water quality was measured *in situ* with a YSI datasonde (YSI Incorporated, Yellow Springs, OH) and with a Secchi disk at the same time and location at which water samples were collected. Measurements were analyzed according to the Chesapeake Bay Program's guidelines (35). Measurements included water temperature, salinity, forms of nitrogen and phosphorus, turbidity, and dissolved and suspended solids, as well as pigments of phytoplankton (i.e., chlorophyll *a* and pheophytin). The ratios of nitrogen and phosphorus were also calculated.

A bathymetric digital elevation model, created by the National Oceanic and Atmospheric Association (NOAA) Chesapeake Bay Office using hydrographic survey soundings (36), was accessed to determine the total depth of each monitoring station. Each station was categorized into seven bathymetry bins, and the midpoints of each bin were used as pseudocontinuous variables (37).

Daily weather data for 2007 to 2010 from 23 surrounding monitors were retrieved from NOAA's National Centers for Environmental Information Global Historical Climatology Network database (Fig. 1) (38). Water samples were assigned to the closest weather monitor and temporally matched with each water sampling date. Precipitation and air temperature variables were lagged up to 7 days before each sampling date. Precipitation variables were converted into binary variables based on whether rainfall was moderate to heavy (>1 mm/h). Daily summaries of wind speed were also retrieved from the weather monitor at the U.S. Naval Academy in Annapolis, Maryland, and were temporally matched with each water sample.

Land use classification for the Chesapeake Bay Watershed was extracted from the Multi-Resolution Land Characteristics Consortium's 2006 national land cover database, which is based on Landsat satellite data (39). To simplify the analysis, only five classifications (water, developed lands, plant and animal agriculture, wetlands, and forest) from this database were considered (Fig. 1). Five-, ten-, and fifteen-mile circular buffers were drawn around each water sample monitoring station (40). For each buffer, the percentage of pixels classified as water was calculated (%Water), and the most prominent land use was identified. Only the 5-mile buffer was considered for the %Water variable, given that larger buffers extended significantly outside the watershed. A sensitivity analysis revealed minimal variation in the classification of surrounding land use by buffer size, and so the 10-mile buffer was chosen for all subsequent analyses.

qPCR for total *V. parahaemolyticus*. The purification methods were in accordance with a modified MoBio power soil protocol as described in reference 41. A species-specific primer/probe combination was employed for the detection of total *V. parahaemolyticus* (42). The assay incorporated a unique internal control for the detection of any inhibitors within each sample (42) (Table 4). Primers and probes for this assay were obtained from Integrated DNA Technologies (IDT, Coralville, IA). The qPCR for *V. parahaemolyticus* was performed by adding 0.50 μl of deoxynucleoside triphosphate (dNTP) solution, 0.500 μl of each primer, 0.45 μl of 5-U/μl Platinum hot start *Taq* (Invitrogen), and a quantity of PCR-grade water sufficient for 25-μl reaction mixtures. Two-stage qPCR cycling parameters were optimized to an initial denaturation of template at 95°C for 60 s, followed by 45 cycles of denaturation at 95°C for 5 s and combined annealing and extension at 66°C for 45 s. On occasion, amplification products were run on a 1.5% agarose gel at 84 V for 1 h 45 min as a quality control measure to ensure proper products were being amplified by comparison to a known molecular weight marker.

TABLE 4 Primer/probe sets used for the detection of total *V. parahaemolyticus* (*tlh*) and clinical strains (*trh/tdh*)

Name	Sequence and probe(s)
*tlh*_F	ACTCAACACAAGAAGAGATCGACAA
*tlh*_R	GATGAGCGGTTGATGTCCAA
*tlh*_TXRD	TxRED-CGCTCGCGTTCACGAAACCGT-3BHQ_2
*trh*_F	TTGCTTTCAGTTTGCTATTGGCT
*trh*_R	TGTTTACCGTCATATAGGCGCTT
*trh*133-23	TET-AGAAATACAACAATCAAAACTGA-MGBNFQ
*tdh*_F	TCCCTTTTCCTGCCCCC
*tdh*_R	CGCTGCCATTGTATAGTCTTTATC
*tdh*269-20	FAM-TGACATCCTACATGACTGTG-MGBNFQ

qPCR for *trh-tdh* from *V. parahaemolyticus*. A combination of two sets of primers and probes was used for the detection of the genes *tdh* and *trh* in *V. parahaemolyticus* (42) (Table 4). All primers and probes were obtained from IDT. As with the total *V. parahaemolyticus* assay, a unique internal control was incorporated simultaneously to test for the presence and influence of inhibition. The qPCRs was performed by using 1.00 μl of dNTP, 0.50 μl of *tdh* and *trh* primers, 0.19 μl of 10.00 μM *tdh*_269-20 and *trh*_133-23 probes, and 0.45 μl of 5-U/μl Platinum hot start *Taq* (Invitrogen) per reaction. Two-stage qPCR cycling parameters were an initial denaturation of template at 95°C for 60 s, followed by 50 cycles of denaturation at 95°C for 5 s and combined annealing and extension at 59°C for 45 s. Quality control procedures were performed as described above.

Assay performance/standard curve. Assay performance testing was carried out as described previously (41). Aspects of the assay that were evaluated included extraction recovery estimates, bottle-to-bottle replication, within-sample repeatability, and assay efficiency calculated from multiple standard curves using the formula $E = -1 + 10^{(-1/slope)}$ (43). In addition, the effects of freezing water samples for transportation were evaluated. Seawater was collected and screened prior to use to ensure no background contamination. In replicate 500-ml sterile Nalgene bottles, *V. parahaemolyticus* cells were added over a 6-log scale. One bottle of each replicate log was immediately extracted as described above, while the second replicate was frozen at −20°C for 1 week. Bottles were subsequently pulled from the freezer, thawed, and extracted in the same manner as for all other samples. The recovery of DNA was estimated by first running a boiled cell suspension through the PCR assay and gel purification. The DNA content was measured using a NanoDrop 1000 instrument (Thermo Scientific, Waltham, MA).

Standard curves were generated for total *V. parahaemolyticus*, whereas virulence was assessed as present or absent. To generate standard curves, cell suspensions were made from pure cultures taken in the exponential growth phase in alkaline phosphate water, and 200 μl from each suspension was plated on tryptone salt agar plates with 3% NaCl to determine the cell count. Starting with 4.8×10^7 cells/ml, 1:10 dilutions were made down to 4.8 cells/ml. Two filters were processed for each dilution by spiking 200 ml of water with 1 ml of the *Vibrio* dilution, and all filters were processed as described above. The water used for the standard curve was tested for the presence of *V. parahaemolyticus* prior to being used to ensure no target bacteria were present. The extracted DNA from all dilutions was then run according to the above qPCR parameters, and the C_T values were plotted against the numbers of total cells in the extraction to determine the standard curve. The limit of detection based on the standard curves was 0.14 CFU/ml, and the limit of quantitation was 1.00 CFU/ml. These standard curves were then applied to the water samples to transform C_T values into CFU/ml. Once complete, the units of *V. parahaemolyticus* abundance were transformed into genomic equivalents of CFU per milliliter (GE/ml) and were used as the primary data outcome for statistical analysis.

Statistical analysis. *V. parahaemolyticus* outcome data, along with environmental determinants, were summarized using quartiles and proportions and were tabulated by sampling season. The proportion of missing data for each determinant was calculated. The distribution of outcome data was also evaluated for normality.

Given the large number of water samples analyzed below the limit of detection (sparse data) and for quantitation, along with samples reaching values of 143.6 GE/ml, an interval-censored regression was used to model *V. parahaemolyticus* abundance. The statistical methods developed for survival analysis (44) are easily transferable to this environmental setting by replacing time-to-event censoring with limits of detection and quantitation from the qPCR analysis. Abundance was therefore reclassified into intervals: 0.00 to 0.14 for measures below the limit of detection and 0.14 to 1.00 for measures between the limit of detection and the limit of quantitation. All other *V. parahaemolyticus* measurements were given an interval whose range was the value of the observed concentration (e.g., 34.50 to 34.50). Analyses were conducted assuming a lognormal distribution for quantified measures.

Univariate regression models were created for the abundance of *V. parahaemolyticus*. Nonparametric local regressions (LOESSs) were plotted to determine if the assumption of linearity for the included effects was appropriate. If trends appeared to be clearly nonlinear, linear B-splines were calculated with knots that were identified by visual inspection of the trends. The variance inflation factor was used to identify categories of colinear variables based on a cutoff value of 10. The multivariate model was determined by removing colinear variables as well as variables found to be redundant during *a priori* exploratory analyses.

The results of the models are reported in their exponentiated form, so as to be interpreted as relative change in the geometric mean (RCGM). Generalized linear hypothesis testing for linear combinations of the B-spline estimates was also performed.

Spatial and temporal analyses. To infer large-scale spatial trends, the Chesapeake Bay tidal waters were split into 10 separate aggregations of the Chesapeake Bay Program's analytical segmentation scheme (45) (Fig. 1). Geographic variation of residuals was mapped to assess the lack of fit by monitoring station. The residuals were also evaluated by season and year. Models were compared using the Akaike information criterion (AIC) to assess whether the addition of space/time variables (main effect and interaction) significantly improved the model fit and to evaluate the change in regression estimates as an indication of unmeasured confounders. Models were stratified by year, season, or region to determine if there was substantial effect modification of environmental associations. When such changes were observed, formal statistical interaction terms were added to the model.

Residual spatial variation (i.e., spatially dependent regression residuals) can be a concern for regression modeling, as such dependence would violate one of the underlying assumptions of residual independence. Semivariograms, a tool from the field of geostatistics (46), were estimated for model residuals to diagnose residual spatial variation. Chesapeake Bay water distances were used in semivariogram analyses to better adhere to the complex geometry of the Bay, the details of which have been reported elsewhere (47).

Mapping and spatial data integration were performed in ArcGIS version 10.3 (48). All statistical analyses were performed in R statistical software (49) using the survival package for interval-censored regression modeling (50, 51), ggplot2 for data visualization (52), and multcomp for B-spline analysis (53), as well as a number of additional packages for analysis support (54–66).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/AEM.01147-17.

SUPPLEMENTAL FILE 1, PDF file, 1.5 MB.

REFERENCES

1. Scallan E, Hoekstra RM, Angulo FJ, Tauxe RV, Widdowson MA, Roy SL, Jones JL, Griffin PM. 2011. Foodborne illness acquired in the United States–major pathogens. Emerg Infect Dis 17:7–15. https://doi.org/10.3201/eid1701.P11101.
2. Newton A, Kendall M, Vugia DJ, Henao OL, Mahon BE. 2012. Increasing rates of vibriosis in the United States, 1996–2010: review of surveillance data from 2 systems. Clin Infect Dis 54(Suppl 5):S391–S395. https://doi.org/10.1093/cid/cis243.
3. Baker-Austin C, Trinanes J, Gonzalez-Escalona N, Martinez-Urtaza J. 2017. Non-cholera vibrios: the microbial barometer of climate change. Trends Microbiol 25:76–84. https://doi.org/10.1016/j.tim.2016.09.008.
4. Vezzulli L, Colwell RR, Pruzzo C. 2013. Ocean warming and spread of pathogenic vibrios in the aquatic environment. Microb Ecol 65:817–825. https://doi.org/10.1007/s00248-012-0163-2.
5. Lima FP, Wethey DS. 2012. Three decades of high-resolution coastal sea surface temperatures reveal more than warming. Nat Commun 3:704. https://doi.org/10.1038/ncomms1713.
6. Vezzulli L, Grande C, Reid PC, Helaouet P, Edwards M, Hofle MG, Brettar I, Colwell RR, Pruzzo C. 2016. Climate influence on *Vibrio* and associated human diseases during the past half-century in the coastal North Atlantic. Proc Natl Acad Sci U S A 113:E5062–E5071. https://doi.org/10.1073/pnas.1609157113.
7. Takemura AF, Chien DM, Polz MF. 2014. Associations and dynamics of *Vibrionaceae* in the environment, from the genus to the population level. Front Microbiol 5:38. https://doi.org/10.3389/fmicb.2014.00038.
8. Johnson CN. 2015. Influence of environmental factors on *Vibrio* spp. in coastal ecosystems. Microbiol Spectr 3:VE-0008-2014. https://doi.org/10.1128/microbiolspec.VE-0008-2014.
9. Kaneko T, Colwell RR. 1975. Adsorption of *Vibrio parahaemolyticus* onto chitin and copepods. Appl Microbiol 29:269–274.

10. Turner JW, Malayil L, Guadagnoli D, Cole D, Lipp EK. 2014. Detection of *Vibrio parahaemolyticus*, *Vibrio vulnificus* and *Vibrio cholerae* with respect to seasonal fluctuations in temperature and plankton abundance. Environ Microbiol 16:1019–1028. https://doi.org/10.1111/1462-2920.12246.

11. DePaola A, Nordstrom JL, Bowers JC, Wells JG, Cook DW. 2003. Seasonal abundance of total and pathogenic *Vibrio parahaemolyticus* in Alabama oysters. Appl Environ Microbiol 69:1521–1526. https://doi.org/10.1128/AEM.69.3.1521-1526.2003.

12. U.S. Food and Drug Administration. 2005. Quantitative risk assessment on the public health impact of pathogenic *Vibrio parahaemolyticus* in raw oysters. U.S. Food and Drug Administration, Washington, DC.

13. Johnson CN, Flowers AR, Noriea NF, III, Zimmerman AM, Bowers JC, DePaola A, Grimes DJ. 2010. Relationships between environmental factors and pathogenic *Vibrios* in the northern Gulf of Mexico. Appl Environ Microbiol 76:7076–7084. https://doi.org/10.1128/AEM.00697-10.

14. Böer SI, Heinemeyer E-A, Luden K, Erler R, Gerdts G, Janssen F, Brennholt N. 2013. Temporal and spatial distribution patterns of potentially pathogenic *Vibrio* spp. at recreational beaches of the German North Sea. Microb Ecol 65:1052–1067. https://doi.org/10.1007/s00248-013-0221-4.

15. Naughton LM, Blumerman SL, Carlberg M, Boyd EF. 2009. Osmoadaptation among *Vibrio* species and unique genomic features and physiological responses of *Vibrio parahaemolyticus*. Appl Environ Microbiol 75:2802–2810. https://doi.org/10.1128/AEM.01698-08.

16. Whitaker WB, Parent MA, Naughton LM, Richards GP, Blumerman SL, Boyd EF. 2010. Modulation of responses of *Vibrio parahaemolyticus* O3:K6 to pH and temperature stresses by growth at different salt concentrations. Appl Environ Microbiol 76:4720–4729. https://doi.org/10.1128/AEM.00474-10.

17. Jacobs JM, Rhodes M, Brown CW, Hood RR, Leight A, Long W, Wood R. 2014. Modeling and forecasting the distribution of *Vibrio vulnificus* in Chesapeake Bay. J Appl Microbiol 117:1312–1327. https://doi.org/10.1111/jam.12624.

18. Blackwell KD, Oliver JD. 2008. The ecology of *Vibrio vulnificus*, *Vibrio cholerae*, and *Vibrio parahaemolyticus* in North Carolina estuaries. J Microbiol 46:146–153. https://doi.org/10.1007/s12275-007-0216-2.

19. Phillips AM, DePaola A, Bowers J, Ladner S, Grimes DJ. 2007. An evaluation of the use of remotely sensed parameters for prediction of incidence and risk associated with *Vibrio parahaemolyticus* in Gulf Coast oysters (*Crassostrea virginica*). J Food Prot 70:879–884. https://doi.org/10.4315/0362-028X-70.4.879.

20. Johnson CN, Bowers JC, Griffitt KJ, Molina V, Clostio RW, Pei S, Laws E, Paranjpye RN, Strom MS, Chen A, Hasan NA, Huq A, Noriea NF, III, Grimes DJ, Colwell RR. 2012. Ecology of *Vibrio parahaemolyticus* and *Vibrio vulnificus* in the coastal and estuarine waters of Louisiana, Maryland, Mississippi, and Washington (United States). Appl Environ Microbiol 78:7249–7257. https://doi.org/10.1128/AEM.01296-12.

21. Håkanson L. 2006. The relationship between salinity, suspended particulate matter and water clarity in aquatic systems. Ecol Res 21:75–90. https://doi.org/10.1007/s11284-005-0098-x.

22. Parveen S, Hettiarachchi KA, Bowers JC, Jones JL, Tamplin ML, McKay R, Beatty W, Brohawn K, DaSilva LV, DePaola A. 2008. Seasonal distribution of total and pathogenic *Vibrio parahaemolyticus* in Chesapeake Bay oysters and waters. Int J Food Microbiol 128:354–361. https://doi.org/10.1016/j.ijfoodmicro.2008.09.019.

23. Kemp WM, Boynton WR, Adolf JE, Boesch DF, Boicourt WC, Brush G, Cornwell JC, Fisher TR, Glibert PM, Hagy JD. 2005. Eutrophication of Chesapeake Bay: historical trends and ecological interactions. Mar Ecol Prog Ser 303:1–29. https://doi.org/10.3354/meps303001.

24. EPA. 2011. Clean Water Act section 303(d): notice for the establishment of the total maximum daily load (TMDL) for the Chesapeake Bay (EPA). EPA, Washington, DC. https://www.gpo.gov/fdsys/pkg/FR-2011-01-05/pdf/2010-33280.pdf.

25. Rehnstam-Holm AS, Godhe A, Härnström K, Raghunath P, Saravanan V, Collin B, Karunasagar I, Karunasagar I. 2010. Association between phytoplankton and *Vibrio* spp. along the southwest coast of India: a mesocosm experiment. Aquat Microb Ecol 58:127–139. https://doi.org/10.3354/ame01360.

26. Oberbeckmann S, Wichels A, Wiltshire KH, Gerdts G. 2011. Occurrence of *Vibrio parahaemolyticus* and *Vibrio alginolyticus* in the German Bight over a seasonal cycle. Antonie Van Leeuwenhoek 100:291–307. https://doi.org/10.1007/s10482-011-9586-x.

27. Rehnstam-Holm AS, Atnur V, Godhe A. 2014. Defining the niche of *Vibrio parahaemolyticus* during pre- and post-monsoon seasons in the coastal Arabian Sea. Microb Ecol 67:57–65. https://doi.org/10.1007/s00248-013-0311-3.

28. Rasmussen TJ, Schmidt HC. 2009. Stormwater runoff: what it is and why it is important in Johnson County, Kansas. U.S. Geological Survey, Reston, VA.

29. Paranjpye RN, Nilsson WB, Liermann M, Hilborn ED, George BJ, Li Q, Bill BD, Trainer VL, Strom MS, Sandifer PA. 2015. Environmental influences on the seasonal distribution of *Vibrio parahaemolyticus* in the Pacific Northwest of the USA. FEMS Microbiol Ecol 91:fiv121. https://doi.org/10.1093/femsec/fiv121.

30. Helsel DR. 2005. Nondetects and data analysis. John Wiley and Sons, New York, NY.

31. Paranjpye R, Hamel OS, Stojanovski A, Liermann M. 2012. Genetic diversity of clinical and environmental *Vibrio parahaemolyticus* strains from the Pacific Northwest. Appl Environ Microbiol 78:8631–8638. https://doi.org/10.1128/AEM.01531-12.

32. Chesapeake Bay Program. 2016. Discover the Chesapeake: Bay 101 facts & figures. http://www.chesapeakebay.net/discover/bay101/facts. Accessed 3 December 2016.

33. Chesapeake Bay Program. 2016. Discover the Chesapeake: the Bay ecosystem, physical characteristics. http://www.chesapeakebay.net/discover/bayecosystem/physical. Accessed 3 December 2016.

34. EPA. 1996. Recommended guidelines for sampling and analyses in the Chesapeake Bay monitoring program. EPA, Washington, DC.

35. Olson M, CBPS. 2012. Guide to using Chesapeake Bay Program water quality monitoring data. Chesapeake Bay Program, Annapolis, MD.

36. DOC, NOAA, NOS, SP. 1998. Chesapeake Bay, VA/MD (M130) bathymetric digital elevation model (30 meter resolution) derived from source hydrographic survey soundings collected by NOAA, Silver Spring, MD.

37. Nowlan P. 2012. Chesapeake Bay bathymetric polygons. NOAA Chesapeake Bay Office, Annapolis, MD.

38. NCEI NOAA. 2016. Global Historical Climatology Network (GHCN). https://www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets/global-historical-climatology-network-ghcn.

39. Fry JA, Xian G, Jin S, Dewitz JA, Homer CG, Limin Y, Barnes CA, Herold ND, Wickham JD. 2011. Completion of the 2006 national land cover database for the conterminous United States. Photogramm Eng Remote Sensing 77:858–864.

40. Price CV. 2016. National Water-Quality Assessment (NAWQA) area-characterization toolbox, v2.1. ArcGIS. https://www.arcgis.com/home/item.html?id=29707fb7f1664538871c65d7e1d9612e%20.

41. Jacobs J, Rhodes M, Sturgis B, Wood B. 2009. Influence of environmental gradients on the abundance and distribution of *Mycobacterium* spp. in a coastal lagoon estuary. Appl Environ Microbiol 75:7378–7384. https://doi.org/10.1128/AEM.01900-09.

42. Nordstrom JL, Vickery MC, Blackstone GM, Murray SL, DePaola A. 2007. Development of a multiplex real-time PCR assay with an internal amplification control for the detection of total and pathogenic *Vibrio parahaemolyticus* bacteria in oysters. Appl Environ Microbiol 73:5840–5847. https://doi.org/10.1128/AEM.00460-07.

43. Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29:e45. https://doi.org/10.1093/nar/29.9.e45.

44. Long JS. 1997. Regression models for categorical and limited dependent variables, 1st ed. SAGE Publications, Thousand Oaks, CA.

45. Chesapeake Bay Program. 2005. Chesapeake Bay Program analytical segmentation scheme: revisions, decisions and rationales 1983–2003, 2005 Addendum. Chesapeake Bay Program, Annapolis, MD. http://www.chesapeakebay.net/documents/3676/chesapeake_bay_program_analytical_segmentation_scheme.pdf.

46. Cressie NC. 1993. Statistics for spatial data, revised ed. John Wiley and Sons, New York, NY.

47. Murphy RR, Perlman E, Ball WP, Curriero FC. 2015. Water-distance-based kriging in Chesapeake Bay. J Hydrol Eng 20:0501403. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001135.

48. Environmental Systems Research Institute. 2011. ArcGIS desktop: release 10.3. Environmental Systems Research Institute, Redlands, CA.

49. Core Team R. 2016. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.r-project.org/.

50. Therneau TM, Grambsch PM. 2000. Modeling survival data: extending the Cox model. Springer, New York, NY.

51. Therneau T. 2015. A package for survival analysis in S, version 2.38. http://cran.r-project.org/package=survival.

52. Wickham H. 2009. ggplot2: elegant graphics for data analysis. Springer-Verlag, New York, NY.

53. Hothorn T, Bretz F, Westfall P. 2008. Simultaneous inference in general parametric models. Biom J 50:346–363. https://doi.org/10.1002/bimj .200810425.

54. Pebesma EJ, Bivand RS. 2005. Classes and methods for spatial data: the sp package. https://cran.r-project.org/web/packages/sp/vignettes/intro _sp.pdf.

55. Fox J, Weisberg HS. 2011. An R companion to applied regression, 2nd ed. Sage, Thousand Oaks, CA.

56. Wickham H. 2011. The split-apply-combine strategy for data analysis. J Stat Softw 40:1–29. https://doi.org/10.18637/jss.v040.i01.

57. Bivand RS, Pebesma E, Gomez-Rubio V. 2013. Applied spatial data analysis with R, 2nd ed. Springer, New York, NY.

58. Dowle M, Srinivasan A, Short T, Lianoglou S, Saporta R, Antonya E. 2015. data.table: extension of data.frame, vR package version 1.9.6. https:// cran.r-project.org/package=data.table.

59. Etten JV. 2015. gdistance: distances and routes on geographical grids, vR package version 1.1-9. https://cran.r-project.org/package=gdistance.

60. Bivand R, Keitt T, Rowlingson B. 2016. rgdal: bindings for the geospatial data abstraction library, vR package version 1.1-10. https://cran.r-project .org/package=rgdal.

61. Harrell FE, Dupont C. 2016. Hmisc: Harrell miscellaneous, vR package version 3.17-4. https://cran.r-project.org/package=Hmisc.

62. Hijmans RJ. 2016. raster: geographic data analysis and modeling, vR package version 2.5-8. https://cran.r-project.org/package=raster.

63. Core Team R. 2016. foreign: read data stored by Minitab, S, SAS, SPSS, Stata, Systat, Weka, dBase, vR package version 0.8-67. R Foundation for Statistical Computing, Vienna, Austria. https://cran.r-project.org/package =foreign.

64. Ribeiro PJ, Diggle PJ. 2016. geoR: analysis of geostatistical data, vR package version 1.7-5.2. https://cran.r-project.org/package=geoR.

65. Templ M, Alfons A, Kowarik A, Prantner B. 2016. VIM: visualization and imputation of missing values, vR package version 4.5.0. https://cran.r -project.org/package=VIM.

66. Wright K. 2016. corrgram: plot a correlogram, vR package version 1.9. https://cran.r-project.org/package=corrgram.