APPLIED AND ENVIRONMENTAL MICROBIOLOGY, May 2008, p. 2915–2920 Vol. 74, No. 9
0099-2240/08/$08.00+0 doi:10.1128/AEM.02139-07

Covariability of *Vibrio cholerae* Microdiversity and Environmental Parameters▽†

Young-Gun Zo,[1,2] Nipa Chokesajjawatee,[1,3] Eiji Arakawa,[4] Haruo Watanabe,[4] Anwar Huq,[1,2] and Rita R. Colwell[1,2]*

Center of Marine Biotechnology, University of Maryland Biotechnology Institute, 701 E. Pratt St., Baltimore, Maryland 21202[1]; Center of Bioinformatics and Computational Biology, University of Maryland Institute of Advanced Computer Studies, University of Maryland College Park, College Park, Maryland 20742[2]; National Center for Genetic Engineering and Biotechnology, 113 Phahonyothin Rd., Klong 1, Klong Luang, Pathumthani 12120, Thailand[3]; and Department of Bacteriology, National Institute of Infectious Diseases, Toyama 1-23-1, Shinjuku ku, Tokyo 162-8640, Japan[4]

Received 18 September 2007/Accepted 22 February 2008

Fine-scale diversity of natural bacterial assemblages has been attributed to neutral radiation because correspondence between bacterial phylogenetic signals in the natural environment and environmental parameters had not been detected. Evidence that such correspondence occurs is provided for *Vibrio cholerae*, establishing a critical role for environmental parameters in bacterial diversity.

A notable discovery arising from recent investigations of the microbial world is the immense phylogenetic diversity of natural bacterial populations that determines the fine architecture of "microdiversity clusters" (1). Bacterial populations in coastal waters form clusters of diverse clonal entities that radiate from a limited number of major phylogenetic branches (17). This finding prompts questions of how such a population structure can arise. Previous studies have suggested that the mechanism of the microdiversity structure is an accumulation of neutral genetic variation, after periodic selective sweeps acting against the nonfittest clonal entities (1, 2, 17), mainly because no discernible correspondence could be observed between DNA sequence polymorphism of bacterial housekeeping genes and environmental variation (17). Considering a neutral diversity mechanism alone, however, seemingly contradicts other facts of bacterial microevolution that have been gleaned from studies undertaken in different fields of microbiology. Simulated microevolution experiments demonstrated that a defined population of a given bacterial strain can develop into two or more variants occupying different ecological niches (6). Convincing evidence for niche-specific selection in the determination of a global community structure of *Prochlorococcus*, an autotrophic bacterium, has been provided by Follows et al. (3). Therefore, the question of whether natural selection within diverse ecological niches leads to the formation of seemingly neutral microdiversity remains an intriguing question with respect to the population structure of heterotrophic bacteria in nature. In the present study, extracellular (or ectocellular), phenotypic or genotypic traits of *Vibrio cholerae*, an indigenous inhabitant of estuaries (10), were analyzed with respect to spatial and temporal correspondence of variability in bacterial population diversity to variation in selected environmental parameters.

A set of 221 *V. cholerae* strains examined in the present study were isolated from the upper Chesapeake Bay from 1998 to 2000 and collected in a census-style survey carried out in our laboratory (10). In brief, *V. cholerae* was isolated at five shore sampling stations (sites from North to South: F, Susquehanna river flat; B, Baltimore Inner Harbor; K, Kent Island; S, estuarine shore on the Rhode River subestuary; and H, Horn Point) biweekly during the summer and monthly in other seasons of the year. Using membrane filters and nylon nets, suspended particulates in the surface water (0- to 1-m depth) were collected and split into three size fractions: >0.2 μm, 20 to 64 μm, and >64 μm. Each of the particulate samples was used as an inoculum in enrichment cultures from which *V. cholerae* colonies were isolated.

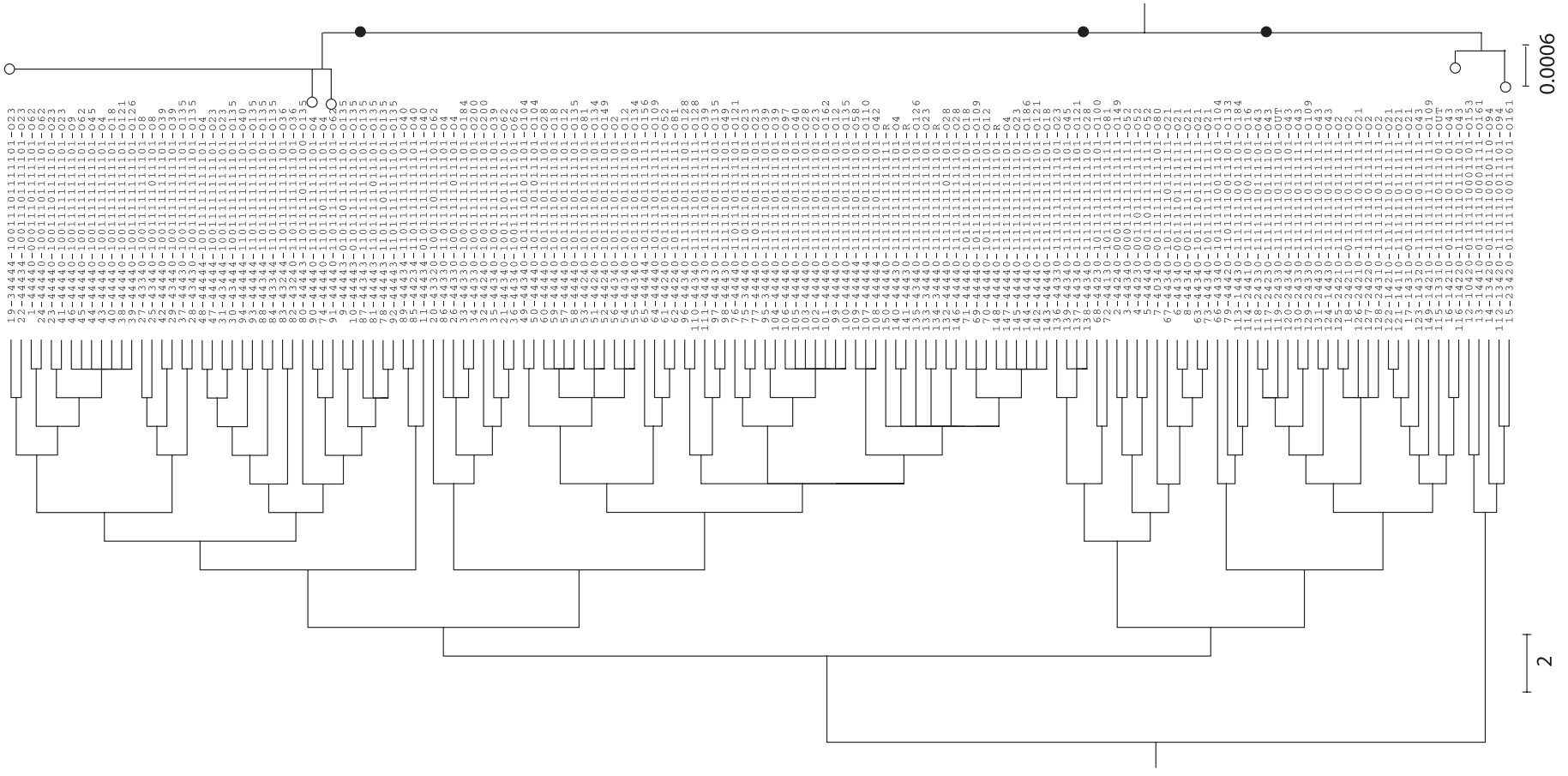
The Sakazaki O serotype (13) of each isolate was determined by using a pool of 210 standard polyclonal antibodies, together with other characteristics listed in Fig. 1. Isolates were also tested for presence or absence of toxin genes and outer membrane protein genes by using a dot blot hybridization method (see Table S1 in the supplemental material for further information). Alleles of genes in each of the isolates were determined by assigning isolates to an allele class according to hybridization signal intensity. The classes were determined by coding the signal intensity semiquantitatively on a scale of 0 to 4, in steps as follows: 0 to 25%, 25 to 50%, 50 to 75%, and 75 to 100% signal intensity compared to the positive control. The results of phenotypic and genotypic tests were coded as binary or five-state characters, e.g., absence (e.g., Δ*hlyA*) and four allele classes (e.g., $hlyA^{+}$, $hlyA^{2+}$, $hlyA^{3+}$, or $hlyA^{4+}$). Isolates sharing an identical set of characters were concluded to belong to the same operational taxonomic units (OTUs) and were treated as identical clones.

The cholera toxin-related genes (*ctxA*, *tcpA*, and *zot*) were absent in all environmental isolates of *V. cholerae*. Phenotypic traits invariable among isolates included the following: growth in nutrient broth containing 0, 1, 3, and 8% NaCl; acid pro-

* Corresponding author. Mailing address: Center for Bioinformatics and Computational Biology, University of Maryland Institute of Advanced Computer Studies, University of Maryland College Park, College Park, MD 20742. Phone: (301) 405-9550. Fax: (301) 314-6654. E-mail: rcolwell@umiacs.umd.edu.

† Supplemental material for this article may be found at http://aem.asm.org/.

▽ Published ahead of print on 29 February 2008.



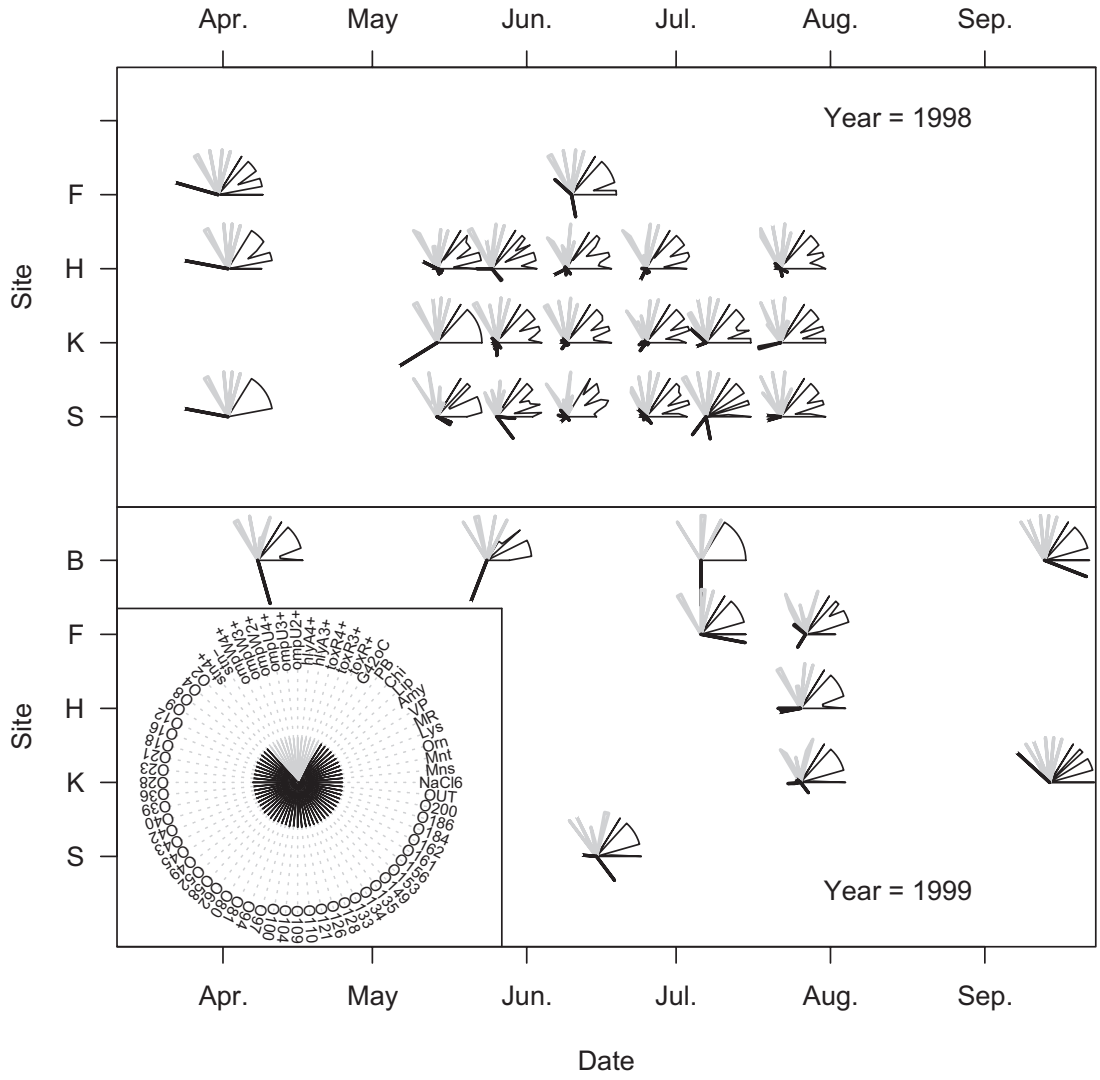


FIG. 2. Prevalence of *V. cholerae* characteristics, shown as star charts. Inset (spines radiating from a common origin): the solid line spine represents scale bars for 100% prevalence of a characteristic. The angle of a star chart apex represents the characteristic indicated by the accompanying label; labels give abbreviations for phenotypic and genotypic characteristics used in Fig. 1 or O-serogroup names. For the star charts, the origin indicates the site and time of sampling; the apex indicates the prevalence of a characteristic, ranging from zero to 100%; the thin, solid lines show the lines connecting apexes (prevalence values) for selected phenotypic characteristics (NaCl6, Mns, Mnt, Orn, Lys, MR, VP, Amy, Lp, Ch, PB, and G42oC [see the legend to Fig. 1 for definitions], counterclockwise starting from the apex with the zero tangent); the thin, gray lines show the lines connecting apexes for genotypic characteristics (*toxR*$^{+}$, *toxR*$^{3+}$, *toxR*$^{4+}$, *hlyA*$^{3+}$, *hlyA*$^{4+}$, *ompU*$^{2+}$, *ompU*$^{3+}$, *ompU*$^{4+}$, *ompW*$^{2+}$, *ompW*$^{3+}$, *ompW*$^{4+}$, Δ*stn*, and *stn*$^{4+}$ in counterclockwise order); the thick, solid lines show the lines connecting apexes for O serogroups (O2-O200 and "OUT", i.e., untypeable O serogroups). On the horizontal axis, tick marks indicate the 15th day of the given month.

duction from sucrose and arabinose; the production of enzymes such as arginine dihydrolase, oxidase, and gelatinase; the hydrolysis of esculin; and sensitivity to 10 and 150 mg of vibriostatic agent O/129. The power of the 18 variable phenotypic, genotypic, and serological traits to differentiate *V. cholerae* strains was estimated by clonal redundancy, calculated as the ratio of the total number of strains to the number of OTUs. From the 221 strains, 150 OTUs could be defined. Therefore,

FIG. 1. Hierarchal clusters of OTUs based on complete linkage analysis of characteristics (left) and maximum-likelihood tree of 16S rRNA sequences (right) of randomly selected strains. The OTU label is presented as follows: arbitrary identification no.-genotypic traits-phenotypic traits-O serogroup. The genotypic traits include transmembrane regulatory gene (*toxR*), hemolysin subunit A (*hlyA*), heat-stable enterotoxin gene (*stn*), and two outer membrane protein genes (*ompU* and *ompW*), and the phenotypic traits include growth in nutrient broth containing 6% NaCl (NaCl6); acid production from mannose (Mns) and mannitol (Mnt); production of enzymes, including ornithine decarboxylase (Orn), lysine decarboxylase (Lys), amylase (Amy), lipase (Lp), and chitinase (Ch); response to the methyl red test (MR); Voges-Proskauer tests (VP); response to 50 U of polymyxin B (PB); and growth at 42°C (G42oC). Symbols: ●, three identical sequences; ○, five divergent sequences. Scale bars: left bar, number of differences; right bar, substitutions per base pair. The abbreviations defined here are used in Fig. 2.

TABLE 1. Decomposition of variance of trait prevalence and OTU incidence by site and hierarchy of time scales[a]

Component[b]	Covariable(s)[c]	Blocks	Whole plot	RDA		CCA	
				Var[c]	*P*	Var[c]	*P*
Site	Biweek	Biweek	None	0.092	0.002	3.4	0.014
Yr	Site	Site	Season	0.099	0.054	3.6	0.052
Season	Yr + site	Yr	Mo	0.161	0.006	5.9	0.508
Mo	Season + site	Season	Biweek	0.202	0.076	8.3	0.042
Biweek	Mo + site	Mo	None	0.075	0.112	3.7	0.198
Residual	Biweek + site	NA[d]	NA	0.371		25.0	
Total				1.000		49.9	

[a] In RDA, the prevalence of bacterial traits in 64 enrichment flasks was used as the response variables. In CCA, the presence of OTUs in each enrichment flask was used as the response variables.

[b] A two-way nested ANOVA model was determined: site × (year > season > month > biweek), where "A > B" indicates an "upper component > lower component" relationship in the hierarchy of the time scale.

[c] For each variance component, the variance was estimated from residuals, after removal of the variance accounted for by the designated covariables. For the site, the temporal variance was removed by setting the finest times scale (biweek) as the covariable. For the times scales, the spatial variance was always removed, along with the variance of the upper component. Var, variance accounted for only by the variance component in the first column, estimated as eigenvalues.

[d] NA, not applicable. The significance of variance estimate (*P*) was tested by permutation test. Blocks, whole plots, and subplots for time scale variables were selected in parallel to the example of spatial hierarchy in section 9.2, "Hierarchical Analysis of Community Variation and Case Study 6," in the study by Lepš and Šmilauer (9). Whole plots were permuted within a block without permuting subplots. This permutation disrupts the ordering of time points only at the time scale of interest (P. Šmilauer, personal communication), resulting in *P* values under the null hypothesis of randomness in time of occurrence of a data point. The time scale of "biweek" corresponds to an ~2-week difference in summer samplings since spring or fall sampling was done only at monthly intervals.

clonal redundancy was estimated to be 1.5, signifying two bacterial types for every three *V. cholerae* isolates. This level of microdiversity is comparable to that determined from 16S rRNA sequence ribotypes (2.1 × 1,067 clones/516 ribotypes) (1), the *hsp60* DNA sequence (1.7 = 232 isolates/141 alleles) (17), and genomic restriction fragment length polymorphism (1.1 = 206 isolates/180 electrophoretic types) (17).

The distance between OTU pairs was calculated as the number of different phenotypic characteristics, or gene loci, and hierarchal cluster analysis was performed by using the complete linkage method (Fig. 1). The higher-order branches were fixed predominantly with specific allele classes of genotypic traits, e.g., *toxR*, *ompU*, and *ompW*. Other genotypic traits, namely, *stn* and *hlyA*, and phenotypic traits, including serotype, were variable in the lower-order clusters. Polymorphism of the 16S rRNA genes (Fig. 1), only a few bases when determined from the evenly scattered selection of seven strain sequences (18), showed congruence with phenotypic divergence and was compatible with the criterion for a microdiversity cluster (i.e., >99% similarity in any pair of sequences).

The composition of the *V. cholerae* population (Fig. S1 in the supplemental material) and its characteristics varied by geographic location and time of sampling (Fig. 2). To determine how much variation in the clonal diversity of *V. cholerae* populations can be attributed to differences in geographic location and time of sampling, redundancy analysis (RDA) (11) or canonical correspondence analysis (CCA) (15), implemented in CANOCO for Windows version 4.5 (16), was performed on either the incidence of OTUs or the prevalence of characteristics, the latter being defined as the ratio of the number of OTUs with the given characteristic to the number of OTUs in the *V. cholerae* population isolated from a specific enrichment flask. The difference in sampling station and sampling date was used as the factor in the two-way factorial analysis of variance (ANOVA), considering each set of three enrichment flasks as replicates for a given sample, since the effect of the difference in inocula was determined to be insignificant (see Table S2 in the supplemental material). Since decomposition of the total temporal variance by different scales of time can account for the temporal dimension in the succession of *V. cholerae* populations, a nested ANOVA design was used for further variance decomposition analyses (Table 1). Among the time scales, season (3-month periods) was the significant factor in the variation in prevalence of a characteristic (16% of the total variance), whereas only variation by month was significant for the occurrence of an OTU (17% of the total variance). This result strongly indicated that turnover in a component of the population was significant at monthly intervals, whereas turnover in a characteristic was slower, i.e., at 3-month intervals. In other words, the succession of bacterial taxa showing microdiversity occurred monthly, while a proportion of the characteristics, e.g., "core" characteristics, are commonly carried by the population for an entire season.

Water temperature, salinity, pH, chlorophyll *a* concentration, and total bacterial number for the 31 water samples that yielded at least one *V. cholerae* isolate in the survey (10) were normalized by transformations and analyzed for correlation with variation in characteristics of *V. cholerae*, using the multiple regression methods of RDA. Most of the variance (6.9% of the total variance of the characteristics; $P < 0.01$) was explained by pH. The main effect of the other variables was not significant. The predominance of pH over the four other parameters was due to colinearity in the seasonal variation of the parameters (10). When the residuals of the prevalence of characteristics, after accounting for pH, were considered using a reduced RDA model with pH as a covariate, ca. 6.0% of the residual, i.e., 5.6% of the total variance, could be accounted for by salinity ($P < 0.046$). To identify characteristics that correlated significantly with pH and salinity, a correlation analysis was done for each bacterial characteristic. The heat-stable enterotoxin (*stn*) gene was found to be significantly correlated with salinity, as was the presence of *V. cholerae* serogroup O39. In the case of pH, a reverse proportional relationship with mannose and O40 serogroup was observed (Fig. 3).

An ecological association of *V. cholerae* with zooplankton was established in earlier studies (14). Zooplankton composi-

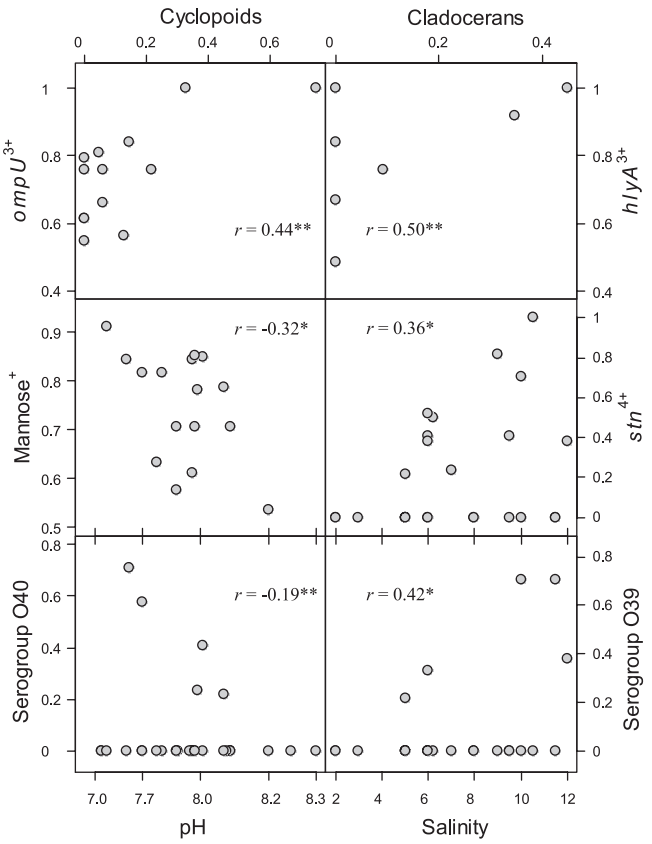


FIG. 3. Distribution of prevalence of selected bacterial traits compared to pH, salinity (middle and bottom rows), and prevalence of zooplankton groups (top row). Labels: *r*, Pearson correlation coefficient; *, significance (*P*) of *r* determined from randomization test of <0.05. The randomization test used the permutation method of toroidal shift, keeping the temporal order of sampling and treating site as permutation block. **, significance of *r* is <0.01.

tion data were available in units of relative abundance for 15 taxa (10), ranging from phylum to suborder (see Fig. S2A in the supplemental material). Zooplankton taxonomic composition was converted to prevalence in a sample by dividing by the total number of zooplankton in a sample. Prevalence values were transformed by square root to calculate the Hellinger distance during the ordination analyses (8). The a priori hypothesis that crustacean zooplankton (notably copepods) provide a microhabitat for specific *V. cholerae* clones had been established, based on the observation of preferential attachment of *V. cholerae* O1 to crustacean exuviae (4, 5, 14). The copepod community in the Chesapeake Bay ecosystem is recognized as one of the most significant trophic components, based on predominance of two species of calanoid copepods (*Acartia tonsa* and *Eurytemora affinis*) in the mesozooplankton communities, both of which demonstrate a strong seasonal and spatial variation (7, 12). Crustaceans were the dominant component of the mesozooplankton community and copepod nauplii were dominant and present in all samples (see Fig. S2A in the supplemental material). The dynamics could be were summarized by two principal axes in the principal component analysis, which classified four groups of taxa (see Fig. S2B in the supplemental material). To test the a priori hypothesis of a biological association of crustaceans and bacterial characteristics, explanation of the variability observed among *V. cholerae* characteristics by dynamics of 10 crustacean taxa was examined. Indeed, the variation observed with respect to the prevalence of members of the crustacean community accounted for nearly half (49.3%) of the total variance of the bacterial characteristics ($P = 0.036$). Forward selection was performed on the prevalence of crustacean taxa to determine the simplest, most effective model for visualizing the mechanism that resulted in this correlation, and cyclopoids and cladocerans were found to be the crustacean taxa significantly accounting for the observed variation in selected bacterial characteristics, either independently (10% variance with $P = 0.030$ and 7% variance with $P = 0.002$ for cyclopoids and cladocerans, respectively) or as a linear additive model (11% variance with $P = 0.040$). None of the other crustacean taxa provided significant accountability for variation in the bacterial characteristics ($P > 0.05$), including the most abundant copepod nauplii ($P = 0.068$). Variance in the bacterial characteristic-crustacean taxon correlation is interpreted as the response of the bacterial population to a shift in crustacean community structure, characterized as a shift from calanoid dominance to cyclopoids or cladocerans (see Fig. S2C in the supplemental material). According to the results of the correlation analysis (Fig. 3), the shift to cyclopoids was correlated with $ompU^{3+}$, whereas the shift to cladocerans was correlated with $hlyA^{3+}$.

In the present study, we examined diversity in a population of *V. cholerae*, with a focus on phenotype and genotype, namely, genes coding for extracellular products involved in interactions with the external physicochemical environment. A significant correspondence was found between phenotypic, genotypic, and serological traits of *V. cholerae* and its environment. Variation in bacterial population structure, namely, seasonal succession or spatial differentiation associated with given bacterial populations, demonstrates a spatiotemporal correspondence with environmental conditions. In the case of the zooplankton community, correspondence occurred at a taxon scale, at which the microdiverse population structure of *V. cholerae* is generated by selective pressure arising from its habitat.

The 16S rRNA sequences determined in the present study were deposited in GenBank under accession no. EF684899 to EF684905.

This research was supported by National Institutes of Health grant 1R01A139129-01 and National Science Foundation grant 01200677.

REFERENCES

1. **Acinas, S. G., V. Klepac-Ceraj, D. E. Hunt, C. Pharino, I. Ceraj, D. L. Distel, and M. F. Polz.** 2004. Fine-scale phylogenetic architecture of a complex bacterial community. Nature **430:**551–554.
2. **Cohan, F. M.** 2002. What are bacterial species? Annu. Rev. Microbiol. **56:**457–487.
3. **Follows, M. J., S. Dutkiewicz, S. Grant, and S. W. Chisholm.** 2007. Emergent biogeography of microbial communities in a model ocean. Science **315:**1843–1846.
4. **Huq, A., E. B. Small, P. A. West, M. I. Huq, R. Rahman, and R. R. Colwell.** 1984. Influence of water temperature, salinity, and pH on survival and growth of toxigenic *Vibrio cholerae* serovar O1 associated with live copepods in laboratory microcosms. Appl. Environ. Microbiol. **48:**420–424.
5. **Huq, A., E. B. Small, P. A. West, R. Rahman, and R. R. Colwell.** 1983. Ecological relationships between *Vibrio cholerae* and planktonic crustacean copepods. Appl. Environ. Microbiol. **45:**275–283.
6. **Kassen, R., and P. B. Rainey.** 2004. The ecology and genetics of microbial diversity. Annu. Rev. Microbiol. **58:**207–231.
7. **Kimmel, D. G., and M. R. Roman.** 2004. Long-term trends in mesozoop-

lankton abundance in Chesapeake Bay, USA: influence of freshwater input. Mar. Ecol. Prog. Ser. **267:**71–83.
8. **Legendre, P., and E. D. Gallagher.** 2001. Ecologically meaningful transformation for ordination of species data. Oecologia **129:**271–280.
9. **Lepš, J., and P. Šmilauer.** 2003. Multivariate analysis of ecological data using CANOCO. Cambridge University Press, Cambridge, United Kingdom.
10. **Louis, V. R., E. Russek-Cohen, N. Choopun, I. N. Rivera, B. Gangle, S. C. Jiang, A. Rubin, J. A. Patz, A. Huq, and R. R. Colwell.** 2003. Predictability of *Vibrio cholerae* in Chesapeake Bay. Appl. Environ. Microbiol. **69:**2773–2785.
11. **Rao, C.** 1964. The use and interpretation of principal component analysis in applied research. Sankhyaá Ser. A **26:**329–358.
12. **Roman, M., X. Zhang, C. McGilliard, and W. Boicourt.** 2005. Seasonal and annual variability in the spatial patterns of plankton biomass in Chesapeake Bay. Limnol. Oceanogr. **50:**480–492.
13. **Shimada, T., E. Arakawa, K. Itoh, T. Okitsu, A. Matsushima, Y. Asai, S. Yamai, T. Nakazato, G. B. Nair, M. J. Albert, and Y. Takeda.** 1994. Extended serotyping scheme for *Vibrio cholerae*. Curr. Microbiol. **28:**175–178.
14. **Tamplin, M. L., A. L. Gauzens, A. Huq, D. A. Sack, and R. R. Colwell.** 1990. Attachment of *Vibrio cholerae* serogroup O1 to zooplankton and phytoplankton of Bangladesh waters. Appl. Environ. Microbiol. **56:**1977–1980.
15. **ter Braak, C. J. F.** 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. Ecology **65:**1167–1179.
16. **ter Braak, C. J. F., and P. Šmilauer.** 2002. CANOCO reference manual and CanoDraw for Windows user's guide: software for canonical community ordination (version 4.5). Microcomputer Power, Ithaca, NY.
17. **Thompson, J. R., S. Pacocha, C. Pharino, V. Klepac-Ceraj, D. E. Hunt, J. Benoit, R. Sarma-Rupavtarm, D. L. Distel, and M. F. Polz.** 2005. Genotypic diversity within a natural coastal bacterioplankton population. Science **307:**1311–1313.
18. **Zo, Y.-G., and R. R. Colwell.** 2008. A simple binomial test for estimating sequencing errors in public repository 16S rRNA sequences. J. Microbiol. Methods **72:**166–179.