Genotypic Diversity Within a Natural Coastal Bacterioplankton Population

Janelle R. Thompson, *et al.*

Science **307**, 1311 (2005);
DOI: 10.1126/science.1106028

18. L. A. Meyerson, K. Saltonstall, L. Windham, E. Kiviat, S. Findlay, *Wetl. Ecol. Management* **8**, 89 (2000).
19. A. H. Al-Mousawi, B. A. Whitton, *Arab Gulf J. Sci. Res.* **1**, 237 (1983).
20. L. A. Richards, Ed., "Diagnosis and improvement of saline and alkali soils" (Agriculture Handbook No. 60. U.S. Department of Agriculture, Washington, DC, 1954).
21. E. V. Maas, *Appl. Agric. Res.* **1**, 12 (1986).
22. E. V. Maas, S. R. Grattan, in *Agricultural Drainage*, R. W. Skaggs, J. Van Schilfgaarde, Eds. (American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, Madison, WI, 1999), pp. 55–110.
23. E. Farber *et al.*, *Geochim. Cosmochim. Acta* **68**, 1989 (2004).
24. A. D. Lemly, *Hum. Ecol. Risk Assess.* **5**, 1139 (1999).
25. A. D. Lemly, *Environ. Monit. Assess.* **28**, 83 (1993).
26. A. D. Lemly, *Ecotoxicol. Environ. Saf.* **10**, 314 (1985).
27. K. M. Banat, F. M. Howari, *J. Arid Environ.*, in press.
28. N. A. Hussain, Ed., *Ahwar of Iraq: An Environmental Approach* (Marine Science Center, Basra University, Basra, Iraq, 1992).
29. M. R. Al-Hilli, thesis, University of Cairo (1977).
30. Funding for this project was provided by the U.S. Agency for International Development through the Iraq Marshlands Restoration Program, implemented by Development Alternatives, Bethesda, MD. The Iraqi scientists at the University of Basra and Baghdad who contributed data for the development of table S3 are regional experts in algae, aquatic plants, fisheries, macroinvertebrates, and mollusca. Given present security concerns about their safety, we have not included their names here. Thanks to A. Douabul of the Iraq Foundation for providing data on Abu Zarag. W. H. Schlesinger, R. S. King, R. DiGiulio and J. W. Pahl reviewed the manuscript at its final stages. Finally, thanks to M. Ho, G. L. Bruland, and R. L. Neighbarger for statistical, graphics, and technical assistance and to H. Partow for helpful discussions and Landsat data on the areas of marsh reflooding.

Supporting Online Material
www.sciencemag.org/cgi/content/full/307/5713/1307/DC1
Materials and Methods
SOM Text
Figs. S1 and S2
Tables S1 to S3
References

28 September 2004; accepted 6 January 2005
10.1126/science.1105750

Genotypic Diversity Within a Natural Coastal Bacterioplankton Population

Janelle R. Thompson,[1,2] Sarah Pacocha,[1,2] Chanathip Pharino,[1] Vanja Klepac-Ceraj,[1,2] Dana E. Hunt,[1] Jennifer Benoit,[1,2] Ramahi Sarma-Rupavtarm,[1] Daniel L. Distel,[3] Martin F. Polz[1]*

The genomic diversity and relative importance of distinct genotypes within natural bacterial populations have remained largely unknown. Here, we analyze the diversity and annual dynamics of a group of coastal bacterioplankton (greater than 99% 16*S* ribosomal RNA identity to *Vibrio splendidus*). We show that this group consists of at least a thousand distinct genotypes, each occurring at extremely low environmental concentrations (on average less than one cell per milliliter). Overall, the genomes show extensive allelic diversity and size variation. Individual genotypes rarely recurred in samples, and allelic distribution did not show spatial or temporal substructure. Ecological considerations suggest that much genotypic and possibly phenotypic variation within natural populations should be considered neutral.

Molecular evidence increasingly demonstrates the remarkable genetic diversity of the microbial world (*1*, *2*), yet ecological interpretation of this diversity remains elusive. This is largely because microbiologists rely on studies of clonal isolates or environmental gene libraries to infer biogeochemical and pathogenic functions of natural bacterial populations. What is missing, however, is quantitative information regarding the environmental prevalence of individual genotypes that would allow inference of their ecological importance or competitive success. It may be expected that ecologically distinct populations display relatively high clonality, because bacterial genomes have a high potential for adaptive mutations, which may lead to purging of genotypic diversity from within the population by selective sweeps (*3*, *4*). However, this view is increasingly difficult to reconcile with recent observations of high levels of differentiation among closely related genomes [e.g., (*5*, *6*)] and the recovery of vast numbers of similar but nonidentical homologous genes from environmental samples (microdiversity) (*7*–*9*). Therefore, questions include whether competition among individual strains is strong enough to result in frequent selective sweeps or instead whether natural populations accumulate large neutral allelic and perhaps even genomic variation (*8*). However, the diversity and prevalence of individual variants within environmental bacterial populations has not been extensively explored, and so questions regarding the ecological importance of genotypic variation remain unanswered.

To analyze genotypic diversity and overall population size quantitatively, we combined culture-dependent and -independent methods to assess the number, extent of variation, and relative frequency of genotypes within a well-defined natural bacterial population. We chose a coastal assemblage of *Vibrio splendidus*, previously identified as a phylogenetically discrete cluster denoted by nearly identical (<1% divergent) 16S rRNA sequences in an analysis of bacterioplankton community structure (*8*). We have proposed that such ribotype clusters represent ecologically differentiated units, i.e., ecotypes or populations (*8*). Thus, we defined the *V. splendidus* cluster as a population of naturally co-occurring genomes that can be tracked quantitatively in the environment and identified in strain collections by their distinct rRNA genes.

Quantification of the *V. splendidus* population over an annual cycle by quantitative polymerase chain reaction (QPCR) (*10*) revealed that it is consistently present as a member of the coastal bacterioplankton community and displays seasonal variation in abundance (Fig. 1A). Concomitant with quantification, we isolated strains from five temporal samples on *Vibrio*-selective media and identified strains by 16S rRNA sequence analysis (Fig. 1B) (*11*). Overall, 20 distinct *Vibrio* (and closely related *Photobacterium*) taxa grew on the media (Fig. 1, B and C), but the majority of isolates (232 of 333) were identified as members of the *V. splendidus* population (red sectors in Fig. 1B). This dominance in all collections, except the cold-water sample (March 2003) (Fig. 1B), roughly parallels the culture-independent quantification by QPCR (Fig. 1A).

Determination of sequence diversity of a universally distributed protein-coding gene (*Hsp60*) among all 333 *Vibrio* isolates showed high heterogeneity but confirmed the monophyly of the *V. splendidus* population detected by the rRNA sequence analysis (*12*). We observed 141 different *Hsp60* alleles among the 232 *V. splendidus* isolates (Fig. 2), and extrapolation using the Chao-1 richness estimator (*13*) suggests a minimum of 436 alleles in

[1]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA. [2]Department of Biology, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA. [3]Ocean Genome Legacy, New England Biolabs, Beverly, MA 01915, USA.

*To whom correspondence should be addressed. E-mail: mpolz@mit.edu

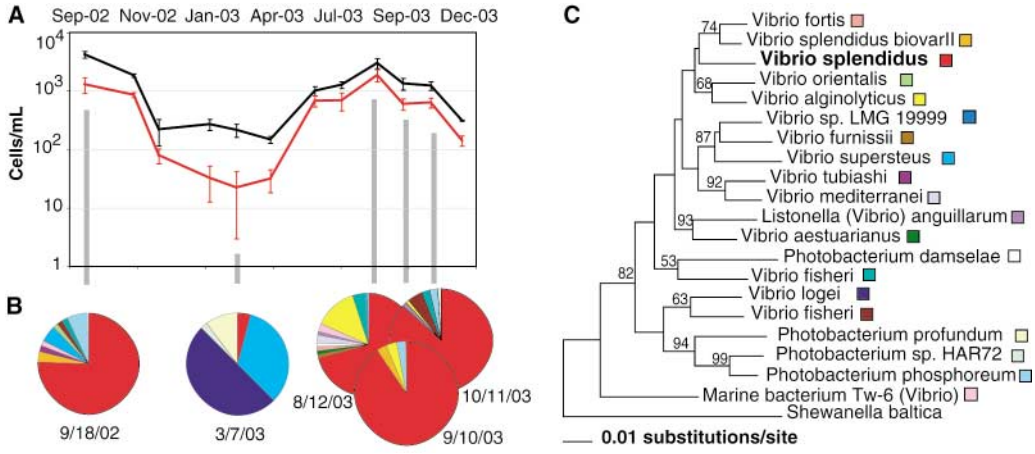

Fig. 1. Diversity and abundance of coastal vibrioplankton (Plum Island Sound, MA) in monthly samples taken over an entire year. (**A**) Quantification of *V. splendidus* (red) and total vibrio (black) populations by culture-independent QPCR. (**B**) Relative proportion of *Vibrio* and *Photobacterium* isolates by phylogenetic association. Color codes correspond to those in (C). (**C**) Phylogenetic relationships among representative *Vibrio* and *Photobacterium* isolates inferred from distance analysis of partial 16S rRNA sequences (bootstrap proportions >50% are indicated above nodes).

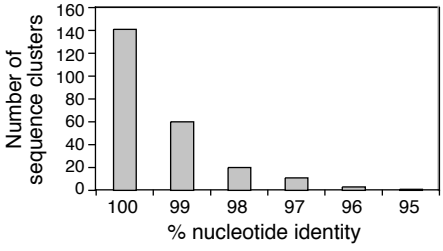

Fig. 2. Number of distinct *Hsp60* clusters among *V. splendidus* isolates observed as cluster cutoff values are decreased from 100 to 95%.

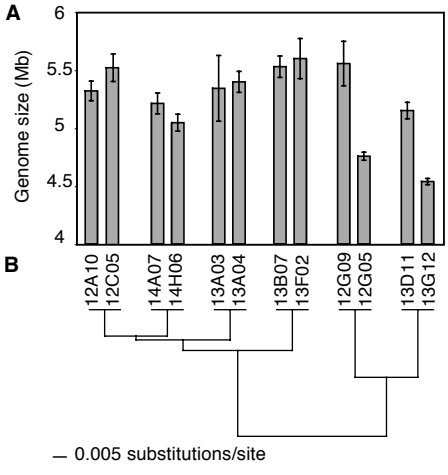

Fig. 3. Genome size estimates and phylogenetic relationships of *Hsp60* sequences for 12 *V. splendidus* isolates chosen as pairs with identical *Hsp60* alleles, encompassing all levels of *Hsp60* variation observed in the strain collection. (**A**) Genome sizes determined by PFGE as averages of six independent estimates, each obtained from single enzyme digests run to resolve large-, medium-, and small-sized bands, respectively, and repeated three times for each of two enzymes (NotI/SfiI or NotI/AscI) per isolate. (**B**) Phylogenetic relationships of *Hsp60* alleles inferred from maximum likelihood analysis with assumption of molecular clock from partial gene sequences. Isolate identifiers correspond to month (12 indicates 12 August 2003; 13, 10 September 2003; 14, 11 October 2003) of isolation and strain name.

the total sampled volume (31.5 ml). Despite these high numbers, the *Hsp60* sequences collapse into a single group at 95% nucleotide consensus (Fig. 2 and fig. S1), with variation primarily limited to neutral third-codon positions [average ratio of nonsynonymous (K_A) to synonymous (K_S) substitutions, K_A/K_S = 0.04]. No single allele showed clear dominance among the isolates, suggesting a relatively even abundance of the strains (fig. S1), and analysis of molecular variance did not reveal evidence for population structure associated with temporal (1 month to 1 year) and spatial (100 µl to 2 ml) scales.

Still greater heterogeneity was revealed within the *V. splendidus* population when the number of distinguishable genotypes among the cultured strains was assayed by pulse field gel electrophoresis (PFGE). PFGE analysis detected 180 different genome patterns among the 206 strains tested, demonstrating that the majority of isolates possess distinct genomes (fig. S1). The Chao-1 estimator (*13*) yielded a total of 1287 genotypes in the samples, considerably exceeding the estimated allelic diversity of the *Hsp60* gene. Moreover, the PFGE analysis provides evidence that, in addition to accumulation of point mutations as observed for the *Hsp60* alleles, a large proportion of genotypes are differentiated by insertions and deletions of large genome fragments. This was suggested by variation among strains in the sum of genome-fragment sizes in the PFGE gels and was explored in detail for a set of 12 strains. Pairs of isolates with identical *Hsp60* alleles were chosen so that sequence identities between pairs reflected overall divergence in these genes (Fig. 3). Among these 12 isolates, genome sizes ranged from 4.5 to 5.6 Mb with only weak correlation of genome size difference to *Hsp60* sequence divergence (r = 0.37) and only minor contribution to size estimates by plasmids (*12*). Even when comparing strains with identical *Hsp60* alleles, most (four of the six pairwise comparisons) showed significant genome size differences (Student's t test, $P < 0.01$; ~170 to 800 kb variation), indicating that specific alleles may be poor markers for distinct genomes.

The high degree of heterogeneity observed among the *V. splendidus* genomes suggests that the average concentration of individual genotypes is small in the sampled environment. To illustrate this, we divided the QPCR-based estimates of population size of *V. splendidus* in samples taken in August, September, and October 2003 (1890, 600, and 640 cells/ml, respectively) (Fig. 1A) by the Chao-1 estimates for the number of *Hsp60* alleles (125, 94, and 279, respectively) and genotypes (465, 553, and 901, respectively) in those same samples. The result suggests that distinct *Hsp60* alleles occurred in the monthly samples at average concentrations of 2 to 15 cells per ml (or at a frequency of 0.3 to 1%), whereas distinct genotypes were present at ~10-fold lower frequency (average concentration for all samples estimated at <1 cell per ml). If the possibility of isolation bias is taken into account, the estimated concentrations (population size/richness) would be even lower, because isolation bias would lead to an underestimation of richness but would not affect population size estimates by QPCR.

What could explain such high diversity of *V. splendidus* genotypes in this environment? The observed pattern suggests that purging of genotypes from within the population (operationally defined as a ribotype cluster) is rare compared with processes introducing variation and that variation persists because it is either favored by selection (e.g., by balancing selection or niche differentiation) or is neutral. Indeed, some proportion of the observed genotypic diversity may reflect the differentiation of (sub)populations that are specialized to particular environmental con-

ditions in the complex life-style of vibrios (including free-living and animal- or particle-associated states). However, ecological considerations suggest that much of the observed genotypic diversity has little adaptive importance in the context of the water column. Given their low estimated concentration (<1 cell/ml), individual genotypes would occupy much less than a trillionth of the volume of a ml of seawater. Because resources are thought to arise in small patches that are unpredictable relative to the location of any given cell (*14*), access by distinct genotypes to conditions allowing rapid growth may be largely stochastic, relegating strong competitive interactions between genotypes to ephemeral microzones. In addition, top-down interactions like predation (*15*) may quickly erase any localized dominance of genotypes. Thus, although individual genotypes may achieve rapid growth in microzones or microcolonies, averaged over the water column their differences do not result in lasting growth advantage (i.e., they are effectively neutral) and so the observed vast genotypic diversity can coexist.

Previous studies have shown that substantial variation in gene content and genome size may occur among closely related genomes (albeit drawn from separate environments) (*6*, *16–18*). Such variation can arise via gene duplication, insertion, and deletion or by horizontal gene transfer (HGT) mediated by phages, plasmid-borne transposons, and integrons (*19*). In fact, HGT is now regarded as a major source of innovation in bacterial evolution (*20–22*), and several cases of environmental differentiation have been linked to specific gene addition or loss (*23–25*). However, it has also been suggested that most acquired sequences do not confer a selective advantage on their host and can be neutral targets for deletion and mutational events, leading to a dynamic genome (*26*, *27*). Indeed, model results indicate that such neutral genome segments are likely to be transient elements represented in only a small fraction of a population (*28*). Our results expand such previous considerations to indicate that large genome modifications, possibly including HGT, are observed with high frequency in genomes that contain identical rRNA and *Hsp60* sequences and coexist within the same natural population.

It will be important to ask whether such extensive genomic variation is a general feature of natural bacterial populations. We have recently observed that two microbial communities are composed of hundreds of microdiverse ribotype clusters (*8*, *9*) and have proposed that these denote bacterial populations that arise by rare selective sweeps followed by effectively neutral diversification (*8*). Consistent with this hypothesis, we show that one such cluster occurs predictably in the bacterioplankton community and contains extensive diversity, much of which may be neutral in the ecological context of the water column. If similar patterns of diversity are common to bacterial communities, caution should be exercised in interpreting the extent to which gene complements or even metabolic traits of individual isolates may reflect the overall properties of populations (*29–31*). Indeed our results suggest that not only the gene content but also quantitative abundance and dynamics of individual traits should be considered when evaluating the ecological importance of differences among coexisting genotypes.

References and Notes

1. M. S. Rappé, S. J. Giovannoni, *Annu. Rev. Microbiol.* **57**, 369 (2003).
2. E. F. DeLong, *Curr. Opin. Microbiol.* **5**, 520 (2002).
3. T. Palys, L. K. Nakamura, F. M. Cohan, *Int. J. Syst. Bacteriol.* **47**, 1145 (1997).
4. S. F. Elena, R. E. Lenski, *Nat. Rev. Genet.* **4**, 457 (2003).
5. E. A. Joyce, K. Chan, N. R. Salama, S. Falkow, *Nat. Rev. Genet.* **3**, 462 (2002).
6. U. Bergthorsson, H. Ochman, *Mol. Biol. Evol.* **15**, 6 (1998).
7. J. C. Venter *et al.*, *Science* **304**, 66 (2004).
8. S. G. Acinas *et al.*, *Nature* **430**, 551 (2004).
9. V. Klepac-Ceraj *et al.*, *Environ. Microbiol.* **6**, 686 (2004).
10. J. R. Thompson *et al.*, *Appl. Environ. Microbiol.* **70**, 4103 (2004).
11. Materials and methods are available as supporting material on *Science* Online.
12. J. Thompson *et al.*, unpublished data.
13. J. B. Hughes, J. J. Hellmann, T. H. Ricketts, B. J. M. Bohannan, *Appl. Environ. Microbiol.* **67**, 4399 (2001).
14. N. Blackburn, T. Fenchel, *Mar. Ecol. Prog. Ser.* **189**, 1 (1999).
15. T. F. Thingstad, *Limnol. Oceanogr.* **45**, 1320 (2000).
16. R. A. Welch *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 17020 (2002).
17. R. A. Alm *et al.*, *Nature* **397**, 176 (1999).
18. J. A. Lindsay, M. T. G. Holden, *Trends Microbiol.* **12**, 378 (2004).
19. J. Hacker, E. Carniel, *EMBO Rep.* **2**, 376 (2001).
20. Y. Boucher *et al.*, *Annu. Rev. Genet.* **37**, 283 (2003).
21. J. G. Lawrence, H. Ochman, *Trends Microbiol.* **10**, 1 (2002).
22. H. Ochman, J. G. Lawrence, E. A. Groisman, *Nature* **405**, 299 (2000).
23. N. T. Perna *et al.*, *Infect. Immun.* **66**, 3810 (1998).
24. S. D. Reid, C. J. Herberlin, A. C. Bumbaugh, R. K. Selander, T. S. Whittam, *Nature* **406**, 64 (2000).
25. G. Rocap *et al.*, *Nature* **424**, 1042 (2003).
26. A. Mira, H. Ochman, N. A. Moran, *Trends Genet.* **17**, 589 (2001).
27. J. G. Lawrence, H. Ochman, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 9413 (1998).
28. O. G. Berg, C. G. Kurland, *Mol. Biol. Evol.* **19**, 2265 (2002).
29. G. E. Fox, J. D. Wisotzkey, J. P. Jurtshuk, *Int. J. Syst. Bacteriol.* **42**, 166 (1992).
30. H. Sass, E. Wieringa, H. Cypionka, H. D. Babenzien, J. Overmann, *Arch. Microbiol.* **170**, 243 (1998).
31. E. Jaspers, J. Overmann, *Appl. Environ. Microbiol.* **70**, 4831 (2004).

32. We thank S. Chisholm, E. DeLong (both MIT), and F. Doolittle (Dalhousie University) for careful readings and comments on the manuscript; S. G. Acinas (MIT) for help with data analysis; J. Fontana and A. Stout (Massachusetts Department of Public Health) for instructions in PFGE optimization; and two anonymous reviewers for their insightful suggestions. This work was supported by grants from U.S. Department of Energy Genomes to Life program and NSF. Sequences have been submitted to GenBank under accession nos. AY836800 to AY837464.

Supporting Online Material

www.sciencemag.org/cgi/content/full/307/5713/1311/DC1
Materials and Methods
Fig. S1

5 October 2004; accepted 4 January 2005
10.1126/science.1106028

Optimization of Virulence Functions Through Glucosylation of *Shigella* LPS

Nicholas P. West,[1]*† Philippe Sansonetti,[2]* Joëlle Mounier,[2] Rachel M. Exley,[1] Claude Parsot,[2] Stéphanie Guadagnini,[3] Marie-Christine Prévost,[3] Ada Prochnicka-Chalufour,[4] Muriel Delepierre,[4] Myriam Tanguy,[2] Christoph M. Tang[1]‡

Shigella, **the leading cause of bacillary dysentery, uses a type III secretion system (TTSS) to inject proteins into human cells, leading to bacterial invasion and a vigorous inflammatory response. The bacterium is protected against the response by the O antigen of lipopolysaccharide (LPS) on its surface. We show that bacteriophage-encoded glucosylation of *Shigella* O antigen, the basis of different serotypes, shortens the LPS molecule by around half. This enhances TTSS function without compromising the protective properties of the LPS. Thus, LPS glucosylation promotes bacterial invasion and evasion of innate immunity, which may have contributed to the emergence of serotype diversity in *Shigella*.**

Pathogenic bacteria have evolved mechanisms to occupy specific niches within hosts while avoiding elimination by innate immune killing. *Shigella* is one of several pathogens that express a type III secretion system (TTSS), a needle-like structure