Resource Partitioning and Sympatric Differentiation Among Closely Related Bacterioplankton

Dana E. Hunt, *et al.*

Science **320**, 1081 (2008);
DOI: 10.1126/science.1157890

Transposon expression in the soma reflects both the silencing of transposons—potentially by either or both posttranscriptional and transcriptional mechanisms—and the tissue specificity of transposon promoters. *Drosophila* somatic cells may contain siRNAs targeting transposons that would not be highly expressed even in the absence of those siRNAs, because the promoters of those transposons are not active in some or all somatic tissues or because they are repressed by additional mechanisms. We analyzed the expression of a panel of transposons in heads from *ago2* and *dcr-2* mutants and in S2 cells depleted of Dcr-1, Dcr-2, or Ago2 by RNAi (Fig. 3 and fig. S8). We found that the steady-state abundance of RNA from the LTR retrotransposons *297* and *412* increased in heads from *dcr-2*^{L811fsX} null mutants (Fig. 3A). Similarly, the steady-state abundance of RNA from the LTR retrotransposons *297*, *412*, *mdg1*, and *roo*, the non-LTR retrotransposon *F-element*, and the SINE-like element *INE-1* increased in *ago2*^{414} mutant heads (Fig. 3B).

In S2 cells, RNA expression from the LTR retrotransposons *297*, *1731*, *mdg1*, *blood*, and *gypsy* and from the DNA transposon *S-element* all increased significantly ($0.00001 < P < 0.002$) when Dcr-2 was depleted or when both Dcr-2 and Dcr-1 were depleted, but not when Dcr-1 alone was depleted (Fig. 3C). Similarly, *ago2(RNAi)* in S2 cells desilenced transposons, including nine LTR and non-LTR retrotransposons and the DNA transposon *S-element* (fig. S8).

Is Ago2 required for the production or accumulation of endo-siRNAs? We sequenced 18- to 29-nt small RNAs from *ago2*^{414} homozygous fly heads and from the same small RNA sample treated to enrich for 3′-terminally modified RNAs. After computationally removing miRNAs, the sequences from the untreated library contained a prominent 21-nt peak (Fig. 4A) that predominantly began with uracil (Fig. 4B), much like miRNAs and unlike siRNAs in wild-type heads, which often began with cytosine (Fig. 1A). Perhaps in the absence of Ago2, only a subpopulation of endo-siRNAs that can bind Ago1 accumulates. The small RNAs from the *ago2*^{414} library enriched for 3′-terminally modified sequences were predominantly 24 to 27 nt long and often began with uracil—a length distribution and sequence bias characteristic of piRNAs, which, like siRNAs, are 2′-*O*-methylated at their 3′ ends. Both the 21-nt small RNAs and the piRNA-like RNAs in the *ago2* mutant heads mapped to transposons, unannotated heterochromatic and unassembled sequences, but the piRNA-like sequences mapped to mRNAs far less frequently than did either the 21-mers or wild-type endo-siRNAs (Fig. 4C). How these piRNA-like small RNAs are generated and whether they contribute to transposon silencing in the fly soma remain unknown.

Note added in proof: The loci described here in figs. S1 and S2 correspond to endo-siRNA–generating hairpins recently identified in (*25*–*27*).

References and Notes

1. L. S. Gunawardane *et al.*, *Science* **315**, 1587 (2007); published online 21 February 2007 (10.1126/science.1140494).
2. J. Brennecke *et al.*, *Cell* **128**, 1089 (2007).
3. V. V. Vagin *et al.*, *Science* **313**, 320 (2006); published online 28 June 2006 (10.1126/science.1129333).
4. H. Yin, H. Lin, *Nature* **450**, 304 (2007).
5. A. Hamilton, O. Voinnet, L. Chappell, D. Baulcombe, *EMBO J.* **21**, 4671 (2002).
6. R. Sunkar, T. Girke, J. K. Zhu, *Nucleic Acids Res.* **33**, 4443 (2005).
7. N. Yang, H. H. J. Kazazian, *Nat. Struct. Mol. Biol.* **13**, 763 (2006).
8. A. Ishizuka, M. C. Siomi, H. Siomi, *Genes Dev.* **16**, 2497 (2002).
9. A. K. Csink, R. Linsk, J. A. Birchler, *Genetics* **138**, 153 (1994).
10. J. Rehwinkel *et al.*, *Mol. Cell. Biol.* **26**, 2965 (2006).
11. K. A. Haynes, A. A. Caudy, L. Collins, S. C. Elgin, *Curr. Biol.* **16**, 2222 (2006).
12. M. Pal-Bhadra, U. Bhadra, J. A. Birchler, *Mol. Cell* **9**, 315 (2002).
13. N. A. Tchurikov, O. V. Kretova, *PLoS ONE* **2**, e476 (2007).
14. M. D. Horwich *et al.*, *Curr. Biol.* **17**, 1265 (2007).
15. A. Pelisson, E. Sarot, G. Payen-Groschene, A. Bucheton, *J. Virol.* **81**, 1951 (2007).
16. K. Saito *et al.*, *Genes Dev.* **21**, 1603 (2007).
17. K. Okamura, A. Ishizuka, H. Siomi, M. C. Siomi, *Genes Dev.* **18**, 1655 (2004).
18. A. Nykanen, B. Haley, P. D. Zamore, *Cell* **107**, 309 (2001).
19. *Drosophila* RNAi Screening Center at Harvard Medical School (http://flyrnai.org/cgi-bin/RNAi_FAQ_lines.pl).
20. Y. S. Lee, R. W. Carthew, *Methods* **30**, 322 (2003).
21. H. Seitz, M. Ghildiyal, P. D. Zamore, *Curr. Biol.* **18**, 147 (2008).
22. P. D. Zamore, T. Tuschl, P. A. Sharp, D. P. Bartel, *Cell* **101**, 25 (2000).
23. S. Desset, N. Buchon, C. Meignin, M. Coiffet, C. Vaury, *PLoS ONE* **3**, e1526 (2008).
24. See supporting material on *Science* Online.
25. B. Czech *et al.*, *Nature* 10.1038/nature07007 (2008).
26. Y. Kawamura *et al.*, *Nature* 10.1038/nature06938 (2008).
27. K. Okamura *et al.*, *Nature* 10.1038/nature07015 (2008).

28. We thank A. Boucher and S. Ma for technical assistance; G. Farley for encouragement, support, and technical assistance; and Roche Applied Science for high-throughput sequencing. P.D.Z. is a W. M. Keck Foundation Young Scholar in Medical Research. Supported by NIH grants GM62862 and GM65236 (P.D.Z.), GM080625 (J.X. and Z.W.), and HG003367 (S.L.); EMBO long-term (ALTF 910-2004) and Human Frontier Science Program (LT00575/2005-L) fellowships (H.S.); and a National Research Service Award predoctoral MD/PhD fellowship from the National Institute on Aging (F30AG030283) (M.D.H.). NCBI Gene Expression Omnibus accession numbers for sequence and abundance data are GSE9389 and GSE11019, respectively.

Supporting Online Material

www.sciencemag.org/cgi/content/full/1157396/DC1
Materials and Methods
Figs. S1 to S8
Tables S1 to S7
References

5 March 2008; accepted 31 March 2008
Published online 10 April 2008;
10.1126/science.1157396
Include this information when citing this paper.

Resource Partitioning and Sympatric Differentiation Among Closely Related Bacterioplankton

Dana E. Hunt,[1]* Lawrence A. David,[2]* Dirk Gevers,[1,3,4] Sarah P. Preheim,[1] Eric J. Alm,[1,5]† Martin F. Polz[1]†

Identifying ecologically differentiated populations within complex microbial communities remains challenging, yet is critical for interpreting the evolution and ecology of microbes in the wild. Here we describe spatial and temporal resource partitioning among *Vibrionaceae* strains coexisting in coastal bacterioplankton. A quantitative model (AdaptML) establishes the evolutionary history of ecological differentiation, thus revealing populations specific for seasons and life-styles (combinations of free-living, particle, or zooplankton associations). These ecological population boundaries frequently occur at deep phylogenetic levels (consistent with named species); however, recent and perhaps ongoing adaptive radiation is evident in *Vibrio splendidus*, which comprises numerous ecologically distinct populations at different levels of phylogenetic differentiation. Thus, environmental specialization may be an important correlate or even trigger of speciation among sympatric microbes.

Microbes dominate biomass and control biogeochemical cycling in the ocean, but we know little about the mechanisms and dynamics of their functional differentiation in the environment. Culture-independent analysis typically reveals vast microbial diversity, and although some taxa and gene families are differentially distributed among environments (*1*, *2*), it is not clear to what extent coexisting genotypic diversity can be divided into functionally cohesive populations (*1*, *3*). First, we lack broad surveys of nonpathogenic free-living bacteria that establish robust associations of individual strains with spatiotemporal conditions (*4*, *5*); second, it remains controversial what level of genetic diversification reflects ecological differentiation. Phylogenetic clusters have been proposed to correspond to ecological populations that arise by neutral diversification after niche-specific selective sweeps (*6*). Clusters are indeed observed among closely related isolates (e.g., when examined by multilocus sequence analysis) (*7*) and in culture-independent analyses of coastal bacterioplankton (*8*). Yet recent theoretical studies suggest

that clusters can result from neutral evolution alone (*9*), and evidence for clusters as ecologically distinct populations remains sparse, having been most conclusively demonstrated for cyanobacteria along ocean-scale gradients (*10*) and in a depth profile of a microbial mat (*11*). Further, horizontal gene transfer (HGT) may erode the ecological cohesion of clusters if adaptive genes are transferred (*12*), and recombination can homogenize genes between ecologically distinct populations (*13*). Thus, exploring the relationship between phylogenetic and ecological differentiation is a critical step toward understanding the evolutionary mechanisms of bacterial speciation (*9*).

In this study, we investigated ecological differentiation by spatial and temporal resource partitioning in coastal waters among coexisting bacteria of the family *Vibrionaceae*, which are ubiquitous, metabolically versatile heterotrophs (*14*). The coastal ocean is well suited to test population-level effects of microhabitat preferences, because tidal mixing and oceanic circulation ensure a high probability of migration, reducing biogeographic effects on population structure. In the plankton, heterotrophs may adopt alternate ecological strategies: exploiting either the generally lower concentration but more evenly distributed dissolved nutrients or attaching to and degrading small suspended organic particles, originating from algal exopolysaccharides and detritus (*3*). Bacterial microhabitat preferences may develop because resources are distributed on the same scale as the dispersal range of individuals, due to turbulent mixing and active motility (*15*). Of potential microhabitats, particles represent abundant but relatively short-lived resources, as labile components are rapidly utilized (on time scales of hours to days) (*16*, *17*), implying that particle colonization is a dynamic process. Moreover, particulate matter may change composition with macroecological conditions (such as seasonal algal blooms). Zooplankton provide additional, more stable microhabitats; vibrios attach to and metabolize chitinous zooplankton exoskeletons (*18*, *19*) but may also live in the gut or occupy niches specific to pathogens. The extent to which microenvironmental preferences contribute to resource partitioning in this complex ecological landscape remains an important question in microbial ecology (*20*).

We aimed to conservatively identify ecologically coherent groups by examining distribution patterns of *Vibrionaceae* genotypes among free-living and associated (with suspended particles and zooplankton) compartments of the plankton-

[1]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA. [2]Computational and Systems Biology Initiative, MIT, Cambridge, MA 02139, USA. [3]Laboratory of Microbiology, Ghent University, Gent 9000, Belgium. [4]Bioinformatics and Evolutionary Genomics Group, Ghent University, Gent 9000, Belgium. [5]Department of Biological Engineering, MIT, Cambridge, MA 02139, USA, and Broad Institute of MIT and Harvard University, Cambridge, MA 02139, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: ejalm@mit.edu (E.J.A.); mpolz@mit.edu (M.F.P.)

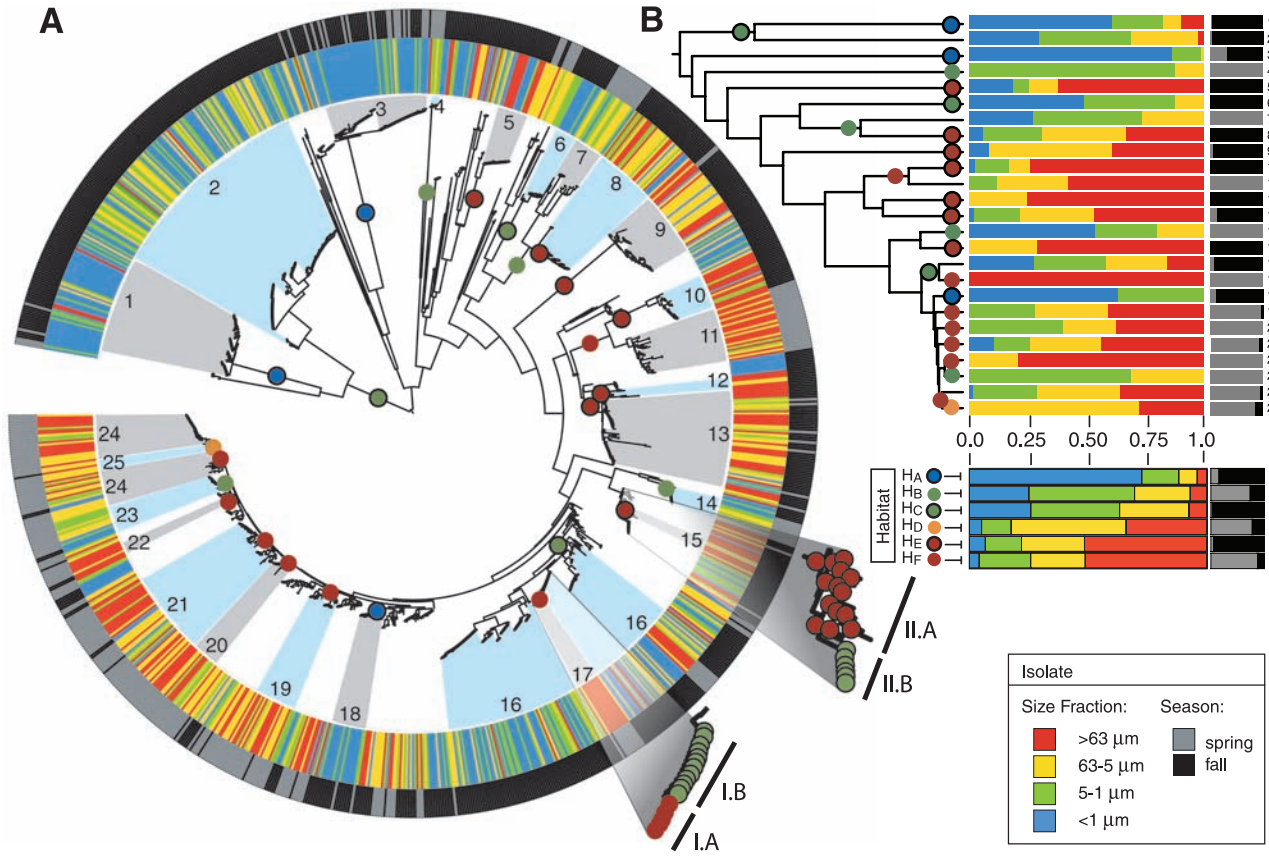

Fig. 1. Season and size fraction distributions and habitat predictions mapped onto *Vibrionaceae* isolate phylogeny inferred by maximum likelihood analysis of partial *hsp60* gene sequences. Projected habitats are identified by colored circles at the parent nodes. (**A**) Phylogenetic tree of all strains, with outer and inner rings indicating seasons and size fractions of strain origin, respectively. Ecological populations predicted by the model are indicated by alternating blue and gray shading of clusters if they pass an empirical confidence threshold of 99.99% (see SOM for details). Bootstrap confidence levels are shown in fig. S10. (**B**) Ultrametric tree summarizing habitat-associated populations identified by the model and the distribution of each population among seasons and size fractions. The habitat legend matches the colored circles in (A) and (B) with the habitat distribution over seasons and size fractions inferred by the model. Distributions are normalized by the total number of counts in each environmental category to reduce the effects of uneven sampling. The insets at the lower right of (A) show two nested clusters (I.A and I.B and II.A and II.B) for which recent ecological differentiation is inferred, including habitat predictions at each node. The closest named species to numbered groups are as follows: G1, *V. calviensis*; G2, *Enterovibrio norvegicus*; G3, *V. ordalii*; G4, *V. rumoiensis*; G5, *V. alginolyticus*; G6, *V. aestuarianus*; G7, *V. fischeri/logei*; G8, *V. fischeri*; G9, *V. superstes*; G10, *V. penaeicida*; G11 to G25, *V. splendidus*.

ic environment under different macroecological conditions (spring and fall) (fig. S1 and table S1). Because the level of genetic differentiation at which ecological preferences develop is not known, we focused on a range of relationships [0 to 10% small subunit ribosomal RNA (rRNA) divergence] among co-occurring vibrios (*21*). Particle-associated and free-living cells were separated into four consecutive size fractions by sequential filtration (four replicate water samples, each subsampled with at least four replicate filters per size fraction); each fraction contained organisms and dead organic material of different origins [detailed in the supporting online material (SOM)]. For simplicity, we refer to these fractions as enriched in zooplankton (≥63 μm), in large (5 to 63 μm) and small (1 to 5 μm) particles, and in free-living cells (0.22 to 1 μm) (fig. S1B). The 1- to 5-μm size fraction was somewhat ambiguous, probably containing small particles as well as large or dividing cells; however, it provided a firm buffer between obviously particle-associated (>5 μm) and free-living (<1 μm) cells. *Vibrionaceae* strains were isolated by plating filters on selective media, previously shown by quantitative polymerase chain reaction to yield good correspondence between genotypes recovered in culture and those present in environmental samples (*21*). Roughly 1000 isolates were characterized by partial sequencing of a protein-coding gene (*hsp60*). To obtain added resolution, between one and three additional gene fragments

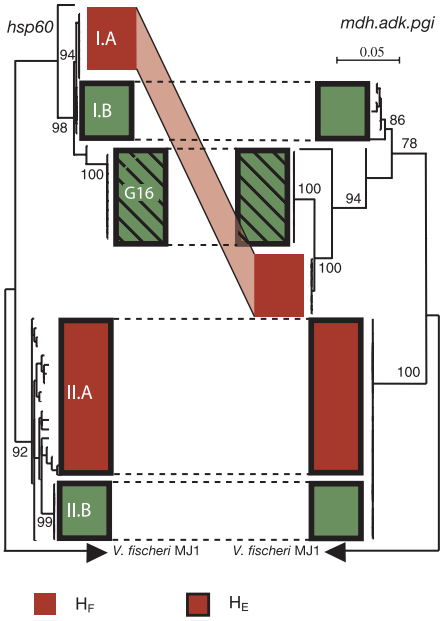

Fig. 2. Multilocus sequence analysis of nested clusters (IA and IB and IIA and IIB) with differential habitat association by comparison of partial *hsp60* (left) and concatenated partial *mdh*, *adk*, and *pgi* (right) gene phylogenies. Habitat predictions (indicated by colored boxes) and the numbering of clusters correspond to Fig. 1. Scale bar is in units of nucleotide substitutions per site.

(*mdh*, *adk*, and *pgi*) were sequenced for over half of the isolates (SOM), including *V. splendidus* strains, the most abundant group (*21*).

Our rationale for testing environmental associations grows out of the following considerations. First, as in most ecological sampling, the true habitats or niches are unknown and can only be observed as projections onto the sampling dimensions ("projected habitats"). Thus, associations can be detected as distinct distributions of groups of strains if habitats/niches are differentially apportioned among samples. Second, the lack of an accepted microbial species concept implies that it is imprudent to use any measure of genetic relationships to define a priori the populations whose environmental association should be assessed. Therefore, we first tested the null hypothesis that there is no environmental association across the phylogeny of the strains. We then refined such estimates by developing a new model to simultaneously identify populations and their projected habitats. Finally, these model-based results were tested with nonparametric empirical statistics.

The initial null hypothesis of no association between phylogeny and ecology is strongly rejected (seasons: $P < 10^{-79}$; size fractions: $P < 10^{-49}$) by comparing the parsimony score of observed environments on the tree to that expected by chance (*22*) (SOM), confirming the visual impression of differential patterns of clustering among seasons and size fractions (Fig. 1A). This result is robust toward uncertainty in the phylogeny, which should diminish but not strengthen associations, and is confirmed by introducing additional uncertainty in the phylogeny (fig. S2). The observed overall association with season and size fraction therefore suggests that water-column vibrios partition resources, but neither provides insights into the phylogenetic bounds of populations or the composition of their habitats.

We therefore developed an evolutionary model (AdaptML) to identify populations as groups of related strains sharing a common projected habitat, which reflects their relative abundance in the measured environmental categories (size fractions and seasons) (SOM). In practice, the model inputs are the phylogeny, season, and size fraction of the strains. It then maps changes in environmental preference onto the tree by predicting projected habitats for each extant and ancestral strain in the phylogeny. Although similar in spirit to existing parsimony, likelihood, and Bayesian methods, which map ancestral states onto trees (*23*), the model accounts for the complexities and uncertainties of environmental sampling. First, projected habitats can span multiple sampling dimensions to account for complex life cycles (such as time spent in multiple true habitats) and problems inherent in environmental sampling: Discrete samples rarely equate to true habitats, and true habitats are frequently misplaced among their typical sample categories (for example, zooplankton fragments may also be found in smaller size fractions). Second, projected habitats can span multiple phylogenetic clusters to allow for the possibility that clusters may arise neutrally or that the relevant parameters differentiating them ecologically have not been measured.

Briefly, AdaptML builds a hidden Markov model for the evolution of habitat associations: Adjacent nodes on the phylogeny transition between habitats according to a probability function that is dependent on branch length and a transition rate, which is learned from the data (SOM) (fig. S3). Subsequently, we optimize the model parameters (the transition rate and the composition of each projected habitat) to maximize the likelihood of the observed data. Finally, we use a simple ad hoc rule for reducing noninformative parameters: We merge habitats that converge to similar distributions (simple correlation of distribution vectors >90%) during the model-fitting procedure (SOM). This reproducibly identified six nonredundant habitats for the observed data set (H_A to H_F in Fig. 1B and fig. S5). Moreover, the algorithm acts conservatively, as suggested by two tests. First, the model did not overfit the data when there was no ecological signal present: When the environments were shuffled, only a single generalist habitat (evenly distributed over all size fractions and seasons) was recovered. Second, when simulated habitats were used to generate environmental assignments, the model usually identified a number of habitats equal to or less than the true number present (fig. S6).

The analysis suggests that a single bacterial family coexisting in the water column resolves into a striking number of ecologically distinct populations with clearly identifiable preferences (habitats). The algorithm identified 25 populations, associated with one of the six habitats defined by distinct distributions of isolates over seasons and size fractions (Fig. 1 and fig. S7). Most clusters have a strong seasonal signal; interestingly, two pairs of highly similar habitats are observed in both seasons (Fig. 1B). The first of the habitat pairs corresponds to populations occurring both free-living and on particles but lacking zooplankton-associated isolates (H_B and H_C); the second indicates a preference for zooplankton and large particles (H_E and H_F) (Fig. 1B). The remaining two habitats were season-specific. Habitat H_A combines all primarily free-living populations in the fall, whereas habitat H_D identifies a second particle- and zooplankton-associated group in spring, but unlike H_E and H_F it has a higher proportion of large particles and maps onto a single small group (G25) (Fig. 1). However, we cannot place high confidence in the absence of the free-living habitat in the spring, because relatively few strains were recovered from that fraction. Moreover, the distribution of individual populations among seasons and size fractions varies considerably, with remarkably narrow preferences for some populations whereas others are more broadly distributed. For example, *V. ordalii* (G3) is almost exclusively free-living in both seasons, whereas *V. alginolyticus* (G5) has a significant representation in both zooplankton and free-living size fractions but

occurs exclusively in the fall (Fig. 1, A and B). The sequences of three additional genes for *V. alginolyticus* isolates were identical, arguing against misidentification due to recombination or additional population substructuring. Similarly, there was good agreement when two different gene phylogenies (*hsp60* and *mdh*) were used to identify habitats for *V. splendidus* (fig. S8), although fewer habitats were identified using the *mdh* tree, most likely because it is less well-resolved. Overall, across all vibrios sampled, association with the zooplankton-enriched and free-living fractions dominated, and although several populations contain particle-associated isolates, only a few appear to be specifically particle-adapted. Because vibrios are generally regarded as particle and zooplankton specialists (*14*), this observed partitioning offers new insight into their ecology.

Thus, in spite of the highly variable conditions of the water column, populations appear to finely partition resources, especially because our habitat estimates are conservative, as clusters occupying the same habitat may be differentiated along additional (unobserved) resource axes. For example, different zooplankton-associated groups may be host- or body region–specific, and the strong seasonal signal of most clusters may be due to a variety of factors; however, temperature is a likely candidate because it has so far arisen as the strongest correlate of microbial population changes both over a seasonal cycle (*24*) and along ocean-scale gradients (*10*). Finally, populations, which appear unassociated in our study, may be true generalists with respect to the resource space sampled or may be adapted to environments not sampled in this study, such as animal intestines or sediments (*14*). Despite these uncertainties, the observed strong partitioning among associated and free-living clusters may have important implications for population biology in the bacterioplankton. As recently suggested (*9*), for attached bacteria, the effective population size (N_e) may be considerably smaller than the census size because colonization serves as a population bottleneck, whereas in free-living clusters, N_e may be closer to the census size. Although computing the true magnitude of N_e in microbial populations remains controversial (*25*), it is an important parameter that determines the relative strength of selection and drift. Thus, attached and free-living populations may evolve under different constraints (*9*).

The phylogenetic structure of populations also provides insights into the history of habitat switches. Deeply branching populations may have remained associated with habitats over long evolutionary time, and shallow branches may have diversified more recently (Fig. 1, A and B). These stable habitat-associated clusters roughly correlate to named species within the *Vibrionaceae*. For example, *V. ordalii* (G3) and *Enterovibrio norvegicus* (G2) both represent clusters without close relatives containing >50 isolates, which are overwhelmingly predicted to follow primarily free-living (H_A) and free-living/particle-associated life-styles (H_C), respectively (Fig. 1A). On the other hand, some very closely related clusters are associated with different habitats; *V. splendidus*, which is composed of strains that are ~99% identical in rDNA gene sequence (*21*), differentiates into 15 microdiverse habitat-associated clusters, of which one is distributed roughly evenly among both seasons, and 9 and 5 predominantly occur in spring and fall, respectively. Thus, *V. splendidus* appears to have ecologically diversified, possibly by invading new niches or partitioning resources at increasingly fine scales.

Recent or perhaps ongoing radiation by sympatric resource partitioning is most strongly suggested for two nested clusters within *V. splendidus*, where groups of strains differing by as little as a single nucleotide in *hsp60* display distinct ecological preferences (Fig. 1A, insets, and table S2). These strains were isolated from multiple independent samples and thus do not represent clonal expansion, suggesting that this may reflect a true habitat switch; nonetheless, homologous recombination could also move alleles between distantly related, ecologically distinct clusters, creating spurious phylogenetic relationships, which can be detected by comparison with other genes. Multilocus sequence analysis shows that for nested cluster I, a close relationship was artificially created because *hsp60* gene phylogeny is discordant with three other genes (Fig. 2). However, this still represents a habitat switch, just at a slightly larger sequence distance, as I.A is nested within the much larger G16 cluster in both the *hsp60* and the *mdh-pgi-adk* phylogenies. For the second nested cluster, the three additional genes confirm partial separation of the subclusters II.A and II.B by a single base pair difference in one of the genes, whereas the other genes consist of identical alleles. This reinforces the idea that subcluster II.A is not incorrectly grouped because of recombination, despite its distinct ecological affiliation (Fig. 2). In combination, these data support the idea that there is ecological differentiation among recently diverged genotypes and show that such changes might be recognized in protein-coding genes as soon as they accumulate (neutral) sequence changes.

How might adaptation to a new habitat relate to speciation, the generation of distinct clusters of closely related bacteria? Mathematical modeling has recently shown that the dynamics of speciation depend on the ratio of homologous recombination to mutation rates (r/m) (*9*). When this ratio per allele exceeds ~1, populations transition from essentially clonal to sexual, with the major consequence that selection is probably required for the formation of clusters (*9*). Our preliminary multilocus sequence analysis on a set of strains with similar taxonomic composition suggests that their r/m is well above that threshold. Thus, our observations of habitat separation for highly similar but clearly distinct genotypes suggest that ecological selection may have triggered phylogenetic differentiation. A plausible mechanism is that differential distribution among habitats (possibly caused by few adaptive loci) is sufficient to depress gene flow between associated genotypes (*9*, *26*). Consequently, mutations will no longer be homogenized but instead accumulate within specialized populations, even for ecologically neutral genes. Over time, genetic isolation may increase because homologous recombination rates decrease log-linearly with sequence distance (*27*). We detected associations with different habitats among sister clades over a wide range of phylogenetic distances, possibly representing populations at various stages of differentiation (Fig. 1A). Although we cannot determine whether clusters represent transiently adapted populations or nascent species, our observations of differential distributions of genotypes suggest that there exists a small-scale adaptive landscape in the water column allowing the initiation of (sympatric) speciation within this community.

Although it has recently been suggested that microbial lineages remain specific to macroenvironments over long evolutionary times (*28*), this study demonstrates switches in ecological associations within a bacterial family coexisting in the coastal ocean. In the *V. splendidus* clade, speciation could be ongoing, but the divergence between most other ecologically defined groups appears large. This is consistent with our previous suggestion that rRNA gene clusters, which are roughly congruent with the deeply divergent protein-coding gene clusters detected here, represent ecological populations (*8*). However, the example of *V. splendidus* highlights the fact that using marker genes to assess community-wide diversity may not capture some ecological specialization. Moreover, different groups of organisms could evolve under different constraints, and the mechanisms suggested here apply to the invasion of new habitats and are thus different from (but compatible with) the widely discussed niche-specific selective sweeps (*29*). Why *V. splendidus* appears to have radiated recently into new habitats whereas other groups appear to be more constant is not known but may be related to its high heterogeneity in genome architecture (*21*). This could indicate a large (flexible) gene pool that, if shared by horizontal gene transfer, gives rise to large numbers of ecologically adaptive phenotypes. It will therefore be important to compare whole genomes within recently ecologically diverged clusters to identify specific changes leading to adaptive evolution.

References and Notes

1. S. J. Giovannoni, U. Stingl, *Nature* **437**, 343 (2005).
2. E. F. DeLong *et al.*, *Science* **311**, 496 (2006).
3. M. F. Polz, D. E. Hunt, S. P. Preheim, D. M. Weinreich, *Philos. Trans. R. Soc. London Ser. B* **361**, 2009 (2006).
4. A. Ramette, J. M. Tiedje, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 2761 (2007).
5. J. B. H. Martiny *et al.*, *Nat. Rev. Microbiol.* **4**, 102 (2006).
6. F. M. Cohan, E. B. Perry, *Curr. Biol.* **17**, R373 (2007).
7. D. Gevers *et al.*, *Nat. Rev. Microbiol.* **3**, 733 (2005).
8. S. G. Acinas *et al.*, *Nature* **430**, 551 (2004).
9. C. Fraser, W. P. Hanage, B. G. Spratt, *Science* **315**, 476 (2007).

10. Z. I. Johnson *et al.*, *Science* **311**, 1737 (2006).
11. M. J. Ferris, M. Kühl, A. Wieland, D. M. Ward, *Appl. Environ. Microbiol.* **69**, 2893 (2003).
12. W. F. Doolittle, R. T. Papke, *Genome Biol.* **7**, 116 (2006).
13. A. C. Retchless, J. G. Lawrence, *Science* **317**, 1093 (2007).
14. J. R. Thompson, M. F. Polz, in *The Biology of Vibrios*, F. L. Thompson, B. Austin, J. Swings, Eds. (American Society for Microbiology Press, Washington, DC, 2006), pp. 190–203.
15. T. Kiørboe, H. P. Grossart, H. Ploug, K. Tang, *Appl. Environ. Microbiol.* **68**, 3996 (2002).
16. L. M. Pomeroy *et al.*, *Bull. Mar. Sci.* **35**, 426 (1984).
17. C. Panagiotopoulos *et al.*, *Organ. Geochem.* **33**, 985 (2002).
18. J. F. Heidelberg, K. B. Heidelberg, R. R. Colwell, *Appl. Environ. Microbiol.* **68**, 5498 (2002).
19. D. E. Hunt, D. Gevers, N. M. Vahora, M. F. Polz, *Appl. Environ. Microbiol.* **74**, 44 (2008).
20. J. Pernthaler, R. Amann, *Microbiol. Mol. Biol. Rev.* **69**, 440 (2005).
21. J. R. Thompson *et al.*, *Science* **307**, 1311 (2005).
22. W. P. Maddison, M. Slatkin, *Evolution* **45**, 1184 (1991).
23. F. Ronquist, *Trends Ecol. Evol.* **19**, 475 (2004).
24. J. R. Thompson *et al.*, *Appl. Environ. Microbiol.* **70**, 4103 (2004).
25. V. Daubin, N. A. Moran, *Science* **306**, 978 (2004).
26. F. M. Cohan, *Genetica* **116**, 359 (2002).
27. J. Majewski, *FEMS Microbiol. Let.* **199**, 161 (2001).
28. C. von Mering *et al.*, *Science* **315**, 1126 (2007).
29. A. Koeppel *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 2504 (2008).
30. We thank E. DeLong and D. Distel for critical reading of the manuscript. This work was supported by grants from the U.S. Department of Energy Genomes to Life program to M.F.P. and E.J.A. and by support from the NSF/National Institute of Environmental Health Sciences Woods Hole Center for Oceans and Human Health, the NSF Biological Oceanography Program, and the Moore Foundation to M.F.P. D.G. is indebted to the Fund for Scientific Research, Flanders (Belgium), for a postdoctoral fellowship and research funding. L.A.D. gratefully acknowledges support from a U.S. Department of Defense National Defense Science and Engineering Graduate fellowship. Sequences have been submitted to GenBank under accession nos. EU653713 to EU655560.

Supporting Online Material

www.sciencemag.org/cgi/content/full/320/5879/1081/DC1
Materials and Methods
Figs. S1 to S10
Tables S1 and S2
References

17 March 2008; accepted 18 April 2008
10.1126/science.1157890

A Polymorphism Within the *G6PC2* Gene Is Associated with Fasting Plasma Glucose Levels

Nabila Bouatia-Naji,[1]* Ghislain Rocheleau,[2]* Leentje Van Lommel,[3] Katleen Lemaire,[3] Frans Schuit,[3] Christine Cavalcanti-Proença,[1] Marion Marchand,[1] Anna-Liisa Hartikainen,[4] Ulla Sovio,[5] Franck De Graeve,[1] Johan Rung,[2] Martine Vaxillaire,[1] Jean Tichet,[6] Michel Marre,[7] Beverley Balkau,[8] Jacques Weill,[9] Paul Elliott,[5] Marjo-Riitta Jarvelin,[5,10] David Meyre,[1] Constantin Polychronakos,[2,11] Christian Dina,[1] Robert Sladek,[2] Philippe Froguel[1,12]†

Several studies have shown that healthy individuals with fasting plasma glucose (FPG) levels at the high end of the normal range have an increased risk of mortality. To identify genetic determinants that contribute to interindividual variation in FPG, we tested 392,935 single-nucleotide polymorphisms (SNPs) in 654 normoglycemic participants for association with FPG, and we replicated the most strongly associated SNP (rs560887, $P = 4 \times 10^{-7}$) in 9353 participants. SNP rs560887 maps to intron 3 of the *G6PC2* gene, which encodes glucose-6-phosphatase catalytic subunit–related protein (also known as IGRP), a protein selectively expressed in pancreatic islets. This SNP was associated with FPG (linear regression coefficient $\beta = -0.06$ millimoles per liter per A allele, combined $P = 4 \times 10^{-23}$) and with pancreatic β cell function (Homa-B model, combined $P = 3 \times 10^{-13}$) in three populations; however, it was not associated with type 2 diabetes risk. We speculate that *G6PC2* regulates FPG by modulating the set point for glucose-stimulated insulin secretion in pancreatic β cells.

Recent innovations in genotyping technology have led to the identification of genetic variants associated with increased risk of type 2 diabetes (T2D) (*1*–*4*). However, the genetic factors contributing to interindividual variation in blood glucose levels in the general population are largely unknown. In addition to affecting metabolic health, these variants may also affect coronary heart disease (CHD) risk, for which a linear correlation between FPG and CHD mortality has been described for both diabetic and nondiabetic individuals (*5*, *6*). Here, we describe the results of a genome-wide association (GWA) study in which we aimed to identify genetic variants involved in glucose homeostasis in the general population.

We analyzed FPG as a quantitative trait in 654 normoglycemic (NG) nonobese individuals by using Illumina Infinium Human1 (Illumina Incorporated, San Diego, CA) and Hap300 BeadArray (Illumina Incorporated, San Diego, CA) (table S1) (*1*, *7*). The strongest association signal was observed with SNP rs560887 [$P = 4 \times 10^{-7}$ adjusted for age, gender, and body mass index (BMI) under the additive model]. This SNP is part of a 17-kb linkage disequilibrium (LD) block on chromosome 2 that encompasses the gene and the 3′ flanking region of the islet-specific glucose-6-phosphatase–related protein [IGRP, also known by the gene name *G6PC2* (glucose-6-phosphatase catalytic unit 2)], a glycoprotein embedded in the endoplasmic reticulum (ER) membrane (*8*). Six SNPs within this LD block were genotyped in our GWA analysis. These included a nonsynonymous SNP (rs492594-L219V), which showed modest association with FPG [$P = 0.04$ for rs492594-L219V, which is in low LD with rs560887 at $r^2 = 0.24$ in the Centre d'Etude du Polymorphisme Humaín (CEPH) families with Northern and Western European ancestry (CEU), according to HapMap data], and a SNP located in the 3′ flanking region of *G6PC2* (rs563694), which showed stronger association ($P = 2 \times 10^{-5}$, $r^2 = 0.76$ with rs560887) (table S2). We used linear regression analysis to adjust for the effect of rs560887 on all SNPs located within 100 kbp upstream and 100 kbp downstream that showed various levels of association with FPG ($P < 0.05$) and found no SNP that remained associated at significance level of 5% (table S2). Imputation of genotypes using the LD structure from HapMap data (CEU population) did not identify any SNP showing potentially stronger association with FPG (Fig. 1). On the basis of these findings, we tested the association of rs560887 with FPG by using a complementary subset (table S1) of the French DESIR (data from the Epidemiological Study on the Insulin Resistance Syndrome) population (DESIR NG stage 2, $N = 3419$) and replicated our original association with FPG (linear regression coefficient $\beta = -0.06$ mmol/l per A allele, $P = 3 \times 10^{-8}$) (Table 1).

IGRP has been proposed to modulate the glycolytic pathway and eventually glucose-stimulated insulin secretion by dephosphorylating glucose-6-phosphate generated by glucokinase, the β cell glucose sensor (*9*). *G6pc2* knockout mice display decreased FPG and normal insulin sensitivity (*10*). In the DESIR NG cohort, the rs560887 A allele [minor allele frequency (MAF) = 0.30] associated with decreased FPG is also associated with a decreased fraction of erythrocyte glyco-

[1]CNRS UMR 8090 Institute of Biology, Pasteur Institute of Lille and Lille 2 Droit et Santé University, 59019 Lille, France. [2]Department of Human Genetics, Faculty of Medicine, McGill University and Génome Québec Innovation Centre, Montreal H3A 1A4, Canada. [3]Gene Expression Unit, Department of Molecular Cell Biology, Katholieke Universiteit, B-3000 Leuven, Belgium. [4]Department of Obstetrics and Gynaecology, 90014 Oulu, Finland. [5]Department of Epidemiology and Public Health, Imperial College, London W2 1PG, UK. [6]Institut Inter Régional pour la Santé (IRSA), 37520 La Riche, France. [7]INSERM U695, Bichat Hospital, 75722 Paris, France. [8]INSERM U780-IFR69, Paris Sud University, 94807 Villejuif, France. [9]Pediatric Endocrine Unit, Jeanne de Flandre Hospital, 59037 Lille, France. [10]Department of Public Health and General Practice, University of Oulu, 90014 Oulu, Finland. [11]Department of Pediatrics, Faculty of Medicine, McGill University, Montreal H3H 1P3, Canada. [12]Genomic Medicine, Hammersmith Hospital, Imperial College London, London W12 0NN, UK.

*These authors equally contributed to this work.
†To whom correspondence should be addressed. E-mail: p.froguel@imperial.ac.uk