2014

Local Genomic Adaptation of Coral Reef-Associated Microbiomes to Gradients of Natural Variability and Anthropogenic Stressors

Linda W. Kelly
San Diego State University

Gareth J. Williams
Scripps Institution of Oceanography

Katie Barott
University of Pennsylvania, kbarott@sas.upenn.edu

Craig A. Carlson
University of California, Santa Barbara

Elizabeth A. Dinsdale
San Diego State University

See next page for additional authors

Recommended Citation

Kelly, L. W., Williams, G. J., Barott, K., Carlson, C. A., Dinsdale, E. A., Edwards, R. A., Haas, A. F., Haynes, M., Lim, Y. W., McDole, T., Nelson, C. E., Sala, E., Sandin, S. A., Smith, J. E., Vermeij, M. J., Youle, M., & Rohwer, F. (2014). Local Genomic Adaptation of Coral Reef-Associated Microbiomes to Gradients of Natural Variability and Anthropogenic Stressors. *PNAS (Proceedings of the National Academy of Sciences), 111* (28), 10227-10232. http://dx.doi.org/10.1073/pnas.1403319111

At the time of this publication, Dr. Barott was affiliated with San Diego State University and the Scripps Institution of Oceanography, but she is now a faculty member at the University of Pennsylvania.

Local Genomic Adaptation of Coral Reef-Associated Microbiomes to Gradients of Natural Variability and Anthropogenic Stressors

Abstract
Holobionts are species-specific associations between macro- and microorganisms. On coral reefs, the benthic coverage of coral and algal holobionts varies due to natural and anthropogenic forcings. Different benthic macroorganisms are predicted to have specific microbiomes. In contrast, local environmental factors are predicted to select for specific metabolic pathways in microbes. To reconcile these two predictions, we hypothesized that adaptation of microbiomes to local conditions is facilitated by the horizontal transger of genes responsible for specific metabolic capabilities. To test this hypothesis, microbial metagenomes were sequenced from 22 coral reefs at 11 Line Islands in the central Pacific that together span a wide range of biogeochemical and anthropogenic influences. Consistent with our hypothesis, the percent cover of major benthic functional groups significantly correlated with particular microbial taxa. Reefs with higher coral cover had a coral microbiome with higher abundances of Alphaproteobacteria (such as Rhodobacterales and Sphingomonadales), whereas microbiomes of algae-dominated reefs had higher abundances of Gammaproteobacteria (such as Alteromonadales, Pseudomonadales, and Vibrionales), Betaproteobacteria, and Bacteriodetes. In contrast to taxa, geography was the strongest predictor of microbial community metabolism. Microbial communities on reefs with higher nutrient availability (e.g., equatorial upwelling zones) were enriched in genes involved in nutrient-related metabolisms (e.g., nitrate and nitrite ammonification, Ton/Tol transport, etc.). On reefs further from the equator, microbes had more genes encoding chlorophyll biosynthesis and photosystems I/II. These results support the hypothesis that core microbiomes are determined by holobiont macroorganisms, and that those core taxa adapt to local conditions by selecting for advantageous metabolic genes.

Keywords
microbial biogeography, marine bacteria, metabolic potential

Disciplines
Biology | Ecology and Evolutionary Biology | Marine Biology | Microbiology

Comments
At the time of this publication, Dr. Barott was affiliated with San Diego State University and the Scripps Institution of Oceanography, but she is now a faculty member at the University of Pennsylvania.

Author(s)
Linda W. Kelly, Gareth J. Williams, Katie Barott, Craig A. Carlson, Elizabeth A. Dinsdale, Robert A. Edwards, Andreas F. Haas, Matthew Haynes, Yan W. Lim, Tracey McDole, Craig E. Nelson, Enric Sala, Stuart A. Sandin, Jennifer E. Smith, Mark J.A Vermeij, Merry Youle, and Forest Rohwer

Local genomic adaptation of coral reef-associated microbiomes to gradients of natural variability and anthropogenic stressors

Linda W. Kelly[a,1], Gareth J. Williams[b], Katie L. Barott[a,b], Craig A. Carlson[c], Elizabeth A. Dinsdale[a], Robert A. Edwards[d], Andreas F. Haas[b], Matthew Haynes[a], Yan Wei Lim[a], Tracey McDole[a], Craig E. Nelson[e], Enric Sala[f], Stuart A. Sandin[b], Jennifer E. Smith[b], Mark J. A. Vermeij[g,h], Merry Youle[i], and Forest Rohwer[a]

[a]Department of Biology, San Diego State University, San Diego, CA 92182; [b]Marine Biology Research Division, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093; [c]Marine Science Institute, Department of Ecology, Evolution, and Marine Biology, University of California, Santa Barbara, CA 93106; [d]Department of Computer Sciences, San Diego State University, San Diego, CA; [e]Center for Microbial Oceanography: Research and Education, Department of Oceanography, University of Hawai'i, Honolulu, HI 96822; [f]National Geographic Society, Washington, DC 20036; [g]Caribbean Research and Management of Biodiversity, Willemstad, Curacao; [h]Aquatic Microbiology, Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, 1090 GE, Amsterdam, The Netherlands; and [i]Rainbow Rock, Ocean View, HI 96737

Edited by Edward F. DeLong, Massachusetts Institute of Technology, Cambridge, MA, and approved June 10, 2014 (received for review February 24, 2014)

Holobionts are species-specific associations between macro- and microorganisms. On coral reefs, the benthic coverage of coral and algal holobionts varies due to natural and anthropogenic forcings. Different benthic macroorganisms are predicted to have specific microbiomes. In contrast, local environmental factors are predicted to select for specific metabolic pathways in microbes. To reconcile these two predictions, we hypothesized that adaptation of microbiomes to local conditions is facilitated by the horizontal transfer of genes responsible for specific metabolic capabilities. To test this hypothesis, microbial metagenomes were sequenced from 22 coral reefs at 11 Line Islands in the central Pacific that together span a wide range of biogeochemical and anthropogenic influences. Consistent with our hypothesis, the percent cover of major benthic functional groups significantly correlated with particular microbial taxa. Reefs with higher coral cover had a coral microbiome with higher abundances of Alphaproteobacteria (such as Rhodobacterales and Sphingomonadales), whereas microbiomes of algae-dominated reefs had higher abundances of Gammaproteobacteria (such as Alteromonadales, Pseudomonadales, and Vibrionales), Betaproteobacteria, and Bacteriodetes. In contrast to taxa, geography was the strongest predictor of microbial community metabolism. Microbial communities on reefs with higher nutrient availability (e.g., equatorial upwelling zones) were enriched in genes involved in nutrient-related metabolisms (e.g., nitrate and nitrite ammonification, Ton/Tol transport, etc.). On reefs further from the equator, microbes had more genes encoding chlorophyll biosynthesis and photosystems I/II. These results support the hypothesis that core microbiomes are determined by holobiont macroorganisms, and that those core taxa adapt to local conditions by selecting for advantageous metabolic genes.

microbial biogeography | marine bacteria | metabolic potential

Coral reefs are complex ecosystems that provide habitats for diverse, interdependent macro- and microorganisms. A coral colony itself is a complex holobiont, each made up of a coral polyp and a suite of prokaryotic microbes, viruses, protists, endolithic fungi and algae, and other invertebrates (1–4). Some coral-associated microbes confer benefits by, for example, remineralizing nutrients that are essential for the coral holobiont (5–9). Others contribute to coral demise by causing a number of specific diseases as well as nonspecific detrimental effects (e.g., hypoxia) (10–12). On degraded reefs, where coral cover is reduced and the benthic surface is dominated by fleshy algae, the microbial community includes higher abundances of copiotrophic microbes, many of which are known pathogens (13). Higher abundances of potential pathogens on reefs are also known to correlate with higher prevalence of coral disease (14), indicating a link between the community structure of reef-associated microbes and coral health.

Previous studies have described the biogeographic distribution of pelagic microbial communities by investigating statistical relationships between pelagic microbes and environmental parameters (15–18). However, application of this approach to coral reef-associated microbes is complicated by a number of factors. First, for microbial members of specific coral holobionts, microbial biogeography is directly linked to the distribution of the coral species. Second, reef-associated microbial communities are influenced by the other benthic macroorganisms present, such as macroalgae—both calcifying and fleshy, which may vary markedly between locations. Third, these microbial communities are subject to abiotic factors— such as variable nutrient, temperature, and hydrodynamic regimes—associated with a particular geographic location. Given this complexity, understanding the drivers that influence the community structure of reef-associated microbes requires unraveling numerous interdependent factors.

The relationships between microbial community structure, the metabolic capacity of the assemblage, and their habitat are

Significance

Microbial communities associated with coral reefs influence the health and sustenance of keystone benthic organisms (e.g., coral holobionts). The present study investigated the community structure and metabolic potential of microbes inhabiting coral reefs located across an extensive area in the central Pacific. We found that the taxa present correlated strongly with the percent coverage of corals and algae, while community metabolic potential correlated best with geographic location. These findings are inconsistent with prevailing biogeographic models of microbial diversity (e.g., distance decay) and metabolic potential (i.e., similar functional profiles regardless of phylogenetic variability). Based on these findings, we propose that the primary carbon sources determine community structure and that local biogeochemistry determines finer-scale metabolic function.

Author contributions: F.R. designed research; L.W.K., K.L.B., C.A.C., E.A.D., R.A.E., A.F.H., M.H., Y.W.L., T.M., C.E.N., E.S., S.A.S., J.E.S., and M.J.A.V. performed research; L.W.K. and G.J.W. analyzed data; and L.W.K., M.Y., and F.R. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequences reported in this paper have been deposited in the MG-RAST Metagenomics Analysis Server, http://metagenomics.anl.gov/linkin.cgi?project=9220 (project name: Pacific Reef Microbiomes).

[1]To whom correspondence should be addressed. Email: lwegley@gmail.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1403319111/-/DCSupplemental.

complex. Numerous taxa share core genes required for survival in the marine habitat. Supplementing these core housekeeping genes in each strain are a varied combination of metabolic genes (the pan-genome) associated with specialized pathways that contribute to fitness under particular local conditions, e.g., limited phosphate availability. These specialization genes do not respect species boundaries and may be found in multiple taxa adapted to similar environmental conditions (19, 20). Due to the mobility of these genes via horizontal gene transfer, the microbes can be considered to share a common gene pool, with specific genes being enriched within communities in particular niche habitats where they increase fitness. As a result, the similar community metabolism (i.e., functional redundancy) can be associated with high phylogenetic variability (21), and likewise communities comprised of similar taxa may differ in metabolic capabilities (22).

The mechanisms that govern community structure and gene flow in complex microbial communities, such as those associated with benthic marine habitats, remain largely unknown to the field of microbial ecology. Coral reefs are of particular interest because of their importance as centers of biodiversity, their contribution to global marine productivity, and their alarming decline. Coral reefs of the Line Islands (LIs) in the central Pacific offer a unique opportunity to investigate these questions as they span a latitudinal gradient from 6° north to 11° south. These islands and atolls (hence forth referred to as atolls) also span across the Equatorial Counter Current and Intertropical Convergence Zone, and thus experience significant variability in nutrient concentrations, temperature and precipitation.

In addition to oceanographic variability, the northern LIs also span a gradient of human disturbance where Teraina, Tabuaeran, and Kiritimati support populations of ~1,000, 2,500, and 5,000 people, respectively. Reefs at these atolls are impacted by subsistence and commercial fishing, as well as some pollution (e.g., sewage, chemicals) and agricultural runoff. Some of the highest known biomass of the fishes for a coral reef ecosystem were observed on the unpopulated atolls (14, 23), where reefs were characterized by the high cover of reef-building corals and crustose coralline algae, abundant coral recruits, and low levels of coral disease (14). In contrast, the populated atolls, most notably Kiritimati, had reefs with as low as 2% coral cover and were associated with a higher abundance of super heterotrophs, many of which are known pathogens (13), and a higher prevalence of coral diseases (14). Because the reefs at the uninhabited atolls have been largely spared from such anthropogenic disturbances, they provide a baseline for a comparative evaluation of the effects of human activity on coral reef-associated microbes. However, to definitively attribute any observed differences to anthropogenic activities, the role of other environmental drivers that differ between atolls must also be examined. For instance, the three inhabited atolls are clustered together in a region spanning <3° latitude, inciting a counterargument that local biogeochemical factors were responsible for reef degradation rather than fishing or other local activities as had been suggested by a prior study (14).

Here we used comparative metagenomics to tease out the key environmental factors driving the composition and metabolism of reef-associated microbial communities in the LIs. Although the 11 atolls are clustered in the same oceanic region, they differ in three key environmental variables that are predicted to influence their microbial communities: nutrient levels, latitudinal distance from the equator, and the percentage of benthic surface occupied by various functional groups of macroorganisms. In this study, we collected reef-associated microbes, then extracted and sequenced the community DNA. Taxonomic and functional annotations were assigned to the resultant reads by comparison with the SEED protein database. We then quantified variation in the structure and metabolic potential of the communities in relation to the three key variables. These comparisons show that (*i*) the microbial taxa present and their relative abundances reflect the benthic community whose carbon-containing exudates provide the primary local energy source, and (*ii*) the presence of various specialized metabolic capabilities correlates with nutrient levels and other latitude-dependent factors.

Results

Studies were conducted at 22 reef sites distributed across 11 LIs spanning 18° latitude (Table S1). At each site, seawater samples were collected at the surface of the benthos for microbial metagenome preparation and from the immediately overlying water for nutrient analysis. The macroorganisms comprising the benthic cover were surveyed. Subsequent analyses assessed the relationships between three predictor variables (benthic macroorganisms, nutrient levels, and latitude) and both the structure and the metabolic capabilities of the microbial communities at these atolls.

Nutrient Concentration. Inorganic nitrogen (nitrate + nitrite) and phosphate concentrations were generally highest near the equator and declined with increasing latitude both north and south (Fig. 1 and Table S2). Nitrate + nitrite concentrations ranged from 0.52 to 4.83 μM, whereas phosphate concentrations varied less (0.15–0.44 μM). Compared with the northernmost (Kingman) and southernmost (Flint) atolls, nitrate + nitrite and phosphate concentrations at equatorial Jarvis were approximately five- and twofold higher, respectively.

Benthic Macroorganisms. The benthic cover was quantified as the percentage covered by each of seven functional groups: hard coral, crustose coralline algae, calcified macroalgae, soft coral, fleshy macroalgae, fleshy turf algae, and "other" (Table S2). A list of the genera within each category is also provided (Table S3). Coral cover varied markedly from 2.2% at one site on Kiritimati to 86.7% at one site on Malden (mean = 44.4%; Table S2). In general, the uninhabited atolls were dominated by reef-building calcifiers including coral, crustose coralline algae, and calcified macroalgae (24), whereas fleshy algae, such as turf and fleshy macroalgae, dominate the inhabited atolls (14).

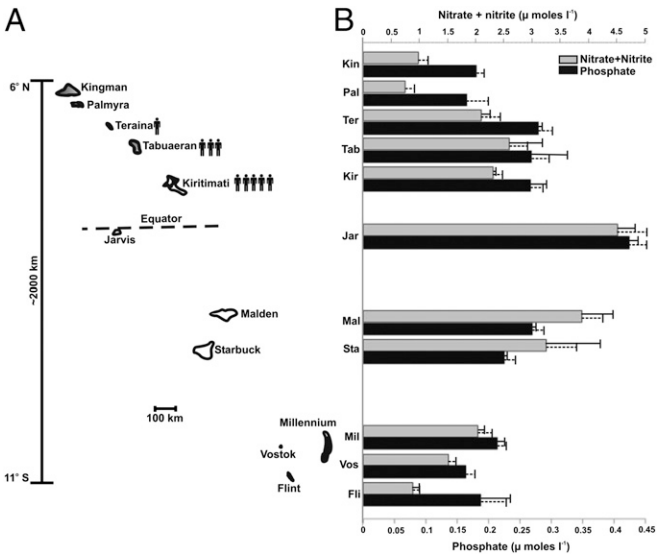

Fig. 1. The LIs and their nutrient concentrations. (*A*) The 11 main atolls sampled in this study. The scale on the left indicates latitude and distance between atolls. Atoll sizes are proportionate, but not to scale. (*B*) Average nutrient concentrations at the 11 atolls. Nutrient concentrations were measured in triplicate for each of the 22 study sites ($n = 66$) and averaged; sites were then averaged for each atoll. Solid and dashed error bars show the SE for atoll and site replicates, respectively. Average values for each site are provided in Table S2.

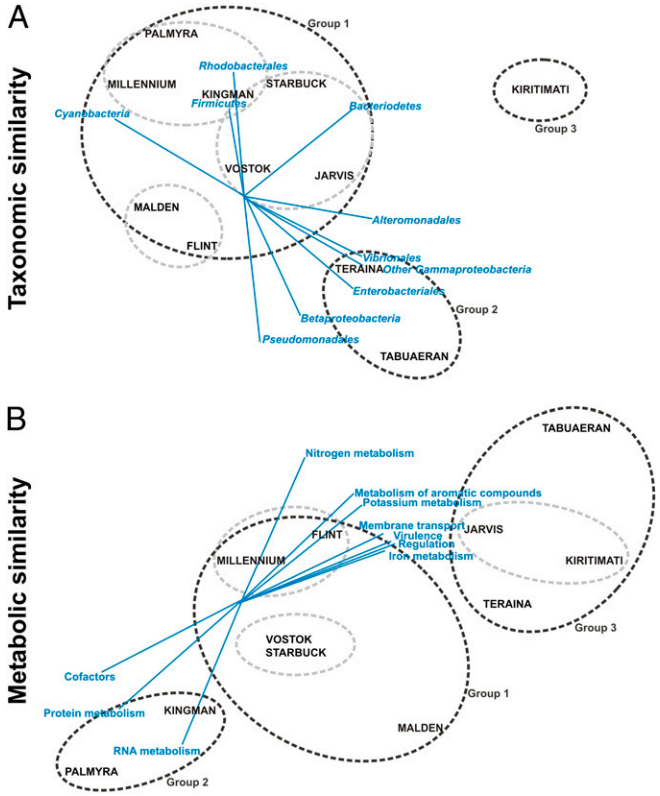

Fig. 2. nMDS plots for the relative abundances of taxonomic similarities (*A*) and metabolic subsystem similarities (*B*). Sites were averaged for each atoll. The 2D stress values are 0.05 and 0.03 for the taxonomic and metabolic similarities, respectively. Dark gray circles indicate significant groupings from the SIMPROF analysis (Figs. S2 and S3; Bray–Curtis similarity, $P < 0.01$). Light gray circles cluster atolls with greatest similarity within each statistically significant group.

Reef-Associated Microbes. DNA isolated from microbes sampled at each site was sequenced to yield 22 metagenomic libraries totaling 2.25 million quality reads (average length 389 bp; Table S1). The sequenced reads were translated in silico into predicted protein sequences; subsequent comparison with the SEED database provided taxonomic annotations for 21–47% of the reads and assignments to functional subsystems for 27–62% of the reads from each site. These annotations were the basis for comparative analyses of the microbial community structure and metabolic capabilities across the LI archipelago.

The relative abundances of the major taxonomic groups were tabulated (Fig. S1), plotted in 2D using nonmetric multidimensional scaling (nMDS; Fig. 2*A*), and analyzed for multivariate structure using similarity profile (SIMPROF) analysis (Fig. S2). By all measures, the geographic location of the atoll was a poor predictor of similarity of microbial community structure. For example, the two northernmost atolls, Kingman and Palmyra, are clustered with the Southern LIs in group 1 (Fig. 2*A*) and were most similar to Millennium, one of the southernmost atolls. Likewise Malden and Flint, separated by nearly 900 km, had similar taxonomic composition. In contrast, the metabolic capabilities (based on level 1 subsystem designations in the SEED; $n = 20$) of microbial communities in geographic proximity were more similar, forming three groups corresponding to the northern, middle, and southern atolls (Fig. 2*B* and Fig. S3). SIMPROF analyses conducted at the site level resulted in a higher number of significant groupings, although each site generally remained located within its own atoll group (Figs. S2 and S3) provided some exceptions, particularly in the metabolic groupings (e.g., Flint 2 clustered with group 3 atolls, Fig. S3). Further analyses were performed to quantify correlations between three key variables and both microbial community structure and metabolism across the LIs.

Community Structure. The correlations visualized by canonical correspondence analysis (CCA) (Fig. 3*A*) illustrate that microbial community structure on LI reefs is closely associated with benthic community composition. Reefs at all of the uninhabited LIs (group 1 in Fig. 2*A* and Fig. S2) associated with a higher percent cover of reef-building calcifiers were characterized by higher abundances of Cyanobacteria, Alphaproteobacteria (i.e., orders Rhodobacterales and Rickettsiales), and Firmicutes. Reefs with the highest hard coral coverage, such as Malden and Flint, had higher abundances of Sphingomonadales and Cyanobacteria (Fig. 3*A*). Although the abundance of the genus *Synechococcus* correlates positively with nutrient concentration in pelagic microbial communities, here it was positively correlated with the percentage of hard coral cover (Table 1; $r = 0.665$, $P = 0.026$). In contrast, hard coral cover showed a strong negative correlation with the abundance of Alteromonadales ($r = -0.819$, $P = 0.002$).

The inhabited group 2 atolls associated with higher percent cover of fleshy macroalgae (Tabuaeran and Teraina; Fig. 3*A*) had greater abundances of Gammaproteobacteria (e.g., orders Enterobacteriales and Pseudomonadales) and Betaproteobacteria. In contrast, the reefs at populated Kiritimati were dominated by fleshy turf algae (58.9–82.4%) and supported a markedly

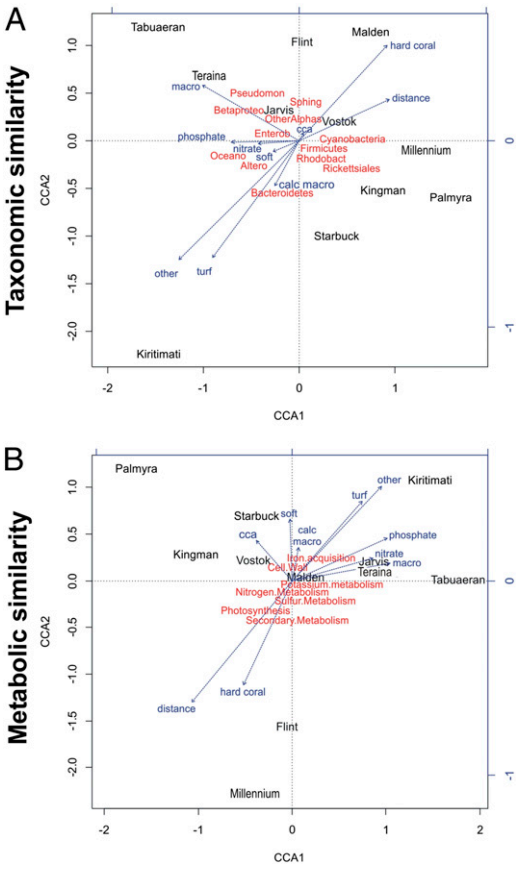

Fig. 3. CCA depicting the correlations between predictor variables (blue) and the relative abundance of taxonomic similarities (*A*) and metabolic similarities (*B*) at each LI. Loading vectors for the taxa and subsystems are shown in red. Altero, Alteromonadales; Betaproteo, Betaproteobacteria; cca, crustose coralline algae; calc macro, calcified macroalgae; dist, distance from the equator in degrees latitude; Enterob, Enterobacteriales; macro, fleshy macroalgae; Oceano, Oceanospirillales; OtherAlphas, other Alphaproteobacteria; Pseudomon, Pseudomonadales; Rhodobact, Rhodobacterales; soft, soft coral; Sphing, Sphingomonadales.

increased abundance of Bacteriodetes (25.1 ± 4.2%, $n = 2$) compared with the other atolls (7.2 ± 3.5%, $n = 20$). Specifically, five genera within the class Flavobacteria (genera *Croceibacter*, *Dokdonia*, *Gramella*, *Leeuwenhoekiella*, and *Polaribacter*) were consistently overrepresented compared with sites on other atolls. Overall, the percent coverage of fleshy turf algae on LI reefs was positively correlated with bacteria from the orders Flavobacteriales and Alteromonadales (Table 1; $r = 0.815$, $P = 0.002$ and $r = 0.682$, $P = 0.021$, respectively). The CCA also depicted a correlation between the percent cover of other benthic organisms and Kiritimati reefs. Although other benthic organisms contributed to <1% of the benthic composition on most LI reefs, the two sites on Kiritimati had a higher percentage of sand, which contributed to the higher percent cover of this category (5.2 ± 0.5%).

A distance-based linear model (DistLM) was used to formally quantify which suite of predictor variables formed the best-fit model (balancing performance with complexity) for explaining variations in microbial communities across LI reefs. Hard coral alone had the largest impact on microbial community structure, explaining 15.2% of the variation between reefs (Table S4).

Community Metabolism. Distance from the equator was the strongest predictor of community metabolism, explaining 18.4% of the variation in microbial metabolic potential (Table S4). The two northern atolls (group 2 in Fig. 2*B*, Kingman and Palmyra) were characterized by high abundances of genes encoding cofactors, RNA metabolism, and protein metabolism. Moving southward, the midlatitude atolls (group 3 in Fig. 2*B*; Jarvis, Kiritimati, Teraina, and Tabuaeran) were characterized by higher abundances of genes for aromatic compound utilization, iron metabolism, membrane transport, nitrogen metabolism, potassium metabolism, regulation, and virulence. All of the southern LIs were combined into one group and had similar community metabolism (group 1; Fig. 2*B*).

The question remained as to which environmental parameters associated with latitude were driving these variations. Nutrient levels varied across the LIs as expected due to the influence of equatorial upwelling (Fig. 1). As such, a number of metabolic pathways (SEED level 3 subsystems) demonstrated significant correlations with local phosphate concentrations across all 11 atolls. These included six pathways positively correlated with phosphate concentration: conjugative transfer, chemotaxis, nitrate and nitrite ammonification, cobalt–zinc–cadmium resistance, multidrug resistance efflux pumps, and Ton and Tol transport (Fig. 4*A* and Table S5). Phosphate concentration was negatively correlated with two metabolic pathways involved in photosynthesis (chlorophyll biosynthesis and photosystems I and II) (Fig. 4*B* and Table S5), and also with the abundance of *Prochlorococcus* (Table 1). Genes for ribosomal proteins were also overrepresented at oligotrophic sites (Fig. 4*B*).

Interisland Comparison. Atolls in close proximity were observed to have similar metabolic capabilities despite differences in their taxonomic composition. For example, microbial communities from the geographically close Jarvis and Kiritimati had similar metabolic profiles (Fig. 2), but the taxonomic profile of Jarvis was most similar to Vostok and Starbuck, whereas that for Kiritimati was the most dissimilar of all (Fig. 2). Conversely, the distant atolls of Kingman and Malden supported taxonomically similar microbial communities that encoded divergent metabolic capabilities. Hence, microbial communities composed of different taxa can encode similar functions, and vice versa.

Table 1. Significance test for linear correlations between taxon abundance and specific predictor variables

Taxon	Predictor variable	*r*	*P*
Flavobacteriales	Turf algae	0.815	0.002
Alteromonadales	Turf algae	0.682	0.021
Alteromonadales	Hard coral	−0.819	0.002
Synechecoccus	Hard coral	0.665	0.026
Gammaproteobacteria	Macroalgae	0.560	0.073
Prochlorococcus	Phosphate	−0.614	0.045
Sphinogomonadales	Nitrate	0.758	0.007
Erythrobacter	Nitrate	0.674	0.023

r, Pearson's coefficient.

Discussion

This study reports, to our knowledge, the first large-scale metagenomic survey of microbial communities associated with coral reefs that simultaneously characterizes both taxonomic composition and metabolic capabilities. We have demonstrated that, at the ecosystem level, benthic macroorganisms most strongly influence the taxonomic composition of the microbial community, whereas metabolic specialization genes carried by these taxa vary between locations and reflect functional adaptations to local oceanographic conditions.

For this study, microbial communities were sampled from 22 coral reef sites at 11 atolls across the LI archipelago; atolls that differed with respect to their benthic community, nutrient levels, and latitude. The microbes collected by our procedure were closely associated with the surface of the benthic macroorganisms (corals and algae). As a result, they included species-specific bacterial components of the coral holobiont (1), as well as specific bacterial taxa associated with some algal functional groups (1, 25). In addition, the microbial communities sampled on these reefs reflected selection by the adjacent benthic macroorganisms, as evidenced by the differences between reef-associated bacterioplankton communities and open ocean communities (26). There is evidence that reef-associated communities undergo selection in shallow reef environments by the locally available labile organic matter exuded by the benthic organisms (27). For example, in an empirical study, Nelson et al. demonstrated that exudates collected from coral and macroalgae selectively fostered growth of distinct bacterioplankton communities (27). Coral exudates promoted communities with higher diversity, including lineages of Alphaproteobacteria with relatively few virulence factors (e.g., Erythrobacteraceae*)*; whereas exudates from fleshy macroalgae selected for less diverse communities with more copiotrophic Gammaproteobacteria lineages (e.g., the families Alteromonadaceae, Pseudoalteromonadaceae, and Vibrionaceae).

Community Structure. The current study confirms and extends earlier findings (27) by demonstrating similar correlations between benthic community composition and the enrichment of specific microbial taxa on coral reefs in situ (Table 1). Consistent with the effects of individual exudates, high coral cover was associated with higher abundances of Alphaproteobacteria, whereas the abundant fleshy macroalgae at Tabuaeran and Teraina were accompanied by more Gammaproteobacteria (e.g., Enterobacteriales and Pseudomonadales). Together, these complementary research approaches indicate that coral- and algae-derived organic exudates enrich for specific types of bacteria living in close association with coral reefs.

Nutrient levels have also been postulated to influence microbial community composition. Here we tested this hypothesis using the natural nutrient gradient present across the LIs. Due to the equatorial Pacific upwelling in this region, phosphate and nitrate are elevated at the equator and decrease with latitude both north and south (Fig. 1). In high-nitrate, low-chlorophyll ecosystems such as this, iron may be the nutrient limiting primary production (28). Other unspecified biogeographic factors also vary with latitude across the LIs. In this study, neither nutrients nor other latitude-dependent variables were included in the best-fit model for determining microbial community structure. Therefore, we propose that on these geographically separate coral reefs, microbial community structure is determined by the available energy source, i.e., the dissolved organic carbon provided in the form of benthic exudates, which provides a mechanism for the correlations

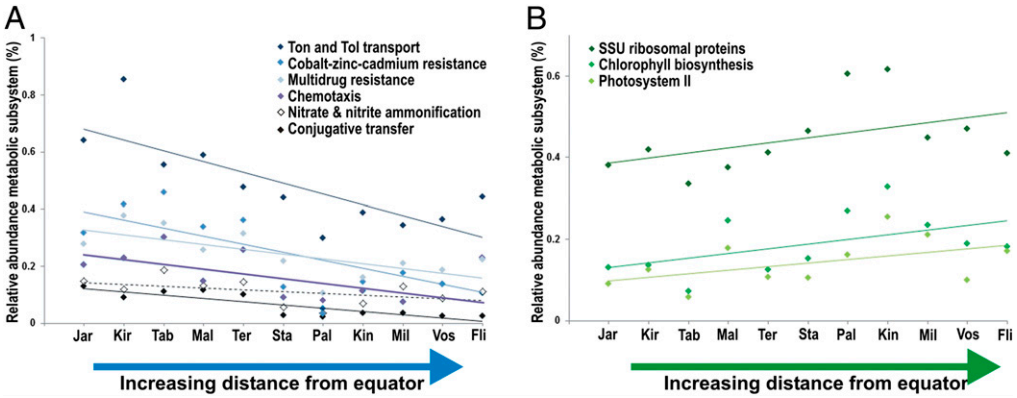

Fig. 4. Metabolic pathways that correlate positively (*A*) and negatively (*B*) with increasing distance from the equator (decreasing nutrient concentrations) across the LIs. Pathways are level 3 subsystem annotations from the SEED database. SSU, small subunit.

observed between the macro- and microbial components of reef communities.

Community Metabolism. In contrast to community structure, the specialized and ecologically relevant metabolic capabilities of these communities reflected local nutrient concentrations. For example, six level 3 metabolic subsystems (SEED database) correlated positively with phosphate concentration across the LIs (Fig. 4*A*). Some of these, such as the TonB system, contribute to nutrient acquisition. The TonB system transports large molecules (e.g., polysaccharides, proteins, and siderophores) in through the outer membrane of Gram-negative bacteria. Its importance in marine environments is evidenced by the presence of these genes in marine bacterial genomes and pelagic metagenomes (29–31), their high levels of expression in metatranscriptome data (32), and the proteomic identification of their products as the predominant membrane proteins in pelagic bacteria (33). In this study, they accounted for nearly 1% of gene function annotations at some high-nutrient sites (Fig. 4*A*). Genes of the conjugative transfer subsystem, also overrepresented at high-nutrient sites, may function in energy and nutrient acquisition via type IV secretion of ectoenzymes and siderophores, and may support active horizontal gene transfer via conjugation. Conversely, the more oligotrophic sites exhibited overrepresentation of two photosynthesis pathways (chlorophyll biosynthesis and photosystems I and II) (Fig. 4*B* and Table S5), as well as greater abundance of *Prochlorococcus,* a key primary producer in oligotrophic oceans (Table 1).

Previous studies have shown that the anaerobic ammonification of nitrate and nitrite (also referred to as dissimilatory nitrate reduction to ammonium or DNRA) is significant for nitrogen metabolism in the diffusive boundary layer, an environment with heterogeneous distribution of dissolved oxygen during the day (12) that then becomes anoxic at night (34). That anaerobes dominate coral-associated microbial communities suggested that this anaerobic nitrogen metabolism may be important on coral surfaces (25). An interesting observation from the nutrient measurements is that atolls with higher nitrate + nitrite availability have lower ammonium concentrations, whereas low nitrate + nitrite atolls have higher ammonium. Nitrate + nitrite to ammonium ratios were 0.26, 0.29, and 0.22 on Malden, Jarvis, and Kiritimati compared with 3.23 and 1.47 on Flint and Kingman, respectively (Table S2). Therefore, the overrepresentation of DNRA may reflect the lower abundances of ammonium at these high-nutrient sites.

Reef-associated microbial communities in high-nutrient environments encoded greater metabolic complexity, suggesting that they carry more specialization genes and thus generally possess larger genomes (Fig. S4). Consistent with this hypothesis, single-copy genes encoding ribosomal proteins were overrepresented at oligotrophic sites (Fig. 4*B*), indicating that the community overall possessed smaller genomes compared with those at high-nutrient sites.

Although both phosphorus and nitrogen concentrations correlated with distance from the equator ($r = -0.74$ and -0.64, respectively; Table S6), neither was as strong a predictor of metabolism, as was latitudinal distance from the equator (as assessed by DistLM analysis). Distance from the equator may serve as a proxy for other influential but unsampled variables, such as seawater temperature, salinity, photosynthetically active radiation, or micronutrient concentrations (e.g., iron). In addition, the limited sampling (one to four sites at each atoll) may have obscured significant correlations to specific nutrients. Had the atoll averages been based on sampling of 20+ sites per atoll, significant correlation with specific nutrients might have been discernible. Nevertheless, the availability of the macronutrients nitrate + nitrite and phosphate are posited to be important factors influencing microbial community metabolism on LI reefs.

Anthropogenic Impacts on LI Reefs. The findings of this study indicate that local human populations influence the reef-associated microbial community indirectly by influencing the composition of benthic macroorganisms. Typically activities such as fishing remove important grazing herbivore species resulting in increased cover of fleshy algae, and this in turn profoundly impacts microbial community structure at the populated atolls (Fig. 2 and Fig. S1). Increased coverage by fleshy algae selects for specific microbes that may be detrimental to coral health (27, 35), thereby opening additional benthic space for further algal colonization (36).

Discordance Between Taxa and Metabolism. Both the abundance of specific taxonomic groups and the community metabolic capabilities of the reef-associated microbial communities varied across the LIs. Both correlated with ecological factors, but did so independent of each other. As a result, atolls as far apart as Kingman and Malden (~1,400 km) hosted taxonomically similar communities, but these communities effectuated different metabolisms. Conversely, the different microbial communities at equatorial Jarvis and Kiritimati encoded similar metabolic specialization genes. This discordance between taxonomy and metabolism is intriguing. We hypothesize that although community structure is attributable to the core genes that classify each taxon, community metabolism reflects the particular complement of specialization genes that comprise the dynamic genome of each strain present. Previously, strain-specific adaptation to different nutrient levels had been documented in marine cyanobacteria for genes involved in phosphate acquisition. The particular genes present and their genomic organization depended on phosphate availability in each isolate's source environment. Strains of *Prochlorococcus* that showed 99.9% similarity of their 16S rRNA genes nevertheless possessed different phosphate metabolism genes located in different genomic locations (19). Conversely, some more divergent strains that occupied environments with similar nutrient regimes shared similar phosphate gene content and organization. Additionally, although *Prochlorococcus*

typically assimilates only ammonium, in regions of nitrogen limitation strains have adapted to use nitrate and nitrite by using genes acquired horizontally from *Synechococcus* (20).

The observed adaptation of microbial community metabolism patterns could have resulted from either gene acquisition and loss or shifts in the relative abundances of strains adapted to different conditions. Traditionally, only changes in strain abundance (i.e., beta diversity) have been considered as possible drivers of rapid adaptation in ecological time. Increased genetic diversity, i.e., evolution, by mechanisms such as horizontal movement of genes between strains or species, has been expected to require evolutionary time. We posit that in these microbial communities, evolution is rapid, occurring in ecological time.

Attempts to identify the evolutionary mechanisms active in this situation have been hampered by the limited representation of marine microbes in databases (37) such as SEED, due to our inability to culture most species (38). The availability of single-cell whole-genome amplification methods (39) promises to enable genomic characterization of unculturable marine microbes, thereby substantially accelerating resolution of this question.

Materials and Methods

Metagenomic sequence reads were compared with the SEED protein database (http://theseed.org/wiki/Main_Page) using BLASTx. For taxonomic annotation, sequences with significant similarities ($E < 10^{-5}$) were assigned to the closest identified microbial representative. For functional annotation, sequences were assigned the function of the closest identified protein and these functions were then grouped into metabolic pathways according to the subsystems in the SEED database. Community structure was compared using the relative abundances of 19 higher-rank microbial taxa (see *SI Materials and Methods* and Table S7 for clarification of taxonomic groups). Similarly, community metabolism was determined by comparing the relative abundance of 20 level 1 subsystem categories in the SEED database.

nMDS analyses were used with the annotated metagenome data to visualize between-atoll similarity in terms of two discrete response variables: community structure and community metabolism. For an initial exploration of potential correlations between the three predictor variables and either microbial community structure or metabolism, a CCA was performed using the R package, vegan. To formally quantify how much variation in the microbial communities or their metabolism could be explained by the predictors measured (continuous variables), a permutational DistLM was used in PERMANOVA+ (www.primer-e.com/permanova.htm).

Full methods and any associated references are available in the *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank Beltran Rodriguez-Mueller, Rob Schmieder, Bahador Nosrat, Federico Lauro, Nao Hisakawa, Jeremy Frank, Bas Dutilh, Katrine Whiteson, Barbara Bailey, and Jim Nulton for mathematical and bioinformatics support; and the Palmyra Atoll Research Consortium and the Nature Conservancy for field support. We are also grateful to Jennifer Martiny for valuable discussions and Heather Maughan for her editing expertise. The microbial samples were collected during two research expeditions to the LIs funded by the National Geographic Society, the Moore Family Foundation, the Hawaii Undersea Research Laboratory of the Coral Reef Conservation Program [a program of the National Oceanic and Atmospheric Administration (NOAA)], and several private donors and during one Reef Assessment and Monitoring Program cruise to Jarvis supported and executed by NOAA-Coral Reef Ecosystem Division. This work was carried out under research permits from the Palmyra Atoll National Wildlife Refuge, US Fish and Wildlife Service, and the Environment and Conservation Division of the Republic of Kiribati. This research was sponsored by National Science Foundation Awards OCE-0927415 and DEB-1046413 (to F.R.), OCE-0927411 (to C.A.C.), and OCE-0417412 (to Moorea Coral Reef Long-Term Ecological Research); and a Canadian Institute for Advanced Research Integrated Microbial Biodiversity Program Fellowship 141679 (to F.R.).

1. Rohwer F, Seguritan V, Azam F, Knowlton N (2002) Diversity and distribution of coral-associated bacteria. *Mar Ecol Prog Ser* 243:1–10.
2. Bourne DG, Munn CB (2005) Diversity of bacteria associated with the coral *Pocillopora damicornis* from the Great Barrier Reef. *Environ Microbiol* 7(8):1162–1174.
3. Koren O, Rosenberg E (2006) Bacteria associated with mucus and tissues of the coral Oculina patagonica in summer and winter. *Appl Environ Microbiol* 72(8):5254–5259.
4. Sunagawa S, Woodley CM, Medina M (2010) Threatened corals provide underexplored microbial habitats. *PLoS ONE* 5(3):e9554.
5. Knowlton N, Rohwer F (2003) Multispecies microbial mutualisms on coral reefs: The host as a habitat. *Am Nat* 162(Suppl 4):S51–S62.
6. Lesser MP, et al. (2007) Nitrogen fixation by symbiotic cyanobacteria provides a source of nitrogen for the scleractinian coral Montastraea cavernosa. *Mar Ecol Prog Ser* 346: 143–152.
7. Fiore CL, Jarett JK, Olson ND, Lesser MP (2010) Nitrogen fixation and nitrogen transformations in marine symbioses. *Trends Microbiol* 18(10):455–463.
8. Wegley L, Edwards R, Rodriguez-Brito B, Liu H, Rohwer F (2007) Metagenomic analysis of the microbial community associated with the coral Porites astreoides. *Environ Microbiol* 9(11):2707–2719.
9. Raina JB, Tapiolas D, Willis BL, Bourne DG (2009) Coral-associated bacteria and their role in the biogeochemical cycling of sulfur. *Appl Environ Microbiol* 75(11):3492–3501.
10. Smith JE, et al. (2006) Indirect effects of algae on coral: Algae-mediated, microbe-induced coral mortality. *Ecol Lett* 9(7):835–845.
11. Barott K, et al. (2009) Hyperspectral and physiological analyses of coral-algal interactions. *PLoS ONE* 4(11):e8043.
12. Haas AF, et al. (2013) Visualization of oxygen distribution patterns caused by coral and algae. *PeerJ* 1:e106.
13. Dinsdale EA, et al. (2008) Microbial ecology of four coral atolls in the northern Line Islands. *PLoS ONE* 3(2):e1584.
14. Sandin SA, et al. (2008) Baselines and degradation of coral reefs in the northern Line Islands. *PLoS ONE* 3(2):e1548.
15. Hewson I, Paerl RW, Tripp HJ, Zehr JP, Karl DM (2009) Metagenomic potential of microbial assemblages in the surface waters of the central Pacific Ocean tracks variability in oceanic habitat. *Limnol Oceanogr* 54(6):1981–1994.
16. Fuhrman JA, et al. (2008) A latitudinal diversity gradient in planktonic marine bacteria. *Proc Natl Acad Sci USA* 105(22):7774–7778.
17. Pommier T, et al. (2007) Global patterns of diversity and community structure in marine bacterioplankton. *Mol Ecol* 16(4):867–880.
18. Martiny AC, Tai APK, Veneziano D, Primeau F, Chisholm SW (2009) Taxonomic resolution, ecotypes and the biogeography of Prochlorococcus. *Environ Microbiol* 11(4): 823–832.
19. Martiny AC, Coleman ML, Chisholm SW (2006) Phosphate acquisition genes in Prochlorococcus ecotypes: Evidence for genome-wide adaptation. *Proc Natl Acad Sci USA* 103(33):12552–12557.
20. Martiny AC, Kathuria S, Berube PM (2009) Widespread metabolic potential for nitrite and nitrate assimilation among Prochlorococcus ecotypes. *Proc Natl Acad Sci USA* 106(26):10787–10792.
21. Burke C, Steinberg P, Rusch D, Kjelleberg S, Thomas T (2011) Bacterial community assembly based on functional genes rather than species. *Proc Natl Acad Sci USA* 108(34):14288–14293.
22. Moore LR, Rocap G, Chisholm SW (1998) Physiology and molecular phylogeny of coexisting Prochlorococcus ecotypes. *Nature* 393(6684):464–467.
23. Williams ID, et al. (2011) Differences in reef fish assemblages between populated and remote reefs spanning multiple archipelagos across the central and western Pacific. *J Mar Biol* 2011:826234.
24. Williams GJ, et al. (2013) Benthic communities at two remote Pacific coral reefs: Effects of reef habitat, depth, and wave energy gradients on spatial patterns. *PeerJ* 1:e81.
25. Barott KL, et al. (2011) Microbial diversity associated with four functional groups of benthic reef algae and the reef-building coral Montastraea annularis. *Environ Microbiol* 13(5):1192–1204.
26. Nelson CE, Alldredge AL, McCliment EA, Amaral-Zettler LA, Carlson CA (2011) Depleted dissolved organic carbon and distinct bacterial communities in the water column of a rapid-flushing coral reef ecosystem. *ISME J* 5(8):1374–1387.
27. Nelson CE, et al. (2013) Coral and macroalgal exudates vary in neutral sugar composition and differentially enrich reef bacterioplankton lineages. *ISME J* 7(5):962–979.
28. Martin JH, et al. (1994) Testing the iron hypothesis in ecosystems of the equatorial Pacific-Ocean. *Nature* 371(6493):123–129.
29. Hopkinson BM, Barbeau KA (2012) Iron transporters in marine prokaryotic genomes and metagenomes. *Environ Microbiol* 14(1):114–128.
30. Tang K, Jiao N, Liu K, Zhang Y, Li S (2012) Distribution and functions of TonB-dependent transporters in marine bacteria and environments: Implications for dissolved organic matter utilization. *PLoS ONE* 7(7):e41204.
31. Dupont CL, et al. (2012) Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *ISME J* 6(6):1186–1199.
32. Ottesen EA, et al. (2011) Metatranscriptomic analysis of autonomously collected and preserved marine bacterioplankton. *ISME J* 5(12):1881–1895.
33. Morris RM, et al. (2010) Comparative metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy transduction. *ISME J* 4(5):673–685.
34. Shashar N, Cohen Y, Loya Y (1993) Extreme diel fluctuations of oxygen in diffusive boundary-layers surrounding stony corals. *Biol Bull* 185(3):455–461.
35. Kelly LW, et al. (2012) Black reefs: Iron-induced phase shifts on coral reefs. *ISME J* 6(3): 638–649.
36. Barott KL, Rohwer FL (2012) Unseen players shape benthic competition on coral reefs. *Trends Microbiol* 20(12):621–628.
37. Yooseph S, et al. (2010) Genomic and functional adaptation in surface ocean planktonic prokaryotes. *Nature* 468(7320):60–66.
38. Rappé MS, Giovannoni SJ (2003) The uncultured microbial majority. *Annu Rev Microbiol* 57:369–394.
39. Stepanauskas R, Sieracki ME (2007) Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. *Proc Natl Acad Sci USA* 104(21): 9052–9057.

Supporting Information

Kelly et al. 10.1073/pnas.1403319111

SI Materials and Methods

Sample Collection. Samples were collected at 22 sites distributed across the 11 major atolls in the Line Island archipelago, 1–4 sites per atoll. Seawater samples of ∼100 L were collected at the surface of representative benthos (including within crevices, when present) across ∼20 m^2 of reef using a modified bilge pump. Samples were collected directly into low-density polyethylene collapsible bags (19 L; Cole-Parmer) and transported to the research vessel within 2 h. Before sampling, all containers, bilge pumps, and tubing were washed once with 1% bleach and 0.1 M NaOH, three times with freshwater, and once with 100 kDa filtered seawater. Samples were filtered through 20 μm Nitex to remove large eukaryotes. The filtrate was concentrated to <500 mL using a 100-kDa tangential flow filter, which retained the unicellular eukaryotes, microbes, and virus-like particles. The microbial fraction was collected by passing this concentrated sample through 0.45-μm Sterivex filters (Millipore, Inc.) and the filters were then stored at −80 °C.

While sampling the reef surface (above), water was also collected for nutrient analysis directly above the same reef area, within 20 cm of the reef surface, using diver-deployable polycarbonate Niskin bottles. Water samples were filtered through 0.2-μm Nuclepore Track-Etched membrane filters (Whatman) into 20 mL high-density polyethylene scintillation vials with cone-shaped plastic lined lids (Fisher Scientific) and then stored at −20 °C. Inorganic nutrient (nitrate + nitrite, nitrite, and phosphate) concentrations were measured using a QuikChem 8000 flow injection analyzer (Lachat Instruments) at the Marine Science Institute Analytical Laboratory (University of California, Santa Barbara).

Characterization of the benthic community was completed using photoquadrats (1). Two 25-m transect lines were quantified per site and ten 0.72-m^2 quadrats were assessed along each transect line using digital underwater photographs. Images were analyzed using the program Photogrid 1.0, where 100 stratified random points were identified to determine benthic community composition at each site. All organisms were characterized to the finest level of resolution possible (genus level for corals and macroalgae and functional group for turfing and crustose coralline algae). All surveys took place at 10 m depth on the fore-reef habitat of each atoll.

DNA Extraction and Metagenomic Library Construction. DNA was extracted and purified using a column purification protocol (NucleoSpin Tissue; Macherey-Nagel), modified to complete the lysis steps in the Sterivex filters. Lysates were removed from the Sterivex filters using a 3-mL Luer-Loc syringe. The rest of the extraction procedure was performed according to the manufacturer's recommendations. Metagenomic libraries were prepared using a GS FLX Titanium Rapid Library Preparation Kit (Roche Applied Sciences) and pyrosequenced using a 454 GS-FLX platform at San Diego State University.

Sequence Library Quality Control and Bioinformatics Analyses. Metagenomic sequence reads were filtered for quality using the Preprocessing and Information of Sequences tool, PRINSEQ (2), uploaded to the MG-RAST server (http://metagenomics.nmpdr.org/metagenomics.cgi), and compared with the SEED protein database using BLASTx (3). For taxonomic annotation, sequences with significant similarities ($E < 10^{-5}$) were assigned to the closest identified microbial representative. For functional annotation, sequences were assigned the function of the closest identified protein and these functions were then grouped into metabolic pathways according to the subsystems in the SEED database (4). These sequences are publically available through the MG-RAST server under the project name Pacific Reef Microbiomes (http://metagenomics.anl.gov/linkin.cgi?project=9220).

Statistics. Nonmetric multidimensional scaling (nMDS) analyses were used with the annotated metagenome data to visualize between-atoll similarity in terms of two discrete response variables: community structure and community metabolism. Community structure was determined by comparing the relative abundances of 19 higher-rank microbial taxa (to limit the number of taxonomic categories to avoid type I errors associated with loss of statistical power in multiple comparisons; see Table S7 for clarification of taxonomic groups), averaged by atoll. Similarly, community metabolism was determined by comparing the relative abundance of 20 level 1 subsystem categories in the SEED database (http://theseed.org/wiki/Main_Page). Significant groupings of atolls depicted by the nMDS were quantified using a similarity profile test based on the Bray–Curtis algorithm ($P < 0.01$) (similarity profile analysis or SIMPROF) (5), using the clustsig package (6) for R (R Development Core Team). Analyses were based on 10,000 random permutations of the annotated metagenomic data. These significant groupings designated by SIMPROF were then superimposed upon the nMDS plots. Individual variables that might be responsible for driving group differences in multivariate space were investigated by calculating Spearman's rank correlations and those with strong correlations (in this study, ≥0.6) plotted as vectors in the nMDS plots.

For an initial exploration of potential correlations between the three predictor variables and either microbial community structure or metabolism, a canonical correspondence analysis (CCA) was performed using the R package, vegan (7). The results from this analysis were visualized by plotting the CCA loading vectors. To formally quantify how much variation in the microbial communities or their metabolism could be explained by the predictors measured (continuous variables), a permutational distance-based multivariate linear model (DistLM) (8) was used in PERMANOVA+ (www.primer-e.com/permanova.htm). To determine their suitability for use in a linear model, collinearity of the predictor variables was tested by calculating pairwise Pearson correlation coefficients. No two predictors exceeded a correlation of 0.75 (Table S6); therefore, all were included in the model. Model selection (balancing performance with parsimony) was based on Akaike's information criterion (AIC) (9) with a second-order bias correction applied (AICc) (10). Significance was determined by comparing the model results obtained with the original data structure to those obtained with 10,000 random permutations of the raw data. Statistical analyses were performed using R Version 2.15.1 (R Development Core Team, www.r-project.org) (11) unless otherwise stated.

The program Xipe (12) was used to determine lower level taxa and level 3 subsystem metabolic pathways that were significantly different ($P < 0.05$; 1,000 iterations) between metagenomic libraries sampled from all 11 atolls. Its bootstrapping technique allows comparison of thousands of gene categories between two metagenomic libraries with a designated confidence threshold (e.g., 95%). The Xipe findings were further tested for correlation with distance from the equator, nutrient concentration, and percentage cover of the seven benthic functional groups by calculating Pearson correlation coefficients (r) in SPSS (IBM Corporation).

1. Sandin SA, et al. (2008) Baselines and degradation of coral reefs in the northern Line Islands. *PLoS ONE* 3(2):e1548.
2. Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27(6):863–864.
3. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215(3):403–410.
4. Meyer F, et al. (2008) The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9:386.
5. Clarke KR, Somerfield PJ, Gorley RN (2008) Testing of null hypotheses in exploratory community analyses: Similarity profiles and biota-environment linkage. *J Exp Mar Biol Ecol* 366(1-2):56–69.
6. Whitaker D, Christman M (2010) clustsig: Significant cluster analysis. R package Version 1.0. Available at www.r-project.org. Accessed June 20, 2014.
7. Oksanen JF, et al. (2012) vegan: Community ecology package. R package Version 2.0-4. Available at http://cran.r-project.org/index.html. Accessed June 20, 2014.
8. McArdle BH, Anderson MJ (2001) Fitting multivariate models to community data: A comment on distance-based redundancy analysis. *Ecology* 82(1):290–297.
9. Akaike H (1973) Maximum likelihood identification of gaussian autoregressive moving average models. *Biometrika* 60(2):255–265.
10. Hurvich CM, Tsai CL (1989) Regression and time-series model selection in small samples. *Biometrika* 76(2):297–307.
11. Gentleman R, Ihaka R (1996) The R language. *Proceedings of 28th Symposium on the Interface of Computing Science and Statistics (Graph-Image-Vision)*, eds Billard L, Fisher NI (Interface Found, Sydney, Australia), pp 326–330.
12. Rodriguez-Brito B, Rohwer F, Edwards RA (2006) An application of statistics to comparative metagenomics. *BMC Bioinformatics* 7:162.

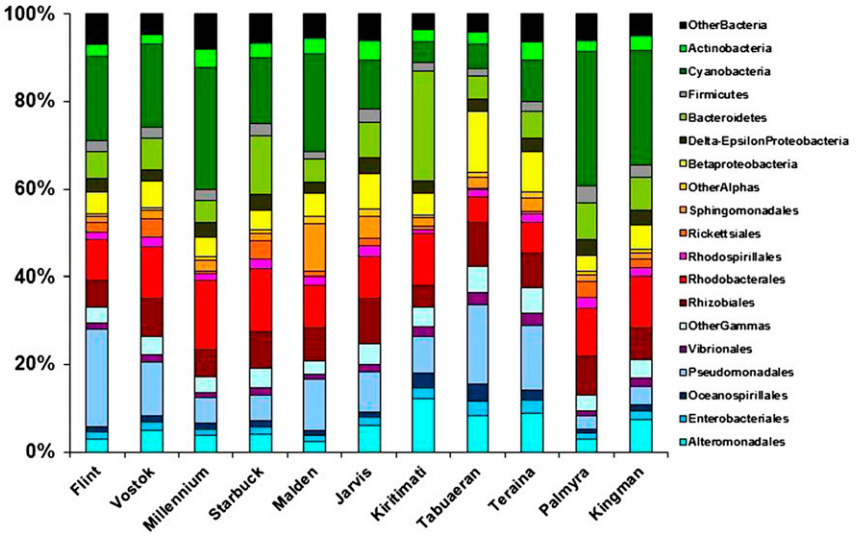

Fig. S1. The relative abundance of bacterial groups across the LI. Reads in the 22 metagenomes were taxonomically annotated by comparison with the SEED database and averaged by atoll. Atolls on the *x* axis are ordered south to north, left to right.

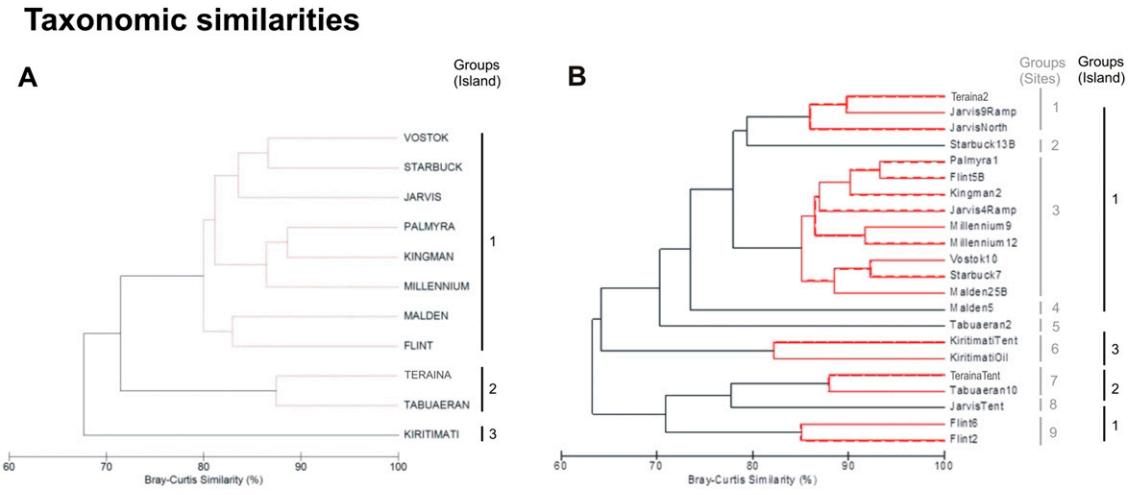

Fig. S2. Multivariate structure for the relative abundance of taxonomic similarities averaged by island (*A*) and at site level (*B*) analyzed using SIMPROF ($P < 0.01$).

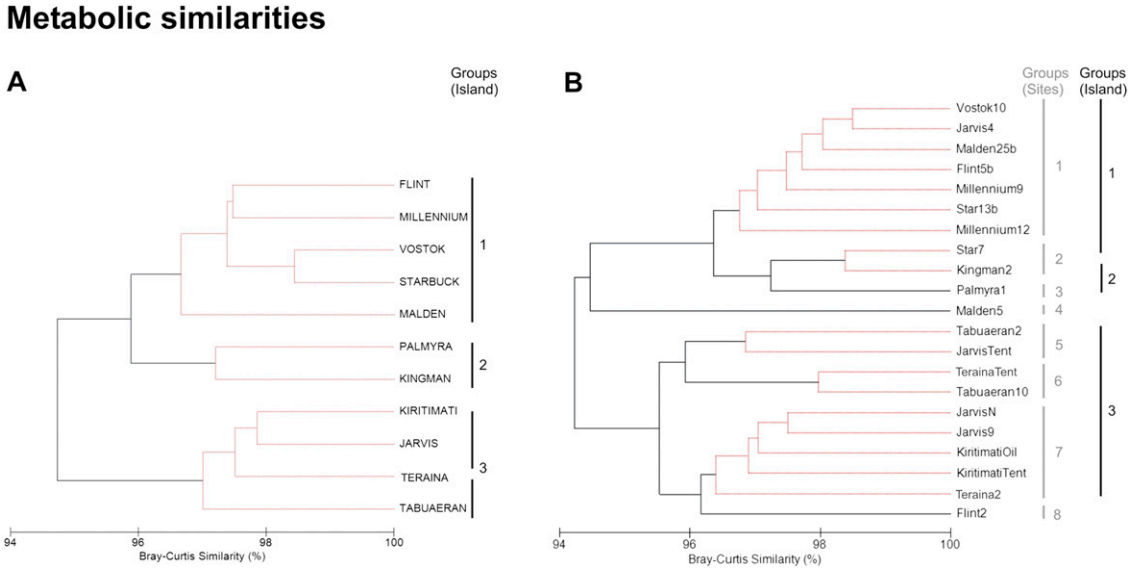

Fig. S3. Multivariate structure for the relative abundance of metabolic groupings averaged by island (*A*) and at site level (*B*) analyzed using SIMPROF ($P < 0.01$).

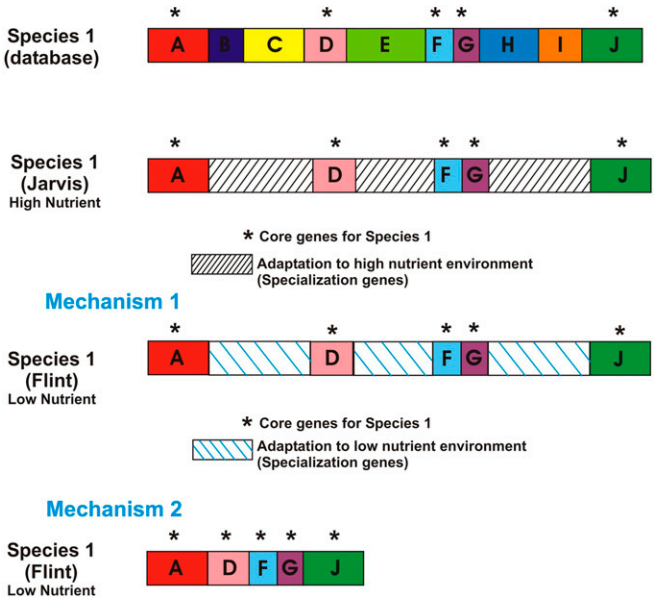

Fig. S4. Conceptual model depicting variation in genome size between strains due to different complements of environment-dependent specialization genes.

Table S1. Metagenomic libraries

Sample name	Date collected	Total no. of reads	Average read, bp	Latitude, °	Longitude, °	% GC content	Total no. of taxon similarities	Total no. of metabolism similarities
Flint 2	03/30/09	24,111	345.71	−11.41924	−151.82739	51	8,454	10,167
Flint 5	03/29/09	98,284	399.55	−11.43911	−151.81964	47	28,931	37,301
Flint 6	03/31/09	39,070	430	−11.44423	−151.81709	47	17,316	20,252
Jarvis 4	04/04/10	171,749	400.17	−0.38188	−159.99800	48	49,941	61,987
Jarvis 9	04/02/10	235,984	407.55	−0.36537	−160.00600	50	63,207	77,340
Jarvis North	11/13/10	66,808	384.81	−0.36902	−160.00819	52	20,340	23,781
Jarvis Tent	11/12/10	49,774	397.8	−0.369017	−160.00819	50	22,465	24,945
Kingman 2	10/31/10	225,914	381.44	6.387	−162.38600	52	47,187	61,909
Kiritimati Oil	11/21/10	156,251	393.74	1.99095	−157.48251	51	54,015	62,427
Kiritimati Tent	11/20/10	30,131	387.29	2.0085833	−157.48945	50	11,457	12,895
Malden 25	04/11/09	164,564	381.72	−4.03326	−154.95094	52	42,411	51,614
Malden 5	04/10/09	48,258	349.25	−3.99531	−154.94452	57	10,993	12,902
Millennium 12	04/19/09	26,895	357.99	−9.90774	−150.19974	53	7,801	9,190
Millennium 9	04/17/09	39,933	373.89	−9.91672	−150.21072	54	13,032	15,772
Palmyra 1	10/25/10	170,135	386.57	5.86646	−162.11346	37	53,623	71,606
Starbuck 13	04/05/09	29,347	401.56	−5.66441	−155.87346	46	11,874	15,052
Starbuck 7	04/06/09	83,014	431.87	−5.62220	−155.88002	42	34,058	45,652
Tabuaeran 10	11/04/10	104,845	396.09	3.82595	−159.34957	56	39,697	46,930
Tabuaeran 2	11/06/10	73,712	411.89	3.84085	−159.36047	55	31,874	36,241
Teraina 2	11/09/10	42,317	385.46	4.70242	−160.39212	53	12,035	14,199
Teraina Tent	11/08/10	285,841	412.73	4.6867167	−160.42023	51	86,972	101,107
Vostok 10	04/01/09	83,219	357.47	−10.05835	−152.30954	44	32,232	41,533
Total	—	2,250,156	—	—	—	—	699,915	854,802

Metadata and library details for the 22 metagenomes generated from the 22 sites sampled at 11 atolls.

Table S2. Predictor variable categories used for CCA and DistLM

Sample name	Hard coral	Crustose coralline algae	Calcified macroalgae	Soft coral	Macroalgae	Turf	Other*	$NO_3^- + NO_2^-$, μM	PO_4^3, μM	NH_4^+	Distance from equator†
Flint 2	75.85	13.10	2.00	0.00	1.15	7.65	0.25	1.09	0.291	4.32	−11.41924
Flint 5	83.00	9.00	0.20	0.00	0.95	6.60	0.25	0.82	0.161	2.46	−11.43911
Flint 6	ND	ND	ND	ND	ND	ND	ND	0.79	0.147	2.13	−11.44423
Jarvis 4	46.30	27.20	0.70	0.00	7.90	17.70	0.30	4.65	0.384	ND	−0.38188
Jarvis 9	57.90	9.30	0.00	0.00	3.40	29.20	0.20	4.54	0.427	ND	−0.36537
Jarvis North	10.70	31.90	1.00	0.00	4.90	50.30	1.30	4.50	0.441	ND	−0.36902
Jarvis Tent	57.65	12.00	0.15	0.35	6.90	21.65	0.95	3.28	0.392	0.289	−0.369017
Kingman 2	14.55	51.20	12.70	0.40	0.95	18.25	1.95	1.44	0.247	0.203	6.387
Kiritimati Oil	2.21	0.00	6.68	0.00	4.00	82.47	4.63	2.26	0.241	1.15	1.99095
Kiritimati Tent	23.36	2.50	7.77	1.64	0.14	58.86	5.73	2.33	0.291	0.436	2.0085833
Malden 25	73.63	5.37	1.84	0.00	1.79	15.95	1.42	3.90	0.264	1.67	−4.03326
Malden 5	86.67	4.56	0.00	0.00	0.22	8.11	0.44	2.82	0.253	1.24	−3.99531
Millennium 12	65.30	11.00	12.30	0.00	1.10	10.10	0.20	2.28	0.216	0.674	−9.90774
Millennium 9	69.30	6.20	7.70	0.00	0.30	15.70	0.80	2.10	0.191	0.803	−9.91672
Palmyra 1	45.70	16.30	5.60	2.00	1.40	23.90	0.60	0.52	0.195	0.365	5.86646
Starbuck 13	25.55	12.35	57.95	0.00	0.10	0.20	3.85	2.87	0.247	1.93	−5.66441
Starbuck 7	21.68	49.42	21.84	0.00	1.21	4.47	1.21	4.83	0.254	3.18	−5.6222
Tabuaeran 10	22.08	30.08	34.62	0.00	8.69	3.92	0.46	2.78	0.299	0.630	3.82595
Tabuaeran 2	39.23	20.18	16.32	0.00	2.41	17.86	4.00	1.60	0.185	0.936	3.84085
Teraina 2	20.96	21.96	0.00	0.76	32.12	20.44	3.76	2.24	0.279	0.458	4.70242
Teraina Tent	8.64	44.93	6.64	6.36	2.36	30.79	0.29	1.92	0.278	0.348	4.6867167
Vostok 10	81.40	14.40	0.35	0.00	0.15	3.70	0.00	1.75	0.158	1.91	−10.05835
Average	44.4	18.7	9.35	0.548	3.91	21.3	1.55	2.51	0.266	1.32	—

Benthic coverage for each functional group is shown as percent cover. Nutrient concentrations are calculated as micromoles per liter. NH_4^+, ammonium; NO_3^-, nitrate + nitrite; PO_4^3, phosphate.

*Other benthic organisms.

†Distance from equator as absolute value of the latitude in decimal degrees.

Table S3. Taxa within the seven functional groups used to classify benthic macroorganisms

Hard coral	
Acropora, Astreopora, Cyphastrea, Cycloseris, Echinophyllia, Favia, Favities, Fungia, Gardineroseris, Halomitra, Herpolitha, Hydnophora, Leptastrea, Leptoseris, Lobophyllia, Montastrea, Montipora, Pavona, Platygyra, Pocillopora, Porites, Psammocora, Sandolitha, Scapophyllia, Sytlophora, Tubastrea, Turbinaria	
Calcified macroalgae	
Galaxaura, Halimeda, Neomeris, Peyssonellia	
Soft coral	
Cladiellqa, Dendronephtya, Lobophytum, Pachyclavularia, Sarcophyton, Sinularia, Stereonephthya	
Fleshy macroalgae	
Avrainvillea, Brown crust, *Caulerpa, Dictyosphaeria, Dictyota, Hypnea, Lobophora, Valonia*	
Other benthic organisms	
Cyanobacteria, Heteractis, Holothurian, Hydroid, *Millepora, Rhodactis,* Sand, Sponge, *Stylaster, Tridacna,* Tunicate, Zoanthid	

Crustose coralline algae and fleshy turf algae were identified as functional groups only.

Table S4. Summarized results of a DistLM for associations of microbial community structure (Taxa) and metabolic function (Metabolism)

Variable	AICc	SS, trace	Pseudo-*F*	*P*	Prop., %*	res.df
		Taxa				
Hard coral	129.52	1,438.1	3.4065	0.0215	15.2	19
		Metabolism				
Distance from equator	53.323	47.953	4.2767	0.0147	18.4	19

Prop., proportion of variance; res.df, degrees of freedom for the residual; SS, sum of squares. The total number of predictors included equals 10 (the percent cover of benthic functional groups, distance from the equator, and nutrient availability).
*The best-fit models are shown, along with the proportion of variability in the multivariate response explained by that variable (Prop.).

Table S5. Significance test for the linear correlations of metabolic pathway abundance with phosphate concentration

Metabolic pathway	*r*	*P*
Conjugative transfer	0.863	**0.001**
Bacterial chemotaxis	0.598	0.052
Nitrate and nitrite ammonification	0.628	**0.038**
Cobalt–zinc–cadmium resistance	0.637	**0.035**
Multidrug resistance	0.617	**0.043**
Ton and Tol transport	0.650	**0.03**
Chlorophyll biosynthesis	−0.552	0.079
Photosystem II	−0.534	0.091
Ribosome SSU bacterial	−0.620	**0.042**

P values < 0.05 are shown in bold.

Table S6. Collinearity among predictor variables using Pearson's coefficient, *r*

	Hard coral	Crustose coralline algae	Other calcified algae	Soft coral	Macroalgae	Turf algae	Other benthic	Nitrate	Phosphate
Hard coral									
Crustose coralline algae	−0.536								
Other calcifying algae	−0.373	0.193							
Soft coral	−0.355	0.339	−0.089						
Macroalgae	−0.287	0.120	−0.151	0.001					
Turf algae	−0.533	−0.196	−0.297	0.213	0.051				
Other benthic	−0.548	0.196	0.294	−0.030	0.211	0.538			
Nitrate	−0.161	0.163	0.038	−0.025	0.105	0.077	−0.077		
Phosphate	−0.266	0.155	−0.175	−0.017	0.258	0.299	−0.088	**0.708**	
Distance from equator	0.487	−0.072	0.054	−0.071	−0.250	−0.531	0.316	**−0.641**	**−0.741**

The correlations between distance from equator and nutrient concentrations are shown in bold.

Table S7. The 19 bacterial taxa included in analyses of community structure

Phyla	Classes	Orders
Actinobacteria		
Bacteroidetes		
Cyanobacteria		
Firmicutes		
Proteobacteria	Alphaproteobacteria	Rickettsiales
		Rhizobiales
		Rhodobacterales
		Rhodospirillales
		Sphingomonadales
		Other
	Betaproteobacteria	
	Gammaproteobacteria	Alteromonadales
		Enterobacteriales
		Oceanospirillales
		Pseudomonadales
		Vibrionales
		Other
	Deltaproteobacteria	
	Epsilonproteobacteria	
Other		

Bacterial taxa were categorized at the phylum level except for the Proteobacteria (which made up 48–87% of the bacterial community). Rarer phyla (those that made up <5% of the relative abundance across all libraries) and unclassified bacteria were designated as "Other bacteria." Because of their higher abundances, Proteobacteria were categorized by class, and Gammaproteobacteria and Alphaproteobacteria (representing 13–64% and 12–36% of the bacterial communities, respectively) were further categorized by order. Rarer Alphaproteobacteria and Gammaproteobacteria orders that made up <1% across all libraries were combined and designated as "Other Alphaproteobacteria" and "Other Gammaproteobacteria."