Supplemental Methods

Satellite measurements

Monthly and eight day average estimates for surface chlorophyll *a* concentration and surface productivity were downloaded from the Coastwatch browser website. The chlorophyll-*a* measurements were from the data categories "Chlorophyll *a*, SeaWiFS, 0.04167 degrees, West US Science Quality + Chlorophyll-a" and "Chlorophyll-a, Aqua MODIS NPP, 0.05 DEGREES, Global, Science Quality". Productivity was extracted from the categories "Primary Productivity, SeaWiFS and Pathfinder, 0.1degrees, Global, EXPERIMENTAL" and "Primary Productivity, NASA Aqua MODIS and Pathfinder, 0.1 degrees, Global, EXPERIMENTAL" (Behrenfeld & Falkowski 1997). Data from SeaWiFS and MODIS together spanned our data set, with overlap in the middle. In cases where the data overlapped, we gave priority to the MODIS data.

Surface eight day averages of photosynthetically active radiation (PAR) (Frouin et al. 2003), colored dissolved organic matter to chlorophyll *a* ratio (CDOM) (Mannino et al. 2008) and particulate organic carbon concentrations (Morel & Gentili 2009) were downloaded from the ocean color data site. A 3x3 grid of pixels was extracted from around SPOT using the SeaDas program (Fu et al. 1998), and the weighted average (using weights from the SeaDas output) of this grid was used in downstream analysis. Monthly sea surface height differential was downloaded from Coastwatch as "Sea Surface Height Deviation, Aviso, 0.25 degrees, Global, Science Quality" (Ducet et al. 2000).

We obtained meteorological data, including minimum and maximum daily air temperatures and precipitation data from the weather station at nearby Avalon airport (33.405°N 118.415°W). Wave height, average wave period, and dominant wave period from a buoy in nearby Santa Monica Bay (33.749°N, 119.053°W) were downloaded from the National Buoy Data Center. Pacific Fisheries and Environmental Laboratory (PFEL) estimates of coastal upwelling, and Sverdrup transport at (33°N, 119°W), along with Multivariate ENSO Index scores were downloaded from the National Oceanographic and Atmospheric Administration (NOAA).

Assigning Taxonomic Identities to ARISA peaks

We assigned taxonomic identity to each ARISA fragment size by identifying which clones from our clone libraries had fragment sizes that fell within the range of peak sizes that were assigned to an ARISA OTU bin. In cases in which an ARISA OTU corresponded in size to more than one clone in

our clone library database, we prioritized our clones based on where they were isolated. For fragments that were more abundant in surface waters than at 890m, we prioritized fragments according to the first number in parentheses, while fragments that were more abundant at 890m than 5m we prioritized according to the second number in parentheses:

(1;3) observed ARISA length of SPOT clones from 5m across all seasons (2;2) SPOT clones from 150m (3;1) SPOT clones from 890m (4;5) Clones from the Pacific Ocean (near Hawaii) and Atlantic Ocean (near the Amazon river outfall) from 5m. (5;7) published cyanobacterial intergenic spacer (ITS) sequences (6;6) observed ARISA lengths from 16S-ITS clones from surface waters of the Indian Ocean: (7;9) in silico amplification of marine isolate genomes from the photic zone(8;4) Clones from the Pacific and Atlantic oceans from 500m and below (9;8) in silico amplification of marine isolate genomes from below the photic zone. Chow et al (2013) provide a full description about these datasets and how they were used to assign identity to ARISA OTUs.

In cases in which more than one clone from the highest priority category fell into a given bin, we selected the clone that had the highest number of instances in our clone libraries.

Environmental parameter variability

We tested for seasonal variability of each measured environmental parameter by applying generalized additive mixed effects models (Wood 2004, 2006). Each variable was modelled according to the equation

$$y = \mu + m_1(\text{time}) + m_2(\text{DoY}) + \varepsilon$$

In this equation “y” is the transformed (for normalization purposes) value of the environmental parameter. m_1(time) is a univariate smooth thin plate regression spline modelling long term variability as a function of the number of days that had elapsed since the beginning of the study. m_2(DoY) (Day of Year) is a cyclic penalized cubic regression smooth spline of one year period. μ is essentially the mean of y, and the m_1 and m_2 functions describe how y deviates from this mean over time. ε is the error term. The model was set up to allow for the data to have a continuous lag-1 autoregressive structure. This model reflects equation 1 in Ferguson et al. (2008) as well as approaches demonstrated elsewhere (Wood 2006, 321–324) and identifies seasonal variability that is not perfectly sinusoidal as well as long term trends that are non-linear. The model was run using the “gamm” function of the “mgcv” R-package (Wood 2011).

For both the seasonal and long term spline function we determined the model's estimated degrees of freedom (EDF) which is essentially a measure of the complexity of the spline. For instance, seasonal splines of EDF of 1 are perfectly sinusoidal while higher EDF relate to unevenly shaped seasonal peaks or local maxima. Long term splines with EDF of 1 are linear, while higher EDF correspond to curved long term splines which may have maximum and minimum values at years within (rather than at the extremes of) the dataset. We also determined p-values for both the seasonal and long term splines, where P is the probability of the null hypothesis that the EDF of the smooth term is actually zero (no prediction by that spline). R2 values for the entire model were also determined.

For each fitted nonparametric regression model, we interrogated the cyclic seasonal spline to determine the month in which that factor had the highest value and the month in which that factor had the lowest value. We interrogated the long-term spline function to determine whether there appeared to be a linear on non-linear change over time and identified the years that appeared to have the highest and lowest values.

To test whether each variable appeared to relate to the Multivariate El-Niño Southern Oscillation Index (MEI) a second GAMM model using MEI instead of year as a predictor variable for the long term spline was fit to each variable. Thus this model was of the form $y = \mu + m_1(MEI) + m_2(DoY) + \varepsilon$. We determined the Akaike information criterion (AIC) of both the original (Year) and modified (MEI) GAMM models. In cases where the second model had a lower AIC than the former, and in which p-values suggested the $m_1(MEI)$ model had good fit we would say that the variable seemed to be driven by variability in MEI.

Seasonal variability of microbial community structure

Graphical approach

We used the "vegdist" function in the "vegan" (Oksanen et al. 2011) package to estimate Bray-Curtis dissimilarity in community structure between all pairs of samples in the dataset, thereby generating a dissimilarity matrix; we calculated similarity matrix by subtracting the dissimilarity scores from one. We determined upper and lower bounds for these similarity scores by examining similarity scores between machine replicates (upper bounds) and randomized samples (lower bounds). Machine replicates were identical samples run on different fragment analysis gel lanes. We determined the machine replicate similarity for every sample in the data set and calculated average machine replicate similarity. Similarities between randomized samples were determined by arbitrarily picking pairs of

samples and then shuffling the orders of the abundances of each OTU. This process was repeated 1000 times and the average value of similarity between randomized samples was recorded.

We determined the temporal difference or lags, in days, between all pairs of samples and the Bray-Curtis similarity between those same pairs of samples. Pairs of samples were binned by their lags in by 30.416 day (the average number of days in a month) intervals and average similarity for pairs of samples in each monthly bin was determined. Accordingly our first bin returned the average Bray-Curtis similarity value for all samples collected between 15 and 45 days apart, the next bin returned the average for all samples between 46 and 76 days apart and so forth. Bray-Curtis similarity scores that oscillated with a period of one year were considered seasonal. A t-test was applied to ask whether samples that were taken one month (15 to 45 days) apart were statistically more similar than samples taken six months apart.

Mantel test approach

Mantel tests were applied to look for seasonality using a 'seasonal difference matrix' (S). S was calculated as follows: "D", a matrix of the difference in serial days between each pair of samples was calculated. 2) "DM", a matrix containing the 365.25 day modulus of each value in D was calculated. "S" was calculated from each value of DM such that if the value was less than 180.625 it was kept and if greater the value was subtracted from 365.25 and the difference was kept.-These 'seasonal difference matrixes' were compared to the community's Bray-Curtis dissimilarity matrix using the "mantel" function in the "ecodist" package for R (Goslee & Urban 2007). Depths where the seasonal matrix significantly correlated to the Bray-Curtis dissimilarity matrix were said to be seasonal.

Interannual variability of microbial community structure

We binned samples by 365.25 day (the average number of days in a year) intervals and applied the same analysis described previously. Thus the first bin would contain the mean of all samples taken between 1 month and 12 months apart, the second all samples taken between 13 and 24 months apart and so on. For each depth we performed ANOVA to ask whether the mean similarity between samples within each bin differed between those bins. In cases in which the ANOVA suggested statistically significant differences, we performed a Tukey corrected t-test for each pair of bin categories to determine which bins had statistically different mean similarity scores. As for the seasonality comparison, we performed Mantel tests to examine the relationship between difference in serial day ("D" as calculated above) and community dissimilarity. Depths in which samples that were more

temporally distant had higher Bray-Curtis dissimilarity scores would be said to show long term change in community structure.

Alpha diversity

Variability between depths

Mean values of Richness of species with greater than 1% 0.1% and 0.01% relative abundance, inverse Simpson index (ISI), Shannon indexes of biodiversity and Pelou's index of evenness were determined at each depth along with 95% confidence intervals of those means. We investigated whether richness at 0.1% and ISI differed between depths using analysis of variance ("AOV" function in the R's "stats" package). A Tukey corrected t-test compared all pairs of depths in order to determine which pairs of depths have different mean richness and ISI.

Relation to season

Richness and ISI were investigated with the same nonparametric regression model used to investigate seasonality. Seasonal and long term spline functions were investigated and depths with seasonal and long term trends were noted. We identified months and years of highest and lowest biodiversity and parameters for the splines used to fit these data.

Relation to community similarity between depths

We examined, for each depth, whether richness and/or ISI was correlated with the similarity of that depths community structure to the community structure at each other depth. Our goal was to identify whether biodiversity at each depth was driven by influence of OTUs from other depths.

Inter-depth community similarity was determined for each pair of depths as the Bray-Curtis similarity between those depths' communities in a given month. Scatterplots of richness vs inter-depth similarity and ISI vs inter-depth similarity were visually investigated to determine whether simple correlations were sufficient to describe relationships between the factors. After determining that no non-linear relationships were present, Pearson correlation tests were applied to determine whether richness and/or ISI at each depth was statistically correlated with the inter-depth community similarity between that depth and each other depth. R values of the Pearson correlations and the 95% confidence intervals of those R values were identified for each comparison. Relationships whose and confidence intervals did not overlap zero were identified as having a statistically significant relation between inter-depth community similarity and biodiversity.

Relation to community change

We queried whether alpha diversity was higher for communities that were changing the most rapidly. To calculate the rate of community change, we compared the Bray Curtis dissimilarity of each month to the community in the previously sampled month and refer to this dissimilarity score henceforth as "Bray-Curtis shift (BCS)". As in the inter-depth similarity analysis, scatterplots of BCS vs Richness and BCS vs ISI at each depth were investigated for non-linear relationships. After determining that no parabolic relationships were present, Pearson correlation tests were applied to determine whether BCS was statistically significantly related to richness and Simpson's index at each depth. R values of the Pearson correlations and the 95% confidence intervals of those R values were identified and depths with confidence intervals not overlapping zero were identified as having a statistically significant relation between community change and biodiversity.

Environmental parameters and community structure: Mantel tests

We applied partial Mantel tests that examine the model $Y = a + bS + cD + dX$ where Y is the Bray Curtis similarity matrix of the community structure, S is the seasonal distance matrix, D is the serial day distance matrix (S and D are described above), and X is the similarity matrix for the variable of interest. Our environmental data set was missing values for a few environmental variables. Because Mantel tests are not able to handle missing data, we filled in our data set using multiple imputation, a method which fills in missing values with numbers that are reasonable estimates but reflect the uncertainty of the data (King et al. 2001). We generated 25 imputed data sets using the "Amelia" R-package (Honaker et al. 2006; Honaker & King 2010) and performed the Mantel test, using the "ecodist" R-package (Goslee & Urban 2007), on each of these imputed data sets. We then report the median rho score of the 25 Mantel tests, and the p-value corresponding to this median rho score. Because we ran many tests in parallel, in addition to calculating p-values for each environmental parameter, we also estimated the false discovery rate "Q" from the p-values at each depth using the "qvalue" R-package (Dabney et al. 2004).

Temporal dynamics of microbial taxa over time

Transformations

Taxonomic groups and individual OTUs were log of odds transformed using the "logit" function, in the "car" R-package (Fox and Weisberg, 2011), with an adjustment factor of 0.001.

Taxonomic Groups

We examined seasonality and long term temporal variability of class level taxonomic groups, more abundant order and family level taxonomic groups, each of the sub clades of SAR11, all of the OTUs of SAR11 Surface group 1, clades of Flavobacteria and genera of Marine group A. To investigate these temporal dynamics we attempted to fit the group's log-odds transformed abundance (Y') using the same nonparametric model used to fit the environmental variables. For each OTU, months and years of both maximal and minimal abundance, estimated degrees of freedom of each spline term and p-values for each spline term were generated. We report all taxonomic groups which were fit by this model with an R^2 value greater than 0.10. False discovery rates (Q) were calculated from the p-values of each category of taxa investigated.

OTUs

The previously described nonparametric regression model was also applied to each of the 100 most abundant OTUs (where Y' is the transformed relative abundance of the OTU under investigation). We recorded the number of bacteria, out of 100 that were fit with R^2 values of 0.1 and 0.2. Of these bacteria, we determined which were fit by the seasonal spline with a p-value of less than 0.05 and which were fit by the long-term spline with a p-value less than 0.05. To determine if the fraction of seasonally variable and interannually variable bacteria (seasonal term $P < 0.05$, $R2 > 0.2$) differed between depths, we applied the "chisq.test" function in R. We recorded the taxonomic identity, ARISA fragment length and statistics for each temporally variable OTU that was fit by the model with an R^2 value of greater than 0.2.