Supplementary File 1: Tables and Figures

Supplementary Tables

| R OF REFERENCE<br>.ES                 | 1   | 2   | m           | 4                              |
|---------------------------------------|---|---|-------------|--------------------------------|
| NG NUMBER<br>) SAMPLI                 | 9 234   | 33 9  | 96 103      | 31 31                          |
| SAMPLIT<br>PERIOI                     | ity 2000-200  | 2001-200  | 1988-200    | 2003-200                       |
| DIVERSITY MEASURE                     | Richness and Shannon divers                                   | Chao richness   | Richness    | Richness                       |
| RAREFACTION<br>DEPTH                  | 4266 and 150  | NA  | NA          | 150                            |
| TAXONOMIC<br>QUANTIFICATION<br>METHOD | <i>de novo</i> clustered OTUs;<br>RDP classification of reads | RDP classification of <i>de</i><br><i>novo</i> clustered OTUs | ARISA OTUs  | RDP classification of reads    |
| <b>DATA TYPE</b>                      | V6 Targeted 454<br>Sequence                                   | 16S Targeted<br>Sanger Sequence                               | ARISA       | Metagenomic<br>Sanger Sequence |
| DATA SOURCE                           | MICROBIS  | POMMIER2007   | FUHRMAN2008 | GOS                            |

#### Table S1: Data sources

| VARIABLE                     | METRIC                    | UNITS              | DEPTH-SPECIFIC | SOURCE |
|------------------------------|---------------------------|--------------------|----------------|--------|
| Apparent Oxygen Utilization* | Monthly Mean              | mL/L               | Yes            | 5      |
| Calcite Concentration        | Annual Mean               | mol/m <sup>3</sup> | No             | 6      |
| Chlorophyll A Concentration  | Annual Mean               | mol/m <sup>3</sup> | No             | 6      |
| Chlorophyll A Concentration* | Annual Range              | mol/m <sup>3</sup> | No             | 6      |
| Chlorophyll A Concentration  | Annual Standard Deviation | mol/m <sup>3</sup> | No             | 7      |
| Chlorophyll A Concentration* | Monthly Mean              | mol/m <sup>3</sup> | No             | 7      |
| Cloud Fraction               | Annual Mean               | %                  | No             | 6      |
| Cloud Fraction               | Annual Standard Deviation | %                  | No             | 7      |
| Cloud Fraction               | Monthly Mean              | %                  | No             | 7      |
| Day Length*                  | Monthly Mean              | h                  | No             | 8      |
| Diffuse Attenuation          | Annual Mean               | m-1                | No             | 6      |
| Dissolved Oxygen             | Monthly Mean              | mL/L               | Yes            | 5      |
| Distance from Land           | NA                        | km                 | NA             | 9      |
| Dust Flux                    | Annual Mean               | $g/m^2$            | No             | 10     |
| Dust Flux                    | Annual Standard Deviation | $g/m^2$            | No             | 10     |
| Dust Flux                    | Monthly Mean              | $g/m^2$            | No             | 10     |
| Insolation*                  | Annual Mean               | $W/m^2$            | No             | 6      |
| Insolation*                  | Annual Standard Deviation | $W/m^2$            | No             | 7      |
| Insolation*                  | Monthly Mean              | $W/m^2$            | No             | 7      |
| Latitude                     | NA                        | Degrees            | NA             | NA     |
| Nitrate Concentration*       | Annual Mean               | µmol/L             | No             | 6      |
| Nitrate Concentration        | Annual Standard Deviation | µmol/L             | No             | 11     |
| Nitrate Concentration*       | Monthly Mean              | µmol/L             | Yes            | 11     |
| Nitrate:Phosphate Ratio      | Monthly Mean              | NA                 | Yes            | 11     |
| Ocean Depth*                 | NA                        | m                  | NA             | 7      |
| Ocean Temperature            | Annual Mean               | °C                 | No             | 6      |
| Ocean Temperature*           | Annual Range              | °C                 | No             | 12     |
| Ocean Temperature            | Annual Standard Deviation | °C                 | No             | 12     |
| Ocean Temperature*           | Monthly Mean              | °C                 | Yes            | 12     |
| Oxygen Saturation            | Monthly Mean              | NA                 | Yes            | 5      |
| pH                           | Annual Mean               | NA                 | No             | 6      |
| Phosphate Concentration*     | Annual Mean               | µmol/L             | Yes            | 6      |
| Phosphate Concentration      | Annual Standard Deviation | µmol/L             | No             | 11     |
| Phosphate Concentration*     | Monthly Mean              | µmol/L             | Yes            | 11     |
| Proximity to Thermocline*    | Monthly Mean              | log(m)             | No             | 13     |
| Pycnocline Depth             | Annual Standard Deviation | m                  | No             | 13     |
| Pycnocline Depth             | Monthly Mean              | m                  | No             | 13     |
| Salinity*                    | Annual Mean               | PSS                | No             | 6      |
| Salinity*                    | Annual Standard Deviation | PSS                | No             | 14     |
| Salinity*                    | Monthly Mean              | PSS                | Yes            | 14     |
| Sea Ice Concentration        | Annual Mean               | %                  | No             | 9      |
| Sea Ice Concentration        | Annual Standard Deviation | %                  | No             | 7      |
| Sea Ice Concentration*       | Monthly Mean              | %                  | No             | 9      |
| Thermocline Depth            | Annual Standard Deviation | m                  | No             | 13     |
| Thermocline Depth            | Monthly Mean              | m                  | No             | 13     |
| Sampling Depth               | NA                        | m                  | NA             | NA     |
| Sampling Year                | NA                        | Julian Day Number  | NA             | NA     |

## Table S2: Environmental covariate metadata

| Table S3: Environmental covariate correlatio | ons |
|--|-----|
|--|-----|

| Calcite Concentration - Annual Mean  | 1 | ÷     | 1    | 1      | : :   | 1    |        | :      | • •   |        | :     | 1      |        | • •            | 1      |        | • •        | 1       | 1      |        | • •   | 1     | 1      | 1       | •          | 1      | • •         | • •     |        | • •            | ÷        |
|--|---|-------|------|--------|-------|------|--------|--------|-------|--------|-------|--------|--------|----------------|--------|--------|------------|---------|--------|--------|-------|-------|--------|---------|------------|--------|-------------|---------|--------|----------------|----------|
| Salinity - Monthly Mean  | 1 | ł     | 1    | 1      | •     | 1    |        | :      | • •   |        | 1     | 1      |        | • •            | 1      |        | • •        | 1       | 1      | 1      | • •   | 1     | 1      | •       | • •        | 1      | • •         | • •     | 1      | • •            | 0.041    |
| nteh humh - entrequer meoO   |   | ł     | ł    | ł      |       | ,    |        |        |       |        | :     |        |        |                |        |        |            | 1       | • •    |        |       |       | i.     |         |            | 1      |             |         | ı.     | ų.             | 0.151    |
| Cloud Fraction - Monthly Mean  |   | ł     | ;    |        |       | ;    |        | :      |       |        | :     |        |        |                |        |        |            |         |        |        |       |       |        |         |            |        |             |         |        | 0.279          | 0.034    |
| Pycnocline Depth - Monthly Mean  |   |       | ;    |        |       | ,    |        |        |       |        |       |        |        |                |        |        |            |         |        |        |       |       |        |         |            |        |             |         | 13     | 182            | 000      |
| Chlorophyl A Concentration - Menthly Mean  |   |       |      |        |       |      |        |        |       |        |       |        |        |                |        |        |            |         |        |        |       |       |        |         |            |        |             |         | 900    | 015            | 212 0    |
| SIN - MUMPY  |   |       |      |        |       |      |        |        |       |        |       |        |        |                |        |        |            |         |        |        |       |       |        |         |            |        |             | . 8     | 0.0    | 191            | 13 0.    |
| AV doment  |   |       |      |        |       |      |        |        |       |        |       |        |        |                |        |        |            | Ì       |        |        |       |       |        |         |            | Ì      |             | - 00    | 10 0.1 | 881            | 88       |
| niek humA - neisamut seallid   | : |       | 1    |        |       |      |        | :      |       |        | :     |        |        |                |        |        |            |         |        |        |       |       |        |         |            | :      |             | 0.0     | 9 000  | 88             | 10       |
| moM luumA - noindoani  | 1 | ì     | 1    | 1      |       | 1    |        |        |       |        | :     |        |        | •              |        |        |            | 1       | •      |        | •     | 1     | 1      |         |            | 1      | - 0.0       | 0.00    | 0.16   | 0.87           | 0.00     |
| naM humh - Annu - Yoni - Salini - Salin | 1 | ł     | 1    | 1      | •     | 1    |        | :      | •     |        | 1     | 1      |        | •              | 1      |        | • •        | 1       | •      |        | •     | 1     | 1      |         | •          | 1      | 0.172       | 0.051   | 0      | 0.148          | 10.0     |
| nieł wyska w obrzy stadysowa w obrzy stadysowa w obrackie w o   | • | ÷     | 1    | 1      | • •   | 1    |        | :      | • •   |        | :     |        |        | • •            | •      |        | • •        | 1       | •      |        | • •   | 1     | 1      |         | • •        | 0      | 0.001       | 0.005   | 0.004  | 0.008          |          |
| AN - thqsd nssoO   | : | ł     | ł    | 1      | • •   | ;    |        | :      | • •   |        | :     |        |        | • •            | •      |        |            | :       | • •    |        | • •   | :     | 1      | • •     | - 000      | 0.12   | 0.058       | 0.052   | 0.015  | 0.038          | 0.11     |
| sgmA humA - suitssqmsT naoO  |   | ł     | ł    | 1      |       | 1    |        | :      |       |        | :     |        |        | • •            |        |        | • •        | 1       | • •    | 1      |       | 1     | ÷      | -       | 0,049      | 0.072  | 0.153       | 0.17    | 0.01   | 0.021          | 0.027    |
| meth laumA - noineineonaO stedqooff  |   | ł     | ł    | i      |       | ,    |        |        |       |        | :     |        |        |                |        |        |            |         |        |        |       |       |        | 0.022   | 1000       | 0.04   | 0.538       | 0.262   | 0.177  | 0.648          | 0.058    |
| Prosphate Concentration - Monthly Mean   | : |       | ;    |        |       | ;    |        | :      |       |        | :     |        |        |                |        |        |            |         |        |        |       | ;     |        | 0.018   | 0000       | 058    | 0.031       | 0.197   | 0.23   | 0.644          | 0.03     |
| nesh yangin - Maning - Mgan  |   |       |      |        |       | ,    |        | :      |       |        | :     |        |        |                |        |        |            |         |        |        |       | ;     | 002    |         |            |        | 80          | 0 100   | 694    | 80.            |          |
| UDBANA DRUBS RUBU - UDBBBOD ARDIN  |   |       |      |        |       |      |        |        |       |        |       |        |        |                |        |        |            |         |        |        |       |       | 27 0   | 18      | 69         | 8      | 87 U        | 002     | 141 0  | 0<br>90        | 87<br>97 |
|  |   |       | Ì    |        |       |      |        |        |       |        |       |        |        |                |        |        |            | Ĵ       |        |        | . 6   |       | 0.0    | 13 0.   | 18 0.0     | 3 0.4  | 01 U.       | 20 0.0  | 48 0.0 | 000            | 31 0.0   |
| AV - Intel moderousid  |   |       |      |        |       |      |        | :      |       |        |       |        |        |                |        |        |            | Ċ       |        |        | 00 0  | 0     | 88     | 88      | 80         | 0      | 3 61<br>2   | 0.0 0.0 | 0.0    | 88:            | 900      |
| Themselve Depth - Mean Themselve   | : | ľ     | 1    | •      | : :   | 1    |        | :      |       |        | •     |        |        |                |        |        | : :        |         |        | - 2    | 000   | 0.05  | 8 0.23 | 000     | 0.02       | -      | 100         | 1 0.00  | 9 0.80 | 0.16           | 6 0.00   |
| noinived budance forms - Armen Devinion  | 1 | ľ     | 1    | 1      |       | 1    |        | 1      |       |        | 1     |        |        |                |        |        | : :        | 1       | 1      | 0.34   | 0.00  | °     | 0.22   | 000     | 10.0       | 0      | 6000        | 0.00    | 0.36   | 0.42           | 0000     |
| Chlorophyll A Concentration - Amuel Range  | 1 | ł     | 1    | 1      | •     | 1    |        | •      | • •   |        | :     | 1      |        | •              | 1      |        | • •        | 1       | 1000   | 0.012  | 000   | 0     | 0000   | 60.09   | 0.14       | 0.218  | 0.612       | 0.038   | 0000   | 51000          | 0.126    |
| naiM yfdinol - Noniniy Mean  | : | ÷     | ;    | 1      | • •   | ;    |        | :      | • •   |        | :     | •      |        | • •            | •      |        | • •        | 1       | 0.002  | 0.236  | 0.237 | 0.001 | 0.83   | 0.025   | 0.002      | 0.048  | 0.014       | 0.264   | 0.234  | 0.637          | 0.001    |
| noiniyə Davhara Bandara Devinion   | : | ÷     | ;    | 1      |       | ;    |        | :      |       |        | :     | •      |        | • •            |        |        | : :        | 0.005   | 0.02   | 0.017  | 0.062 | 0.002 | 0.001  | 0       | 0.154      | 0.245  | 0.185       | 0.124   | 0.018  | 0.014          | 0.24/    |
| moM humnA - Hq   | : | ÷     |      |        |       | ,    |        |        |       |        | :     |        |        |                |        |        | 0.054      | 0.164   | 0.038  | 0.013  | 0.062 | 0.001 | 0.178  | 0.004   | 0.022      | 0.058  | 0.114       | 0.0042  | 0.016  | 0.104          | 0.028    |
| nisht hunuk - neitennom 2 steat  |   |       |      | ,      |       | ,    |        | :      |       |        | :     |        |        |                |        |        | 200        | 1987    | 0 112  | 0.203  | 2100  |       | 0.81   | 0018    | . 8        | 018    | 0083        | 2000    | 0.186  | 5590           | 77003    |
| model foarmek - zuiet read   |   |       |      |        |       | ,    |        |        |       |        |       |        |        |                |        | 037    | 700        | 0.28    | 048    | 027    | 200   |       | 830    | 10      | 00         | 023    | 810         | 620     | 610    | 041            | 88       |
| Sea lee Concentration - Monthly Mean   |   |       |      |        |       |      |        |        |       |        |       |        |        |                | 500    | 103    | 17.7       | 682 0   | 014 0  | 0027   | 828   | 100   | 125 0  | 014     | 610<br>610 | 115    | 075 0       | 002 0   | 013 0  | 218 0          | 002      |
| ursee áutoose - usorzuun uzšáxo mareldy  |   |       |      |        |       |      |        |        |       |        |       |        |        | . 8            | 0.800  | 502 0. | 120        | 39 0.   | 01 0.  | 53 0.  | 0 0   | 0.150 | 0.4 0. | 0.88    | 0 90       | 07 0   | 07 0.       | 04 0.   | 0.0    | 9 G I          | 500      |
|  |   |       |      |        |       | ĺ    |        |        |       |        |       |        |        |                | 12 0.0 | 10 12  | 17         | 18 0.   | 12 07  | 0.0    | 10 20 | 0     | 20 0   | 20      | 2000       | 62 03  | 21 0.0      | 15 0.0  | 16 0.1 | 8 2 I<br>8 2 I | 8_       |
| animial index is from to advisori  |   |       | ľ    |        |       | ĺ    |        |        |       |        | 1     |        |        | 1000           | 10     | 2 0.2  | 200        | 0.0     | 8 0.0  | 0      | 000   |       | 00     | 00      | 8          | 5 0.0  | 5 0'0 1     | 1 0.0   | 4 0.1  | 290            | 100      |
| noisivyd budoużł lenniń - sutersomyT nisoO   | : | ľ     | 1    |        | : :   |      |        | -      |       |        | •     |        | 5 0.03 | 0.0            | 6 0.0  | 2 0.0  | 0'0<br>* 0 | 0.0     | 0.0    | 0.0    | 000   | 0     | 0.0    | 0.9     | 000        | 0.0    | 5 0.0       | 0.10    | 0.0    | 0.0            | 900      |
| neinived hubmaß laumA - neinenvono3 studgooff  | 1 | ľ     | 1    | 1      |       | 1    |        | 1      |       |        | :     | 0.05   | 0.12   | 0.02           | 0000   | 0.22   | 000        | 0.21    | 0.06   | 10.01  | 1000  | 0000  | 0.30   | 8000    | 1000       | 0.12   | 0.18        | 0.06    | 0.03   | 0.31           | 9000     |
| neam hanna - Annan Mean  | 1 | ľ     | 1    | 1      |       | 1    |        | :      | • •   |        | 1 10  | 0      | 0.124  | 004600         | 0.084  | 0.375  | 0000       | 0.362   | 0.745  | 0.118  | 0.155 | •     | 0.351  | •       | 0000       | 0.057  | 16100       | 0.067   | 0.135  | 0.396          | 0000     |
| nasté vídnost - negyző bevíossið   | • | ÷     | 1    | 1      |       | ;    |        | :      | • •   | ;      | 0.388 | 0000   | 0.609  | 0.047          | 0.035  | 0.564  | 0.055      | 0.513   | 0.024  | 0.137  | 0000  | 0.003 | 0.56   | 0.018   | 0.051      | 0.223  | 0.102       | 0.021   | 0.153  | 0.921          | 0.003    |
| ntebb yhtende - zult izud  | ; | ÷     | 1    | 1      | • •   | ;    |        | :      | • •   | 0.025  | 0.063 | 0,015  | 0.013  | 10.00          | 0.857  | 0.024  | 0.005      | 0.02    | 0.02   | 0.015  | 2000  | •     | 0.022  | 0.009   | 0,028      | 0.02   | 0.014       | 0.025   | 0.014  | 0.028          | 0.007    |
| Ocean Temperature - Monthly Mean   | : | ł     | ł    | 1      |       | 1    |        | :      | 0.026 | 0.94   | 0.398 | 0.02   | 0.673  | 0.123          | 0.041  | 0.645  | 0.123      | 0.538   | 0.012  | 0.22   | 0.014 | •     | 0.635  | 0.017   | 0,008      | 0.149  | 26070       | 0.042   | 0.246  | 0.967          | 0.118    |
| noinived budants fammé - diqed ambonryf  |   | ł     | ;    | ,      |       | ,    |        | 1000   | 0.014 | 0.296  | 0.208 | 0.01   | 0.373  | 0.01           | 0.02   | 0.163  | 0.00       | 0.129   | 0.001  | 0.23   | 8000  | 0.003 | 0.116  | 0.005   | 0.000      | 0.001  | 0.356       | 0.01    | 0.341  | 0.321          | 0.002    |
| Chlorophyll A Concentration - Amual Standard Deviation   |   | ł     | ;    |        |       | ;    |        | 0.002  | 0.001 | 0.009  | 0     | 900    |        | 000            | 0.087  | 0.002  | 100        | 0.01    | 0.707  | 0.012  | 6700  |       | 2000   | 0.037   | 800        | 0.021  | 0.002       | 0.56    | 0.005  | 2000           | 0.12     |
| Cloud Fraction - Amund Sumdard Deviation   |   |       |      |        |       | ,    | 026    | 104    | 242   | 1062   | 1213  | 100    | 1043   |                | 1063   | 102    | 0.064      | 075     | 143    | 60.0   | 200   | 0     | 1065   | 100     | 190        | 003    | 1071        | 1048    | 1068   | 105            | 002      |
| Sea los Concentration - Annual Mean  |   |       |      |        |       | 014  | 003 0  | 035    | 800   | 362 0  | 015   | 280    | 336 (  | 108            | 10     | 11     | 280        | 105     | 600    | 908    | 900   | 100   | 22     | 610     |            | 142    | 221 0       | 002     | 002    | 19             | 5 6      |
| NORMACI DIRECTOR (REQUINE - XIER ISPAT   |   |       |      |        | . 80  | 1 0  | 0 1/0  | 820    | 20    | 0 60   | 100   |        | 0.7 0  | 000            | 24     | 033    |            | 0 0     | 88     | 0.4    |       |       | 0.00   | 002     | R a        | 0 610  | 0 10<br>992 | 002 0   | 0.05   | 28             | B00      |
| 1936 Summer - Second - Street  |   |       |      |        | 0 25  |      | 0.0    | V0 80  | 03 0  | 38 0.0 | G8 0. | 18 0.4 | 33 0.0 | 10 50          | 0.0    | 58 0.1 | 000        | 86 0.4  | 78 0.0 | 46 0.0 | 10 07 | 15    | 83 0.4 | 239 0.1 | 00 00      | 13 0.4 | 0 0'        | 78 0.4  | 29 0.0 | 32 0.          | 00 00    |
| not dimot - notempt repro-   |   | ľ     | 1    | 1 2 0  | 000 0 | 1    | Î      | 0.0    | 0.0   | 7 0.0  | 6 0.0 | 0.0    | 5 0.0  | 80 89<br>0 1 0 | 7 0.0  | 10 9   | 0.0        | 3 0.1   | 0.0 1  | 6 0.1  | 0.0   | 8 0.0 | 1.0    | 3 0.0   | 0.0        | 8 0.0  | 9 0.0       | 1 0.0   | 10 10  | 0.0 8          | 0.0      |
| maM ylrimoM - noisileani   | 1 | ľ     | •    | 3 0.14 | 010   | 0.00 | 2 0    | 4 0.16 | 20.00 | 0.3    | 2 0.2 | 100    | 2 0.21 | S 0.16         | 0.00   | 270 5  | 1010       | \$ 0.26 | 0.00   | 7 0.30 | 0 15  | 0.34  | 2 0.25 | 0.01    | 0.00       | 2 0.08 | 9 0.00      | 0.00    | 8 0.30 | 0770           | 0.00     |
| minited balant? humb minimum. of sal   | 1 | ľ     | 0.06 | 0.09   | 0.37. | 0    | 0.000  | 0.02   | 000   | 0.31   | 0.04  | 0.00   | 0.08   | 0.13           | 000    | 0.16   | 0.12       | 0.15    | 0.02   | 0.00   | 0.02  | •     | 0.19   | -       | 0.04       | 0.14   | 0.000       | 0.00    | 0.00   | 0.32           | 0.00     |
| Chlorophyll A Concentration - Amust Maste  | 1 | 0.043 | 0.01 | 0.00   | 0.000 | 0.01 | 0.488  | 0000   | 0000  | 0.031  | 0.00  | 0.044  | 0.01   | 1000           | 0.008  | 0      | 0.13       | •       | 0.614  | 0.008  | 200   | •     | 000    | 0.081   | 0.134      | 0.28   | 2010        | 0.032   | 0.008  | 0.016          | 0.150    |
|  |   |       |      |        |       |      | uopes. |        |       |        |       |        |        |                |        |        |            |         |        |        |       |       |        |         |            |        |             |         |        |                |          |

| Chlorophyll A Concentration - Annual Mean<br>Sea lee Concentration - Annual Sandard Deviation | usolation - Monthly Mean | Doygen Saturation - Monthly Mean | bust Flux - Annual Standard Deviation | ea lee Concentration - Annual Mean | Cloud Fraction - Annual Standard Deviation | (histophyll A Concentration - Annual Standard Dev<br>verseline Denti - Annual Standard Deviation | Scen Temperature - Monthly Mean | Nust Flux - Monthly Mean | Nasolved Oxygen - Monthly Maan | boou risciou - Annuai Astan<br>boobse Carcentation - Annual Sandard Deviatio | Neur Temperature - Annual Standard Devistion | nsolation - Armal Standard Deviation | Apparent Oxygen Utilization - Monthly Mean | but Flor - Annual Mean | eitrate Concentration - Annual Mean | H - Antual Mean | alinity - Annual Sandard Deviation | atrate Concentration - Monthly Mean | Thiotophyll A Concentration - Annual Range | nemotine Depth - Annual Standard Deviation<br>hemotine Denti - Monthly Mean | Distance from Land - NA | ütrate Concentration - Annual Sandard Deviation | Xey Longth - Monthly Mean | hosphate Concentration - Monthly Mean | Tosphate Concentration - Annual Mean<br>Veryo Tamoacoura - Annual Processi | bean Depth - NA | atrate-Phosphate Ratio - Monthly Mean | alinity - Arnual Mean | reolation - Annual Mean | Xituse Autoustion - Annual Mean | Churchell A Constraint Marikh Mari | venotine Depth - Monthly Mean | Soud Fraction - Monthly Mean | bean Temperature - Annual Mean | alinky - Meathly Mean | calcite Concentration - Annual Mean |  |
|---|--------------------------|----------------------------------|---------------------------------------|------------------------------------|--|--|---------------------------------|--------------------------|--------------------------------|--|--|--------------------------------------|--|------------------------|-------------------------------------|-----------------|------------------------------------|-------------------------------------|--|---|-------------------------|---|---------------------------|---------------------------------------|--|-----------------|---------------------------------------|-----------------------|-------------------------|---------------------------------|------------------------------------|-------------------------------|------------------------------|--------------------------------|-----------------------|-------------------------------------|--|
|---|--------------------------|----------------------------------|---------------------------------------|------------------------------------|--|--|---------------------------------|--------------------------|--------------------------------|--|--|--------------------------------------|--|------------------------|-------------------------------------|-----------------|------------------------------------|-------------------------------------|--|---|-------------------------|---|---------------------------|---------------------------------------|--|-----------------|---------------------------------------|-----------------------|-------------------------|---------------------------------|------------------------------------|-------------------------------|------------------------------|--------------------------------|-----------------------|-------------------------------------|--|

| VARIABLE                    | METRIC                    | GLOBALLY<br>REPRESENTATIVE | GLOBALLY<br>REPRESENTATIVE | GLOBALLY<br>REPRESENTATIVE | PROPORTION OF AREA |
|-----------------------------|---------------------------|----------------------------|----------------------------|----------------------------|--------------------|
|                             |                           | SAMPLE K-S                 | SAMPLE SIZE                | SAMPLE P-VALUE             | WITH MESS <-20     |
| Apparent Oxygen Utilization | Monthly Mean              | 0.2117                     | 234                        | < 0.001                    | 0.0002             |
| Calcite Concentration       | Annual Mean               | 0.4456                     | 234                        | < 0.001                    | 0.0006             |
| Chlorophyll A Concentration | Annual Mean               | 0.4804                     | 234                        | < 0.001                    | 0.0002             |
| Chlorophyll A Concentration | Annual Range              | 0.4666                     | 234                        | < 0.001                    | 0.0005             |
| Chlorophyll A Concentration | Annual Standard Deviation | 0.1832                     | 91                         | < 0.005                    | 0.0003             |
| Chlorophyll A Concentration | Monthly Mean              | 0.5136                     | 229                        | < 0.001                    | 0.0001             |
| Cloud Fraction              | Annual Mean               | 0.3181                     | 234                        | < 0.001                    | 0                  |
| Cloud Fraction              | Annual Standard Deviation | 0.2363                     | 234                        | < 0.001                    | 0.0275             |
| Cloud Fraction              | Monthly Mean              | 0.3203                     | 234                        | < 0.001                    | 0                  |
| Day Length                  | Monthly Mean              | 0.2495                     | 234                        | < 0.001                    | 0                  |
| Diffuse Attenuation         | Annual Mean               | 0.5183                     | 234                        | < 0.001                    | 0.0001             |
| Dissolved Oxygen            | Monthly Mean              | 0.273                      | 234                        | < 0.001                    | 0                  |
| Distance from Land          | NA                        | 0.6947                     | 234                        | < 0.001                    | 0.1099             |
| Dust Flux                   | Annual Mean               | 0.4569                     | 234                        | < 0.001                    | 0.0051             |
| Dust Flux                   | Annual Standard Deviation | 0.0365                     | 234                        | NS                         | 0.0118             |
| Dust Flux                   | Monthly Mean              | 0.0962                     | 234                        | NS                         | 0.0093             |
| Irradiance                  | Annual Mean               | 0.3651                     | 234                        | < 0.001                    | 0                  |
| Irradiance                  | Annual Standard Deviation | 0.4599                     | 234                        | < 0.001                    | 0.0002             |
| Irradiance                  | Monthly Mean              | 0.0797                     | 234                        | NS                         | 0                  |
| Latitude                    | NA                        | 0.587                      | 234                        | < 0.001                    | 0                  |
| Nitrate Concentration       | Annual Mean               | 0.2695                     | 234                        | < 0.001                    | 0                  |
| Nitrate Concentration       | Annual Standard Deviation | 0.4436                     | 234                        | < 0.001                    | 0.005              |
| Nitrate Concentration       | Monthly Mean              | 0.103                      | 234                        | < 0.05                     | 0.0003             |
| Nitrate:Phosphate Ratio     | Monthly Mean              | 0.0007                     | 234                        | NS                         | 0.0074             |
| Ocean Depth                 | NA                        | 0.5811                     | 234                        | < 0.001                    | 0.0001             |
| Ocean Temperature           | Annual Mean               | 0.3336                     | 234                        | < 0.001                    | 0                  |
| Ocean Temperature           | Annual Range              | 0.4911                     | 234                        | < 0.001                    | 0.0006             |
| Ocean Temperature           | Annual Standard Deviation | 0.4549                     | 234                        | < 0.001                    | 0.0024             |
| Ocean Temperature           | Monthly Mean              | 0.3045                     | 234                        | < 0.001                    | 0                  |
| Oxygen Saturation           | Monthly Mean              | 0.1761                     | 234                        | < 0.001                    | 0.0002             |
| pH                          | Annual Mean               | 0.1759                     | 234                        | < 0.001                    | 0.0161             |
| Phosphate Concentration     | Annual Mean               | 0.1954                     | 234                        | < 0.001                    | 0                  |
| Phosphate Concentration     | Annual Standard Deviation | 0.3478                     | 234                        | < 0.001                    | 0.0123             |
| Phosphate Concentration     | Monthly Mean              | 0.1037                     | 234                        | < 0.05                     | 0.0004             |
| Proximity to Thermocline    | Monthly Mean              | 0.3355                     | 234                        | < 0.001                    | 0                  |
| Pycnocline Depth            | Annual Standard Deviation | 0.1627                     | 221                        | < 0.001                    | 0.047              |
| Pycnocline Depth            | Monthly Mean              | 0.2037                     | 234                        | < 0.001                    | 0.0372             |
| Salinity                    | Annual Mean               | 0.2895                     | 234                        | < 0.001                    | 0                  |
| Salinity                    | Annual Standard Deviation | 0.1074                     | 234                        | NS                         | 0.0045             |
| Salinity                    | Monthly Mean              | 0.2389                     | 234                        | < 0.001                    | 0                  |
| Sea Ice Concentration       | Annual Mean               | 0.0804                     | 234                        | NS                         | 0.0001             |
| Sea Ice Concentration       | Annual Standard Deviation | 0.073                      | 234                        | NS                         | 0                  |
| Sea Ice Concentration       | Monthly Mean              | 0.0353                     | 234                        | NS                         | 0                  |
| Thermocline Depth           | Annual Standard Deviation | 0.2955                     | 221                        | < 0.001                    | 0.0346             |
| Thermocline Depth           | Monthly Mean              | 0.2814                     | 234                        | <0.001                     | 0.0192             |

### Table S4: Environmental covariate coverage

| TOPIC ADDRESSED                      | Primary analysis | Effects of classification method on diversity patterns | Effects of modeling method on diversity patterns | Effects of sampling depth on diversity patterns | Effects of diversity metric on diversity patterns | Phylum-level diversity patterns | Range map                         | Range map          | Range map          | Range map          |
|--------------------------------------|------------------|--|--|---|---|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|-----------------------------------|--------------------|--------------------|--------------------|
| TAXONOMIC SCOPE                      | Bacteria         | Bacteria   | Bacteria   | Bacteria  | Bacteria  | Cyanobacteria                   | Alphaproteobacteria             | Actinobacteria                  | Gammaproteobacteria             | Bacteroidetes                   | Verrucomicrobia                 | Prochlorococcus/<br>Synechococcus | Pelagibacter       | Polaribacter       | Sphingopyxis       |
| DIVERSITY<br>MEASURE                 | Richness         | Richness   | Richness   | Richness  | Shannon   | Richness                        | Richness                        | Richness                        | Richness                        | Richness                        | Richness                        | Relative Abundance                | Relative Abundance | Relative Abundance | Relative Abundance |
| RAREFACTION<br>DEPTH                 | 4266             | 4266   | 4266   | 150   | 4266  | 4266                            | 4266                            | 4266                            | 4266                            | 4266                            | 4266                            | 4266                              | 4266               | 4266               | 4266               |
| MODEL                                | Linear           | Linear   | MARS   | Linear  | Linear  | Linear                          | Linear                          | Linear                          | Linear                          | Linear                          | Linear                          | Logistic                          | Logistic           | Logistic           | Logistic           |
| SEQUENCE<br>CLASSIFICATION<br>METHOD | De Novo          | RDP  | De Novo  | De Novo   | De Novo   | De Novo                         | De Novo                         | De Novo                         | De Novo                         | De Novo                         | De Novo                         | RDP                               | RDP                | RDP                | RDP                |
| ANALYSIS                             | I                | II   | III  | IV  | ^   | VI                              | VII                             | VIII                            | IX                              | Х                               | IX                              | XII                               | XIII               | XIV                | XV                 |

## Table S5: Modeling approaches

# Table S6: Best models, Analysis I

| NUMBER OF<br>PARAMETERS   | STATISTIC<br>MINIMIZED  | COVARIATES <sup>1</sup>   |
|---|---|---|
| 0   | CV R <sup>2</sup>   | -   |
| 0   | AIC/BIC   |   |
| 1   | CV R <sup>2</sup>   | DaylengthSq   |
| 1   | AIC/BIC   | DaylengthSq   |
| 2   | CV R <sup>2</sup>   | DaylengthSq, PhosphateAnnualSq  |
| 2   | AIC/BIC   | DaylengthSq, PhosphateAnnualSq  |
| 3   | CV R <sup>2</sup>   | DaylengthSq, PhosphateAnnualSq, ThermoclineDistance   |
| 3   | AIC/BIC   | DaylengthSq, PhosphateAnnualSq, ThermoclineDistance   |
| 4   | CV R <sup>2</sup>   | DaylengthSq, PhosphateAnnualSq, ThermoclineDistance, ChlorophyllRange   |
| 4   | AIC/BIC   | DaylengthSq, PhosphateAnnualSq, ThermoclineDistance, ChlorophyllRange   |
| 5   | CV R <sup>2</sup>   | PhosphateAnnualSq, ThermoclineDistance, TempRange, Insolation, InsolationStDev  |
| 5   | AIC/BIC   | PhosphateAnnualSq, ThermoclineDistance, TempRange, Insolation, InsolationStDev  |
| 6   | CV R <sup>2</sup>   | PhosphateAnnual, ThermoclineDistance, Insolation, InsolationAnnual, ChlorophyllRange, NitrateAnnual   |
| 6   | AIC/BIC   | PhosphateAnnual, ThermoclineDistance, Insolation, InsolationAnnual, ChlorophyllRange, NitrateAnnual   |
| 7   | CV R <sup>2</sup>   | PhosphateAnnual, ThermoclineDistance, Insolation, InsolationAnnual, ChlorophyllRange, NitrateAnnual, Salinity   |
| 7   | AIC/BIC   | PhosphateAnnual, ThermoclineDistance, Insolation, InsolationAnnual, ChlorophyllRange, NitrateAnnual, Salinity   |
| 8   | CV R <sup>2</sup>   | PhosphateAnnual, ThermoclineDistance, Insolation, InsolationAnnual, ChlorophyllRange, NitrateAnnual, Salinity, Chlorophyll  |
| 8   | AIC/BIC   | PhosphateAnnual, ThermoclineDistance, Insolation, InsolationAnnual, ChlorophyllRange, NitrateAnnual, Salinity, SalinityStDev  |
| Abbreviations<br>Chlorophyll: n<br>Chlorophyll: n<br>Daylength: me<br>Insolation me<br>Insolation stupe<br>Nitrate: mean n<br>Nitrate: mean n<br>Nitrate: mean n<br>Nitrate: mean n<br>Nitrate: mean n<br>Phosphate: mean<br>PhosphateAnn<br>PhosphateAnn<br>Salinity Stupe:<br>Seaice: sea ice<br>Temp: mean of<br>TempRange: n<br>ChempRange: n<br>ChempRange: n<br>ChempClinpD | nean chlorphyll a cc<br>nge: mean annual ra<br>an daylengtin in mo<br>gquare of mean dayl<br>an insolation in mon<br>lal: mean annual ni<br>v: annual standard<br>dirtate concentration<br>mean annual ni<br>tepth of ocean<br>oxygen utilization i<br>an phosphate concer<br>al: mean annual salir<br>annual standard<br>i mean annual salir<br>annual standard<br>concentration in m<br>cean temperature in<br>nean annual range o<br>tearance: log of disa | ncentration in month sampled<br>ange of chlorophyll a concentration<br>ange of chlorophyll a concentration<br>ength in month sampled<br>beviation of insolation<br>in month sampled<br>e concentration<br>n month sampled<br>nration in month sampled<br>osophate concentration<br>an annual phosphate concentration<br>mpled<br>ity<br>viation of insolation<br>onth sampled<br>f sea surface temperature<br>f sea surface temperature |

#### Table S7: Selected Models

| ANALYSIS  | NUMBER OF<br>PARAMETERS  | COVARIATES <sup>1</sup>  |
|---|--|--|
| I   | 3  | DaylengthSq, PhosphateAnnualSq, ThermoclineDistance  |
| п   | 4  | DaylengthSq, PhosphateAnnualSq, ThermoclineDistance, ChlorophyllRange  |
| IV  | 3  | DaylengthSq, PhosphateAnnualSq, ThermoclineDistance  |
| V   | 3  | DaylengthSq, PhosphateAnnualSq, SalinityStDev  |
| VI  | 5  | ChlorophyllRange, InsolationAnnual, PhosphateAnnual, SalinityStDev, DaylengthSq  |
| VII   | 2  | DaylengthSq, PhosphateAnnual   |
| VIII  | 7  | ChlorophyllRange, Daylength, PhosphateAnnual, Seaice, Temp, OceanDepth, Salinity   |
| IX  | 4  | Insolation, NitrateAnnual, SalinityAnnual, PhosphateAnnualSq   |
| х   | 3  | InsolationAnnual, InsolationStDev, SalinityStDev   |
| XI  | 5  | Daylength, InsolationAnnual, NitrateAnnual, Salinity, PhosphateAnnualSq  |
| XII   | 3  | InsolationAnnual, SalinityAnnual, Temp   |
| XIII  | 3  | Insolation, PhosphateMean, OceanDepth  |
| XIV   | 3  | InsolationAnnual, Temp, DaylengthSq  |
| XV  | 2  | Salinity, TempRange  |
| Chlorophylli<br>Daylength: n<br>DaylengthSq<br>Insolation: n<br>InsolationAn<br>InsolationAn<br>NitrateAnnu<br>OceanDepth:<br>Oxygen: mea<br>Phosphate: m<br>PhosphateAr<br>PhosphateAr<br>Salinity: mea<br>SalinityStDe<br>Sealce: sea is<br>Temp: mean<br>TempRange:<br>Thermocline | tange: mean annual ra<br>sean daylength in mon<br>sean daylength in mon<br>sean taylength in mon<br>sequare of mean dayle<br>ean insolation in mon<br>nual: mean annual ins<br>depth of ocean<br>n oxygen utilization in<br>ean phosphate concen<br>nual: mean annual phu<br>nualSq: square of mea<br>salinity in month san<br>al: mean annual salini<br>v: annual standard dev<br>ee concentration in mo<br>ocean temperature in n<br>mean annual range of<br>Distance: log of distan | nge of chlorophyll a concentration<br>th sampled<br>sength in month sampled<br>olation<br>eviation of insolation<br>n month sampled<br>c concentration<br>i month sampled<br>ration in month sampled<br>sphate concentration<br>n annual phosphate concentration<br>spled<br>ty<br>iation of insolation<br>nth sampled<br>sea surface temperature<br>ce between sampling depth and mean thermocline depth in month sampled |

Supplementary Figures



Figure S1: Sampling locations for data used in constructing maps. Models with zero to eight parameters were fitted using MICROBIS data. Predictive performance of the models was evaluated using both internal measures of model performance (AIC, BIC, and PRESS) and three independent data sets, collected at the locations shown in red, green, and yellow (see Table S1). Analyses were based on 377 samples (234 MICROBIS, 30 GOS, 9 POMMIER2007, 103 FUHRMAN2008) collected from 164 distinct locations.



Figure S2: Maps of global marine bacterial diversity. Sequences are classified using RDP, and diversity is modeled using a linear model (Analysis II; see Table S5). (A) In December, OTU richness peaks in temperate latitudes in the Northern Hemisphere. (B) In June, OTU richness peaks in temperate latitudes in the Southern Hemisphere. Predicted richness during the spring and fall is intermediate, with roughly globally uniform richness near the equinoxes. These predictions hold regardless of the method used to classify reads (see Table S5 and Fig. 1)



Figure S3: Maps of global bacterioplankton diversity. Sequences are classified using *de novo* clustering, and diversity is modeled using a multiple adaptive regression splines model (Analysis III; see Table S5). (A) In December, OTU richness peaks in temperate latitudes in the Northern Hemisphere. (B) In June, OTU richness peaks in temperate latitudes in the Southern Hemisphere. Predicted richness during the spring and fall is intermediate, with roughly globally uniform richness near the equinoxes. These predictions hold regardless of the modeling method (see Table S5 and Fig. 1).



Figure S4: Maps of global marine bacterial diversity. Sequences are classified using *de novo* clustering and a linear model is used (see Table S5). A sub-optimal, sevenpredictor model is used (Table S6 and Fig. S8). (A) In December, OTU richness peaks in temperate latitudes in the Northern Hemisphere. (B) In June, OTU richness peaks in temperate latitudes in the Southern Hemisphere. Predicted richness during the spring and fall is intermediate, with roughly globally uniform richness near the equinoxes. These predictions are robust to the exact predictors used; see the map in Fig. 1, which was generated using the optimal, three-parameter model given in Table S6.



Figure S5: Maps of global marine bacterial diversity. Sequences are classified using *de novo* clustering, and diversity is modeled using a linear model, but sequencing depth is shallow (150 reads; Analysis IV; see Table S5). (A) In December, OTU richness peaks in temperate latitudes in the Northern Hemisphere. (B) In June, OTU richness peaks in temperate latitudes in the Southern Hemisphere. Predicted richness during the spring and fall is intermediate, with roughly globally uniform richness near the equinoxes. These predictions hold regardless of the rarefaction depth (see Table S5 and Fig. 1).



Figure S6: Maps of standard errors of estimated richness for Analysis I.



Figure S7: **MESS plots for Analysis I.** Sequences are classified using *de novo* clustering and a linear model is used (see Table S5). MESS plots give an estimate of where extrapolation, rather than interpolation, is necessary to project an SDM to geographic space. Negative MESS values indicate that at least one predictor variable is outside of the observed range. Overall, the variables used in the selected model were mostly within the observed range, indicating that maps were primarily based on interpolation rather than extrapolation.



Figure S8: Model selection for Analysis I. Sequences are classified using *de novo* clustering and richness is modeled using a linear model (see Table S5). The graphs in (A), (B), (C), respectively, show the best possible cross validation, BIC, and AIC values for the given number of predictors. (D) and (E) show the performance of the best cross validation model when used to predict independent data: POMMIER2007 and FUHRMAN2008, respectively. (F) Rate at which 95% prediction intervals cover observed values for the GOS data, when the model is fit using RDP-classified sequences at a rarefaction depth of 150 sequences. In all cases, the model was fit using MICROBIS data. All of training data measures (A, B, and C) point to a six-predictor model. However, the independent measures indicate that a six-predictor model is overfit, and that a three-predictor model is preferable. We, therefore, implemented the three-predictor model.



Figure S9: Leave-one-out cross validation for Analysis I. Sequences are classified using *de novo* clustering and a linear model is used.



Figure S10: Leave-one-out cross validation for Analysis III. Sequences are classified using *de novo* clustering and a multiple adaptive regression splines model is used.



Figure S11: Leave-one-out cross validation for Analysis II. Sequences are classified using RDP and a linear model is used.



Figure S12: Leave-one-out cross validation for Analysis IV. Sequences are classified *de novo* clustering at shallow sampling depth and a linear model is used.



Figure S13: Leave-one-out cross validation for the Shannon diversity analysis.



Figure S14: Leave-one-out cross validation for models of diversity within phyla (Analyses VI to XI). Panels are for (A) cyanobacteria, (B) verrucomicrobia, (C) gammaproteobacteria, (D) alphaproteobacteria, (E) actinobacteria, and (F) bacteroidetes.



Figure S15: Leave-one-out cross validation for models of relative abundance of select genera (Analyses XII to XV). Panels are for (A) *Pelagibacter*, (B) *Prochorococcus* and *Synechococcus* (grouped together by RDP), (C) *Polaribacter*, and (D) *Sphingopyxis*.



Figure S16: Model selection for Analysis II. Sequences are classified using RDP, and richness is modeled using a linear model (see Table S5). The graphs in (A), (B), (C), respectively, show the best possible cross validation, BIC, and AIC values for the given number of predictors. (D) and (E) Performance of the best cross validation model when used to predict independent data: POMMIER2007 and FUHRMAN2008, respectively. All training data measures (A, B, and C) point to a six-, seven-, or eight-predictor model. However, the independent measures indicate that these models are overfit, and that a three-predictor model is preferable. We therefore implemented the three-predictor model.



Figure S17: Model selection for Analysis IV. Sequences are classified using *de novo* clustering, and richness is modeled using a linear model, and sequencing depth is shallow (150 reads; see Table S5). The graphs in (A), (B), (C), respectively, show the best possible cross validation, BIC, and AIC values for the given number of predictors. (D) and (E) Performance of the best cross-validation model when used to predict independent data: POMMIER2007 and FUHRMAN2008, respectively. (F) Rate at which 95% prediction intervals cover observed values for the GOS data, when the model is fit using RDP-classified sequences. In all cases, the model was fit using MICROBIS data. All training data measures (A, B, and C) point to a six- or seven-predictor model. However, the independent measures indicate that these models are overfit, and that a three-predictor model is preferable. We therefore implemented the three-predictor model.



Figure S18: Model selection for mapping Shannon diversity (Analysis V). The graphs in (A), (B), (C), respectively, show the best possible cross validation, BIC, and AIC values for the given number of predictors. (D) Performance of the best cross validation model when used to predict independent data: POMMIER2007. (The model is fit with MICROBIS data.) All training data measures (A, B, and C) point to a six- or seven-predictor model. However, the independent measure indicates that these models are overfit, and that a three-predictor model is preferable. We therefore implemented the three-predictor model.



Figure S19: Model selection for mapping richness within cyanobacteria (Analysis VI). The graphs in (A), (B), (C), respectively, show the best possible cross validation, BIC, and AIC values for the given number of predictors. (D) Performance of the best cross validation model when used to predict independent data POMMIER2007. (The model was fit using MICROBIS data.) All training data measures (A, B, and C) point to a eight-predictor model. However, the independent measures indicate that an eightpredictor model is overfit, and that a five-predictor model is preferable. We therefore implemented the five-predictor model.



Figure S20: Model selection for mapping richness within alphaproteobacteria (Analysis VII). The graphs in (A), (B), (C), respectively, show the best possible cross validation, BIC, and AIC values for the given number of predictors. (D) Performance of the best cross validation model when used to predict independent data POMMIER2007. (The model was fit using MICROBIS data.) All training data measures (A, B, and C) point to a six-predictor model. However, the independent measures indicate that a six-predictor model is overfit, and that a two-predictor model is preferable. We therefore implemented the two-predictor model.



Figure S21: Model selection for mapping richness within actinobacteria (Analysis VIII). The graphs in (A), (B), (C), respectively, show the best possible cross validation, BIC, and AIC values for the given number of predictors. (D) Performance of the best cross validation model when used to predict independent data POMMIER2007. (The model was fit using MICROBIS data.) All training data measures (A, B, and C) point to a seven-predictor model as does the independent data, so we implemented this model.



Figure S22: Model selection for mapping richness within gammaproteobacteria (Analysis IX). The graphs in (A), (B), (C), respectively, show the best possible cross validation, BIC, and AIC values for the given number of predictors. (D) Performance of the best cross validation model when used to predict independent data POMMIER2007. (The model was fit using MICROBIS data.) All training data measures (A, B, and C) point to a six-predictor model. Although the independent data are uninformative here  $(R^2 < 0.1 \text{ for all models})$ , based on the previous analyses, the six-predictor model is likely overfit. Thus, we implemented the four-predictor model, the largest model such that adding another predictor increases the cross validation  $R^2$  less than 2.5% (see Supplementary Methods).



Figure S23: Model selection for mapping richness within bacteroidetes (Analysis X). The graphs in (A), (B), (C), respectively, show the best possible cross validation, BIC, and AIC values for the given number of predictors. (D) Performance of the best cross validation model when used to predict independent data POMMIER2007. (The model was fit using MICROBIS data.) All training data measures (A, B, and C) point to a five-predictor model. However, the independent measures indicate that a five-predictor model is overfit, and that a three-predictor model is preferable. We therefore implemented the three-predictor model.



D

3 4 5 6 7 8

Figure S24: Model selection for mapping richness within verrucomicrobia (Analysis XI). The graphs in (A), (B), (C), respectively, show the best possible cross validation, BIC, and AIC values for the given number of predictors. (D) Performance of the best cross validation model when used to predict independent data POMMIER2007. (The model was fit using MICROBIS data.) All training data measures (A, B, and C) point to a model with eight or more predictors. However, the independent measures indicate that a this model is overfit, and that a one- or five-predictor model is preferable. Because the five-predictor model had better performance with the training data, we implemented it.



Figure S25: Model selection for mapping the relative abundance of *Prochlorococcus* and *Synechococus* (Analysis XII). The taxa *Prochlorococcus* and *Synechococus* were grouped together by RDP, so we analyzed them in a single analysis. The graphs in (A), (B), (C), respectively, show the best possible cross validation, BIC, and AIC values for the given number of predictors. Although a small marginal improvement could be gained by using a more complex model, we chose a conservative three-predictor model based on the overfitting observed using the independent data in the other analyses (Supplementary Methods).



Figure S26: Model selection for mapping the relative abundance of *Pelagibacter* (Analysis XIII). The graphs in (A), (B), (C), respectively, show the best possible cross validation, BIC, and AIC values for the given number of predictors. Although a small marginal improvement could be gained by using a more complex model, we chose a conservative three-predictor model based on the overfitting observed using the independent data in the other analyses (Supplementary Methods).



Figure S27: Model selection for mapping the relative abundance of *Polaribacter* (Analysis XIV). The graphs in (A), (B), (C), respectively, show the best possible cross validation, BIC, and AIC values for the given number of predictors. Although a small marginal improvement could be gained by using a more complex model, we chose a conservative three-predictor model based on the overfitting observed using the independent data in the other analyses (Supplementary Methods).



Figure S28: Model selection for mapping the relative abundance of *Sphingopyxis* (Analysis XV). The graphs in (A), (B), (C), respectively, show the best possible cross validation, BIC, and AIC values for the given number of predictors. Although a small marginal improvement could be gained by using a more complex model, we chose a conservative two-predictor model based on the overfitting observed using the independent data in the other analyses (Supplementary Methods).



Figure S29: Maps of global marine bacterial Shannon diversity. Sequences are classified using *de novo* clustering, and diversity is modeled using a linear model (Analysis V; see Table S5). Shannon diversity is high when all OTUs are equally prevalent, and low when only a few OTUs dominate the community. (A) In December, Shannon diversity peaks in temperate latitudes in the Northern Hemisphere. (B) In June, Shannon diversity during the spring and fall is intermediate, with roughly globally uniform richness near the equinoxes. Predicted Shannon diversity patterns are qualitatively similar to predicted richness patterns (Fig. 1), and suggest that summertime blooms may contribute to global diversity patterns when measured by certain parameters.

Figure S30: Latitudinal gradient in observed data. OTU Richness (log scale) is plotted versus latitude for MICROBIS samples from each of four seasons. A strong latitudinal gradient is observed in the winter months, which generally reverses in summer. These plots do not account for distance from thermocline and other variables which affect OTU Richness and which were included in our SDM predictions. Their influence accounts for the fact that we predict a stronger reversed gradient in summer than is apparent in these plots.





Figure S31: Sampling bias in marine surveys. The latitude of sampling locations from MICROBIS and the three validation data sets are plotted against the day length on the day of the survey. The range of daylengths at each latitude are shown in grey. (A) MICROBIS has more samples in the Northern versus Southern hemispheres, but nearly covers the range of possible latitude and daylength combinations. Specifically, MICROBIS sampled high latitudes in both winter and summer. (B) In contrast, GOS sampled only low latitudes, and POMMIER and FUHRMAN did not sample high latitudes in winter.



Figure S32: Independent model validation. In (A) and (B), a linear model was fit using MICROBIS sequences classified with a *de novo* clustering (Analysis I; see Table S5). This model was then used to predict observations in two independent data sets; "FUHRMAN2008" and "POMMIER2007" (Fuhrman et al., 2008; Pommier et al., 2007). In (A), the model fit with MICROBIS had good predictive power, similar to the predictive power to models fit using FUHRMAN2008 in the original publication (Fuhrman et al., 2008) ( $R^2 = 0.2$ ). In (B) the model fit with MICROBIS has excellent predictive power. Thus, the model that we used has good predictive power, despite the independent data sets being collected at different locations at different times, using different methodology and with different sequencing depths S1. The model had relatively little predictive power for the Global Ocean Survey data, but (C) 95% prediction intervals covered observed richness values 96.8% of the time when the model is fit using RDP-classified MICROBIS sequences at a rarefaction depth of 150 sequences.



Figure S33: **Range maps of representative genera.** This is a color version of Figure 2 in the text.



Figure S34: **Patterns of OTU richness within bacterial phyla.** Columns show maps for different phyla; rows for different seasons. Maps are for phyla not shown in Fig. 3; see caption for a complete description.



Figure S35: **Patterns of OTU richness within bacterial phyla.** This is a color version of Figure 3 in the text.



Figure S36: Bacterial diversity hotspots and sea surface temperature. (a) Hotspots of marine bacterial richness overlaid on a map of sea surface temperature (SST) increases, quantified by comparing weekly SST values between the late 1980s and the early 2000s (Halpern et al., 2008). Hotspots are outlined with black borders, and are defined as the 10% of ocean surface with the greatest diversity in December and June (primarily in the Northern and Southern hemispheres, respectively). (b) The distribution of SST increases across the entire ocean and within December and June diversity hotspots.



Figure S37: Bacterial diversity hotspots and acidification. (a) Hotspots of marine bacterial richness overlaid on a map of ocean acidification, measured as the change in aragonite saturation state (ASS) (Halpern et al., 2008). Hotspots are outlined with black borders, and are defined as the 10% of ocean surface with the greatest diversity in December and June (primarily in the Northern and Southern hemispheres, respectively). (b) The distribution of ASS values across the entire ocean and within December and June diversity hotspots.



Figure S38: Bacterial diversity hotspots and ultraviolet radiation. (a) Hotspots of marine bacterial richness overlaid on a map of ultraviolet (UV) radiation anomalies, measured as the rate of deviations greater than one standard deviation from the average Joules per square meter per month (Halpern et al., 2008). Hotspots are outlined with black borders, and are defined as the 10% of ocean surface with the greatest diversity in December and June (primarily in the Northern and Southern hemispheres, respectively). (b) The distribution of UV anomaly values across the entire ocean and within December and June diversity hotspots.