Supplementary File S1: Additional details on 16S metabarcoding

Nucleic acid extraction

For algal samples, approximately 100 mg (wet weight) of algal material were ground in liquid nitrogen and DNA extraction was performed with the Nucleospin plant II mini kit (Macherey-Nagel, Düren, Germany) according to the manufacturer's instructions. DNA from bacteria found in the culture medium was extracted using a protocol following Barbeyron *et al.* (1984). Fifty ml of algal medium were filtered through 40 µm sterile cell strainers and centrifuged for 20 min at 4°C and 10,000g. Bacterial pellets were then re-suspended in 150 µL of 50mM Tris-HCl (pH8) containing 25% (w/v) sucrose. After addition of 30 µL of lysis buffer (50mM Tris-HCl pH 8, 5mM EDTA, 15 g/L lysozyme, 0.5 g/L RNAse A) samples were incubated for 10 min at RT. Then 40 µL of 400 mM EDTA (pH 8) were added, and the samples incubated for 10 min on ice. Next, 150 µL of aqueous SDS (25% w/v) and 250 µL of proteinase solution (50 mM Tris-HCl pH 8, 50 mM EDTA, 100 mM NaCl, 80mg/L proteinase K) were added. After 1 h of incubation at 37°C, proteins were removed from the reaction by extraction with 1 volume of phenol/chloroform (1:1 v:v, pH 8), and remaining phenol was removed by extraction with only chloroform. Then, DNA was precipitated with 0.8 volumes of isopropanol followed by a washing step with 70% ice-cold ethanol. Finally, pellets were dried at RT for 10 min, and re-suspended in 50 µL of TE (10mM Tris-HCl pH 8, 1mM EDTA).

Primer design

Suitable primer pairs were selected from an *in silico* assay (Wang and Qian 2009), according to their predicted coverage, the length of the predicted amplicon, and the taxonomic resolution to be expected from the variable regions to be amplified (Chakravorty *et al.* 2007). Promising candidate pairs were then re-tested using TestPrime 1.0 (Klindworth *et al.* 2013) and the ARB Silva SSU NR database 114, and blasted against the *Ectocarpus siliculosus* nuclear as well as chloroplast and mitochondrial genomes. The most promising primer pair 341F and 806R (Table S4), was then manually optimized by generating a central mismatch with the *E. siliculosus* chloroplast and other cyanobacteria, and the new specificity tested once more with TestPrime 1.0. Their performance was similar to that of the primer pair used by Caporaso *et al.* (2011), which was predicted to amplify 94.9% of bacterial sequences with 2 mismatches, compared to 95.0% for our primer pair.

Library preparation and sequencing

Sequencing libraries were constructed directly by PCR following Caporaso *et al*. (2012) as shown in Supplementary Figure S2. A primer comprising the Illumina P5 adapter, a linker sequence, and 341F was used in a PCR reaction together with a primer comprising the Illumina P7 adapter, a linker, an 8 bp index, and the primer 806R. Indexes were chosen randomly from a selection generated by Hamady et al. (2008). All oligonucleotides were ordered from Eurogentec (Seraing, Belgium) without chemical modification of bases; a complete list of primers is available in Table S4. For each sample, two 25 µL PCR reactions with two different polymerases (PFU polymerase, Promega, Madison, USA; and PFX polymerase, LIFE, Carlsbad, USA) were carried out using the forward primer construct 341Fv3 and one

selected reverse construct (806Rcomp001-806Rcomp064) for each sample. For both reactions 0.5 µL of sample (approximately 5 ng) were used as template, and final primer concentrations were 250 nM. For the PFX reactions $MgSO_4$ was added at a final concentration of 2 mM. The following PCR protocols were used for the PFU/PFX polymerases respectively: initial denaturation 2 min at 95°C; 30 cycles of denaturation 45 sec at 95°C, annealing 30 sec at 50°C, extension 75 sec/45 sec at 72°C/68°C, and final extension 5min at 72°C/68°C. Two µL of PCR product were then loaded on a 1.5% agarose gel to verify that a product of ca. 500 bp length had been amplified, replicate reactions for each sample pooled, and pools purified using Agencourt AMPure XP beads (Beckman-Coulter, Brea, USA) according to the manufacturer's instructions. The same protocol was also used for a no template control (the sample was replaced by distilled water) and a positive control (the sample was replaced by DNA from the marine bacterium *Zobellia galactanivorans*), except that for the no template control 45 instead of 30 amplification cycles were used.

The positive control sample was quantified spectrophotometrically using a Nanodrop 2000, and a 10x dilution series of this sample with concentrations ranging from 50 nM to 0.5 pM was used for the relative quantification of the remaining samples by qPCR. All qPCR reactions were carried out on a single 384 well plate on a LightCycler480 (Roche Diagnostics, Basel, Switzerland). Each sample was diluted 1000-fold in two replicate dilutions, and each dilution was quantified twice on the plate. The reaction mix consisted of 5 µL LightCycler480 SYBR Green I Master mix (Roche), 0.5 µL of each of the Illumina P5 and P7 oligonucleotides (Supplementary Table S3; 500 nM final concentration), and 4 µL of diluted sample. The following run protocol was used: initial activation 10 min at 95°C; 35 cycles of 30 sec at 95°C denaturation, 10 sec at 55°C annealing, 45 sec at 60°C extension. Quantification of the samples using the fit-point method implemented in the LightCycler 480 software was then used to prepare an equimolar pool of all samples. Finally, the pooled library was again quantified by qPCR using the KAPA library quantification kit (Clinisciences, Nanterre, France) according to the manufacturer's instructions, and quality-checked on a Bioanalyzer 2100 with a DNA-1000 kit (Agilent, Santa Clara, USA). For sequencing, this library was diluted to 6 pM, spiked with 20% PhiX as internal standard, and analyzed on a MiSeq (Software v2.2.0.2) using a MiSeq Reagent kit V2 (500 cycles) and the sequencing primers listed in Supplementary Table S3. In total, approximately 5 million sequence pairs were obtained.

Sequence processing and OTU calling

An overview of the cleaning procedure is available in Supplementary Figure S4. Low quality bases (q-value <25) were trimmed and sequences < 200 bp length removed using the fastq_quality_trimmer script provided with the FastX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/), and non-paired sequences removed from the dataset using the custom script "get_pairs.py" available on the Roscoff instance of Galaxy (http://abims.sb-roscoff.fr/galaxyproject). Cleaned sequence pairs were then assembled using PANDASeq v2.4 (Masella et al. 2012) with default parameters and requiring a minimal contig length of 300 bp. Assembled sequences were imported to mothur v.1.31.2 (Schloss et al. 2009), which was used to identify unique sequences. A taxonomic reference database was generated by extracting bacterial sequences from the Arb-SILVA SSU-Ref nr database v115, cutting the alignment to the positions amplified by our primers, and removing gap-only positions. Our assembled reads were aligned to this reference in mothur using a kmer size of 8. Only sequences that could be aligned with the

reference at a threshold of 50% were kept for further processing, and any tailing bases were removed. 99.7 % of the 324,578 sequences removed during the alignment corresponded to PhiX reads not identified by the Illumina software. Aligned sequences were subjected to a preclustering step allowing for 1% mismatches (i.e. 4 nt), and chimeric sequences were removed using Uchime and default parameters (Edgar et al. 2011). The remaining sequences were then classified using the mothur implementation of RDP classifier (Wang et al. 2007), the aforementioned reference database, and a bootstrap-cutoff of 70, and organelle sequences were removed from the dataset. Despite the mismatch with the algal chloroplast sequence in the reverse primer, 32% of the obtained sequences corresponded to chloroplasts: 0.1% in the medium samples and 54% in the algal samples. To reduce memory requirements for clustering, we also removed singlet sequences (i.e. sequences that occurred only once in the entire dataset) at this step using the costume script "removeSingletons.jar". OTUs were called at a 97% identity level using the "cluster" command and the average neighbor method. Only OTUs represented at least five times over all of the examined samples were considered for further analyses. Finally, 26 OTUs supported by 2,323 reads corresponding to probable contaminations (i.e. OTUs that were more abundant in our no template control than in other samples) were removed.

Prediction of metagenomes

We used the Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) suite (Langille *et al.* 2013) to obtain an overview of the genomic and metabolic features represented by the bacterial communities in our samples. To associate OTUs with known bacterial genomes pre-calculated in PICRUSt, closed OTU picking against the Greengenes 16S rRNA gene database (version May 2013) was carried out. As this functionality was not implemented in mothur, cleaned and aligned sequences were exported from mothur using the custom script "CountToQIIME.jar" prior to the removal of singletons. Further analyses were then carried out with QIIME 1.7.0 (Caporaso *et al.* 2010) and the "pick_closed_reference_otus.py" script, setting the similarity to 0.97. This allowed assigning 91 % of the sequences to predefined OTUs. The resulting OTU table was normalized (correction for different 16S rRNA gene copy numbers) and used for metagenome inference of KEGG terms using PICRUSt 1.0.

References

Barbeyron T, Kean K, Forterre P. (1984). DNA adenine methylation of GATC sequences appeared recently in the *Escherichia coli* lineage. J. Bacteriol. 160:586–590.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. (2010). QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 7:335–336.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J. 6:1621–1624.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc. Natl. Acad. Sci. U. S. A. 108 Suppl :4516–4522.

Chakravorty S, Helb D, Burday M, Connell N, Alland D. (2007). A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. J. Microbiol. Methods 69:330–339.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. (2011). UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27:2194–2200.

Hamady M, Walker JJ, Harris JK, Gold NJ, Knight R. (2008). Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. Nat. Methods 5:235–237.

Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, et al. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res. 41:e1.

Langille MGI, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, et al. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat. Biotechnol. 31:814–821.

Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD. (2012). PANDAseq: paired-end assembler for illumina sequences. BMC Bioinformatics 13:31.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. 75:7537–41.

Wang Q, Garrity GM, Tiedje JM, Cole JR. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. 73:5261–7.

Wang Y, Qian P-Y. (2009). Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. Field, D, ed. PLoS One 4:e7401.