

Electronic Supplementary Material, Text for “Interaction capacity as a potential driver of community diversity”

Masayuki Ushio

Contents:

1. Thermal cycle profiles
2. Internal standard DNA sequences
3. Summary of sequence quality
4. Rarefaction curve
5. Negative control results
6. Positive control results
7. Total DNA quantification
8. Quantitative PCR (qPCR)
9. Shotgun metagenomic analysis
10. A meta-analysis
11. Potential limitations of the present study
12. Software package versions
13. Supplementary video
14. Core dataset

1. Thermal cycle profiles

Common protocols

The first-round PCR (first PCR) was carried out with a 12- μ l reaction volume containing 6.0 μ l of 2 \times KAPA HiFi HotStart ReadyMix (KAPA Biosystems, Wilmington, WA, USA), 0.7 μ l of forward primer, 0.7 μ l of reverse primer (each primer at 5 μ M used in the reaction; with adaptor and six random bases), 2.6 μ l of MilliQ water, 1.0 μ l of DNA template and 1.0 μ l of internal standard DNA mixture. Thermal cycle profile, primer set and internal standard DNA sequences are provided in following subsections. Triplicate first PCRs were performed, and these replicate products were pooled in order to mitigate the PCR dropouts. The pooled first PCR products were purified using AMPure XP (PCR product : AMPure XP beads = 1:0.8; Beckman Coulter, Brea, California, USA). The pooled, purified, and 10-fold diluted first PCR products were used as templates for the second-round PCR.

The second-round PCR (second PCR) was carried out with a 24- μ l reaction volume containing 12 μ l of 2 \times KAPA HiFi HotStart ReadyMix, 1.4 μ l of each primer (each primer at 5 μ M in the reaction volume), 7.2 μ l of MilliQ water and 2.0 μ l of template. Different combinations of forward and reverse indices were used for different templates (samples) for massively parallel sequencing with MiSeq. The thermal cycle profile after an initial 3 min denaturation at 95°C was as follows (12 cycles): denaturation at 98°C for 20 s; annealing at 68°C for 15 s; and extension at 72°C for 15 s, with a final extension at 72°C for 5 min.

Twenty microliters of the indexed second PCR products were mixed, and the combined library was again purified using AMPure XP (PCR product: AMPure XP beads = 1:0.8). Target-sized DNA of the purified library (*ca.* 440 bp for prokaryote 16S rRNA; *ca.* 320 bp for eukaryote 18S rRNA; *ca.* 470-600 bp for fungal ITS; *ca.* 510 bp for animal COI) was excised using E-Gel SizeSelect (ThermoFisher Scientific, Waltham, MA, USA) (Note that “the target size” includes Illumina P5/P7 adapter, Rd1SP/Rd2SP sequencing primer and sample-specific index sequences of which total length is 137 bp). The double-stranded DNA concentration of the library was quantified using a Qubit dsDNA HS assay kit and a Qubit fluorometer (ThermoFisher Scientific, Waltham, MA, USA). The double-stranded DNA concentration of the library was then adjusted using MilliQ water and the DNA was applied to the MiSeq (Illumina, San Diego, CA, USA). The prokaryote 16S rRNA, eukaryote 18S rRNA, fungal ITS and animal COI libraries were sequenced using MiSeq Reagent Kit V2 for 2 \times 250 bp PE, MiSeq Reagent Kit V2 for 2 \times 150 bp PE, MiSeq Reagent Kit V3 for 2 \times 300 bp PE and MiSeq Reagent Kit V2 for 2 \times 250 bp PE, respectively.

Primer sequences

For the first PCR, taxa-specific univarsal primers were combined with the MiSeq sequencing primers and six random bases (Ns) to improve the quality of MiSeq sequencing. For the second PCR, MiSeq adaptor and sequencing primers were combined with index sequences (eight bases denoted by X in the following table) to identify each sample. See sample metadata for index sequences of each sample.

Table S2. Taxa-specific 1st PCR primers

Target taxa	Primer name	Sequencing primer - NNNNNN - Universal primer
Prokaryote (16S rRNA)	515F	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG NNNNNN GTGYCAGCMGCCGCGGTAA
	806R	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG NNNNNN GGACTACNVGGGTWTCTAAT
Eukaryote (18S rRNA)	Euk_1391f	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG NNNNNN GTACACACCGCCCGTC
	EukBr	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG NNNNNN TGATCCTTCTGCAGGTTACCTAC
Fungi (ITS)	ITS1_F_KYO1	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG NNNNNN CTHGGTCATTTAGAGGAATAA
	ITS_KYO2	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG NNNNNN TTYRCTRGGTTCTTCATC
Animal (mitochondria COI)	mlCOIintF	ACACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNNN GGWACWGGWTGAACWGTWATYCCYCC
	HCO2198	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT NNNNNN TAACTTCAGGGTGACCAAAAAATCA

References for the taxa-specific 1st PCR primers are as follows: 515F-806R (Bates *et al.* 2011; Caporaso *et al.* 2011), Euk_1391f-EukBr (Amaral-Zettler *et al.* 2009), ITS1_F_KYO1-ITS_KYO2 (Toju *et al.* 2012), mlCOIintF (Leray *et al.* 2013) and HCO2198 (Folmer *et al.* 1994).

Table S3. 2nd PCR primers

Primer name		MiSeq adaptor - XXXXXXXX - Sequencing primer	
2nd PCR primer for 16S, 18S and ITS	Forward	AATGATACGGCGACCACCGAGATCTACAC	XXXXXXXXXX
	Reverse	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG	XXXXXXXXXX
2nd PCR primer for COI	Forward	AATGATACGGCGACCACCGAGATCTACAC	XXXXXXXXXX
	Reverse	CAAGCAGAAGACGGCATACGAGAT	XXXXXXXXXX
		GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT

Taxa-specific thermal cycle profiles of the first-round PCR**Prokaryote 16S rRNA (515F-806R)**

The thermal cycle profile after an initial 3 min denaturation at 95°C was as follows (35 cycles): denaturation at 98°C for 20 s; annealing at 60°C for 15 s; and extension at 72°C for 30 s, with a final extension at the same temperature for 5 min.

Eukaryote 18S rRNA (Euk_1391f-EukBr)

The thermal cycle profile after an initial 3 min denaturation at 95°C was as follows (35 cycles): denaturation at 98°C for 20 s; annealing at 62°C for 15 s; and extension at 72°C for 30 s, with a final extension at the same temperature for 5 min.

Fungal ITS (ITS1_F_KYO1-ITS_KYO2)

The thermal cycle profile after an initial 3 min denaturation at 95°C was as follows (35 cycles): denaturation at 98°C for 20 s; annealing at 55°C for 15 s; and extension at 72°C for 30 s, with a final extension at the same temperature for 5 min.

Animal COI (mlCOIintF-HCO2198)

The thermal cycle profile after an initial 3 min denaturation at 95°C was as follows (35 cycles): denaturation at 98°C for 20 s; annealing at 55°C for 15 s; and extension at 72°C for 30 s, with a final extension at the same temperature for 5 min.

2. Internal standard DNA sequences

Five artificially designed and synthesized internal standard DNAs, which are similar but not identical to the corresponding region of any existing target organisms (e.g., the V4 region of prokaryotic 16S rRNA), were included in the library preparation process to estimate the number of DNA copies (Ushio *et al.* 2018; Ushio 2019, i.e., quantitative MiSeq sequencing). They were designed to have the same primer-binding regions as those of known existing sequences and conserved regions in the insert region. Variable regions in the insert region were replaced with random bases so that

no known existing sequences had the same sequences as the standard sequences. The numbers of standard DNA copies were adjusted appropriately to obtain a linear regression line between the copy numbers of the standard DNAs and their sequence reads from each sample. Sequences and the copy numbers of the standard DNAs are listed in the following tables.

Ushio et al. (2018) showed that the use of standard DNAs in MiSeq sequencing provides reasonable estimates of the DNA quantity in environmental samples when analyzing fish environmental DNA. Indeed, the copy numbers of standard DNAs were highly correlated with the number of sequence reads in this study (for most samples, $R^2 > 0.9$; Fig. S3a,b, suggesting that the number of sequence reads is generally proportional with the number of DNA copies in a single sample. However, it should be noted that it corrects neither for sequence-specific amplification efficiency, nor for species-specific DNA extraction efficiency. In other words, the method assumes similar amplification efficiencies across sequences, and similar DNA extraction efficiencies across microbial species, which is apparently not valid for complex environmental samples. Therefore, we need careful interpretations of results obtained using that method.

Table S4. Sequences and the copy numbers of standard DNAs of prokaryote 16S rRNA

Primer name	Sequence
STD_pro1 5,000 copies/ μ l	GTGCCAGCAGCCGCGTAATACGTAGGGGGCAAGCGTTATCCGGAATTACTGGGCGTAAA GGGAAGGTAGGCGGAAGCTGAAGTCATGTGTGAAAACGCCTGGCTCAACTTAGCTCAGGG TCGCTGAAATCGTTGGATTCTCTGTGAACAAGGTCCTCGGAAATTTCTTGCTAGAGCAAT AAGGGCTGGCGGATAAGGCGAACTTTGCGGCGAAGGCACGGAGCATAGGCAGGGCCCTA ACTGGACCGTTAAAGCGTGGGGAGCAAAACAGGATTAGATACCCTGGTAGTCC
STD_pro2 10,000 copies/ μ l	GTGCCAGCAGCCGCGTAATACGTAGGGGGCAAGCGTTATCCGGAATTACTGGGCGTAAA GGGTATGTAGGCGGTAAAGGAAGTTACGAGTGAAATTACAGGGCTCAACCGATAAGTCGT GGCCAAGAGGCGGAGCCTGCAGTAATCGAGAGGTGGGCGGAATCGGGTAAAAAGTCAACCT AAGGAAACAACCTTTCAAAGGAAGCCAGATCGCGAAGGCGGCCTCGTAGTTAGGTCCTTCT CTGGACAAACGAAAGCGTGGGGAGCAAAACAGGATTAGATACCCTGGTAGTCC
STD_pro3 25,000 copies/ μ l	GTGCCAGCAGCCGCGTAATACGTAGGGGGCAAGCGTTATCCGGAATTACTGGGCGTAAA GGGTAGTAGGCGGGCAATTAAGTTATAGGTGAAAAGTAATGGCTCAACTTCGCGAACTC CGGACGCAGAAAGGTGCTGTGGCGTTAAAGGGGCTCCCGAACCAGCAACAAACCTGA AAGCCGTTCCGATGTTTCGGAACGTCTAATGCGAAGGCTCAAGTCCTAGTAAGTGGCGAT GTCCTAATCACAAGCGTGGGGAGCAAAACAGGATTAGATACCCTGGTAGTCC
STD_pro4 50,000 copies/ μ l	GTGCCAGCAGCCGCGTAATACGTAGGGGGCAAGCGTTATCCGGAATTACTGGGCGTAAA GGGAGTGTAGGCGGCCACGTAAGTTCCCGGTGAAATCGAGCGGCTCAACGTGCTCCTCGC CCGAGGTATCTAAGCCTTGTAATCAGCCTAGGAACCCGAAAGACTTTGTAGACATGCT AAAACGTGACCTGGTGTATGAATATCTCCAGCGAAGGCCGCACAACCGCACTCCACGCTT TAGTGGGTTTAAAGCGTGGGGAGCAAAACAGGATTAGATACCCTGGTAGTCC
STD_pro5 100,000 copies/ μ l	GTGCCAGCAGCCGCGTAATACGTAGGGGGCAAGCGTTATCCGGAATTACTGGGCGTAAA GGGCCGGTAGGCGGCTCCCGAAGTCGGTAGTGAAATTCTGAGGCTCAACTAAAACCACGA CATTGGAACGAGGTGTGTTAGCTTGATACCCGGCCAGTGGAACAGTGCTGCTCACTTCAT AACTCGGGTACGAAATAAGGAAGACGTCCCGCGAAGGCTTTGGCAGGGCATGCCAGGGCA TGGCTTTTGTAAAGCGTGGGGAGCAAAACAGGATTAGATACCCTGGTAGTCC

Table S5. Sequences and the copy numbers of standard DNAs of eukaryotic 18S rRNA

Primer name	Sequence
STD_euk1 10,000 copies/ μ l	GTACACACCGCCCGTCGCTCCTACCGATTGTGTGTGCGGGTGAAGGGACCGGATAGGTACT TGGGGTTGTTTCCCAAGCTATTAAGTAGAGAACTTTACTAAACCAGCACACATAGAGGAAG GTGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCAGAAGGATCA
STD_euk2 5,000 copies/ μ l	GTACACACCGCCCGTCGCTCCTACCGATTGTGTGTGCGGGTGAAGCGGCAGGATAGGCCTC GAGGCTCGATTCCAGATATGTCGACAGAGAACTTATCTAAACCGACACAAGTAGAGGAAG GTGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCAGAAGGATCA
STD_euk3 2,500 copies/ μ l	GTACACACCGCCCGTCGCTCCTACCGATTGCTTGTGCCGGTGAAGAGGATGGATGGGTGAG CAGGGGATATTCCCCAGGGCAGAAGGCGAGAACTTTCGTAAACCGTCGCATATAGAGGAAG GTGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCAGAAGGATCA
STD_euk4 1,250 copies/ μ l	GTACACACCGCCCGTCGCTCCTACCGATTGCCTGTGCGGGTGAAGAGGATGGATCGGCACT ACGGTGTGTTCCCGTTTGGCCGGCGGGAGAAATTTAAGTAAACCAGCGCAATTAGAGGAAG GTGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCAGAAGGATCA
STD_euk5 250 copies/ μ l	GTACACACCGCCCGTCGCTCCTACCGATTGTGTGTGCGGGTGAAGGGAAAGGATGGGGACT GCGGCGCTTTTCCCGCAACACTTAAAGAGAAATTCATAAACCTACACAATTAGAGGAAG GTGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCAGAAGGATCA

Table S6. Sequences and the copy numbers of standard DNAs of fungal ITS

Primer name	Sequence
STD_asco1 20 copies/ μ l	CTCGGTCATTTAGAGGAAGTAAAAGTCGTAACAAGGTCTCCGTAGGTGAACCTGCGGAGG GATCATTACCGAGCGAGGGACGGAGATGAAGCCTTGACACTTTGTGTCCGACACGGTTTG CTTCGGGGGCGATTCTGCCGCAAAAGTTGCATTCCCCAAGATATTCGTCAAAACACTGCA TCAACACGTCGGAACATACTGTTAATGTTTCAAAACTTTCAACAACGGATCTCTTGGTTC TGGCATCGATGAAGAACGCAGCGAA
STD_asco2 40 copies/ μ l	CTCGGTCATTTAGAGGAAGTAAAAGTCGTAACAAGGTCTCCGTAGGTGAACCTGCGGAGG GATCATTACCGAGCGAGGGAAAGTCCCCTACTACGTCAGTCTTTGTGTCTAGCAAAGTTG CTTCGGGGGCGACAATGCCGTACTCTCCGATTCCCCTCATGCACGATCAAAACACTGCA CATAGACGTCGGGGTAATCCGGTAATTCTCGAAAACTTTCAACAACGGATCTCTTGGTTC TGGCATCGATGAAGAACGCAGCGAA
STD_asco3 100 copies/ μ l	CTCGGTCATTTAGAGGAAGTAAAAGTCGTAACAAGGTCTCCGTAGGTGAACCTGCGGAGG GATCATTACCGAGCGAGGGAACTTCTGTCCGCGGACCTCCTTTGTGATCCGAAGAGGTTG CTTCGGGGGCGATTTTGCCGGGGTTGACGCATTCCCAGATACTATCTCAAAACACTGCA TGATGACGTCGGTAGAGTCCTAGAACGAAGTAAAACTTTCAACAACGGATCTCTTGGTTC TGGCATCGATGAAGAACGCAGCGAA
STD_asco4 200 copies/ μ l	CTCGGTCATTTAGAGGAAGTAAAAGTCGTAACAAGGTCTCCGTAGGTGAACCTGCGGAGG GATCATTACCGAGCGAGGGAGGTTGTCATGCGTCCCGAACTTTGTGAGTAGTTACTTTTG CTTCGGGGGCGAGTATGCCGAGGCGGAGGCATTCCCCCTAGCCGGGGTCAAAACACTGCA CCTTAACGTCGGGTGAACGACCAAAAGCACTAAAACTTTCAACAACGGATCTCTTGGTTC TGGCATCGATGAAGAACGCAGCGAA
STD_asco5 400 copies/ μ l	CTCGGTCATTTAGAGGAAGTAAAAGTCGTAACAAGGTCTCCGTAGGTGAACCTGCGGAGG GATCATTACCGAGCGAGGGAATTTAACGAGTCTGGGGGCTTTGTGTGCGCTCCGGTTTG CTTCGGGGGCGATGCTGCCGGCGTGACGGCATTCCCAGGTATCGCGTCAAAACACTGCA TATCGACGTCGGAGACCGTCAGTAAAGAGGCAAAACTTTCAACAACGGATCTCTTGGTTC TGGCATCGATGAAGAACGCAGCGAA

Table S7. Sequences and the copy numbers of standard DNAs of animal mitochondrial COI

Primer name	Sequence
STD_mlCOI1 200 copies/ μ l	GGAACAGGATGAACAGTTTACCCTCCACTGTCAACCAGTATTGCTCACAGAGGAGCTTCC GTCGATTTAGCTATTTTTTCTTTACATTTAGCAGGCATCTCATCGATTATAGGGGCAATT AACTTTATTACAACCGTTATTAATATACGAATGTGATGCATAAACGCGGATCGAATACCT TTATTTGTTTGATCAGTCGTAATTACCAGGCTGATTTTTCTATTATCCCTACCCGTGCTA GCCAGAGTGATCTCTATGCTTTCGTCTGATCGTAACTTAAATACATCATTCTACGACCCT GCTGGTGGAGGGGACCCGATTTTGTATCAGCACCTATTTTGATTTTTTGGTCACCCTGAA GTTTA
STD_mlCOI2 100 copies/ μ l	GGAACAGGATGAACAGTTTACCCTCCACTCTCTTCACCAATTGCTCACCGCGGAGCATCA GTGGATTAGCCATTTTTTCGTTTCATTTAGCAGGCATCTCCTCTATTTTAGGTGCATTA AACTTTATTACAACCGTTATTAATATACGATGACCAACAATAGTTAGAGATCGAATACCT TTATTTGTTTGATCAGTAGCTATTACCAGAGTCTTCTTTATATTATCACTACCAGTTCTA GCTAGAGTAATCTCTATGCTTTCGTCTGATCGTAACTTAAATACATCATTCTACGACCCG GCCGGGGAGGAGACCCCATTTTCTATCAGCACCTATTTTGATTTTTTGGTCACCCTGAA GTTTA
STD_mlCOI3 50 copies/ μ l	GGAACAGGATGAACAGTTTACCCTCCACTATCCCCATCTATTGCACACCATGGAGCCTCT GTTGATTAGCAATTTTTTCTTTACATTTAGCCGGATTGTCCTCTATTGTAGGGGCTATA AACTTTATTACAACCGTTATTAATATACGAATGTCCTCTATATAGGTAGATCGAATACCT TTATTTGTCTGATCTGTGGTCATTACGAGAGTGTTCCTAATTCTATCCCTACCCGTGCTA GCCAGAGTGATCTCTATGCTTTCGTCTGATCGTAACTTAAATACATCATTCTACGACCCT GCGGGAGGGGCGACCCAATTTTGTATCAGCACCTATTTTGATTTTTTGGTCACCCTGAA GTTTA
STD_mlCOI4 25 copies/ μ l	GGAACAGGATGAACAGTTTACCCTCCACTTTCACCAAGTATTGCTCACGGGGGAGCCTCT GTCGATTAGCAATTTTTTCTCTTCATTTAGCTGGTATTTTCATCAATTTTAGGCGCACTT AACTTTATTACAACCGTTATTAATATACGAGCACCCGAGATATGCTGTGATCGAATACCT TTATTTGTTTGATCTGTGGAGATTACCAGGGTTATCCTACTTCTATCGCTACCCGTCTTA GCTAGAGTAATCTCTATGCTTTCGTCTGATCGTAACTTAAATACATCATTCTACGACCCG GCAGGTGGTGGGACCCATTTTTTATCAGCACCTATTTTGATTTTTTGGTCACCCTGAA GTTTA
STD_mlCOI5 5 copies/ μ l	GGAACAGGATGAACAGTTTACCCTCCACTTTCCTCTATAATTGCTCACTGAGGAGCATCT GTTGATTAGCCATTTTTTCTCTACATTTAGCGGGTATATCATCTATTCTAGGAGCAGTA AACTTTATTACAACCGTTATTAATATACGAACGGCCTTAATAGTTCCGGATCGAATACCT TTATTTGTCTGATCAGTCGCCATTACGAGCGTCTTTCTTTTATTATCCCTACCCGTCTTA GCGAGAGTAATCTCTATGCTTTCGTCTGATCGTAACTTAAATACATCATTCTACGACCCG GCTGGTGGTGGTACCCGATTTTGTATCAGCACCTATTTTGATTTTTTGGTCACCCTGAA GTTTA

3. Summary of sequence quality

I performed MiSeq run four times to generate sequences from the four metabarcoding regions (i.e., 16S rRNA, 18S rRNA, ITS and COI). Summary of the sequence runs is provided below.

Table S8. Summary of sequence quality

MiSeq run name	Cycles	Target region	Pass filter (%)	>Q30 (%)	Yield (Gbp)	Total reads	Reads PF
RMR-076	500	16S rRNA	95.69	84.77	7.69	15,636,422	14,962,924
CMR-002	300	18S rRNA	89.64	94.86	5.67	20,158,344	18,070,998
RMR-078	600	ITS	95.13	79.73	16.08	27,533,984	26,191,796
RMR-099	500	COI	90.07	89.62	9.45	20,410,736	18,383,068
Total					38.89	83,739,486	77,607,998

Sequence reads from standard DNAs and non-standard DNAs are provided below.

Table S9. Sequence reads from standard DNAs and non-standard DNAs

MiSeq run name	Standard DNAs (all libs)	Non-standard DNAs (all libs)	Standard DNAs (sample only)	Non-standard DNAs (sample only)	Standard / Total
RMR-076	3,259,980	4,750,032	2,516,936	4,718,786	35%
CMR-002	5,391,480	7,437,546	4,485,282	7,394,159	38%
RMR-078	6,180,704	7,732,723	5,186,395	7,710,798	40%
RMR-099	3,343,130	7,693,367	2,759,167	7,472,133	27%
Per sample	26,418	40,136	24,305	44,384	35%

“all libs” indicates that sequence reads of all libraries including true samples, negative and positive controls are included. “sample only” indicates that sequence reads of only true samples are included (see sample metadata provided in <https://github.com/ong8181/interaction-capacity/SupplementaryData>).

4. Rarefaction curve

Rarefying sequence reads is a common approach to evaluate microbial diversity. I examined rarefaction curves and found that the sequencing captured most of the prokaryotic diversity (Fig. S2). Also, a previous study suggested that simply rarefying microbial sequence data is inadmissible (McMurdie & Holmes 2014). Another important point is that my analyses involve conversion of sequence reads to DNA copy numbers and thus are different from other commonly used analyses in microbiome studies. Considering these conditions and the previous study, sequence reads were subjected to most of the downstream analysis in the present study without performing further cor-

reactions using rarefaction or other approaches.

5. Negative control results

Prior to the library preparation, work-spaces and equipment were sterilized. Filtered pipet tips were used, and separation of pre- and post-PCR samples was carried out to safeguard against cross-contamination. To monitor cross-contamination during field sampling and library preparation process, 10 PCR-level negative controls without internal standard DNAs, 22 PCR-level negative controls with internal standard DNAs, and 30 field-level negative controls were employed for each MiSeq run. PCR-level negative controls were made by adding MilliQ water instead of extracted DNA. For field-level negative controls, a 500-ml plastic bottle containing MilliQ water was carried to the field with the other five 500-ml plastic bottles, and taken back to the laboratory after the water sampling. Then, the MilliQ water in the bottle were filtered using the cartridge filter (ϕ 0.22- μ m and ϕ 0.45- μ m). The filter cartridges were treated in an identical way with the other samples during the DNA extraction and library preparation processes. Field-level negative controls were collected consecutively during the first 15 sampling events (from 24 May 2017 to 6 June 2017, and the first pre-monitoring sampling on 22 May 2017). After the first 15 sampling events, field-level negative controls were collected once in a week (i.e., on Monday).

All of the negative controls generated negligible sequences compared with the field samples and positive controls for the four MiSeq runs (Fig. S2), suggesting that there was no serious cross-contamination during the water sampling, DNA extraction and library preparation processes.

6. Positive control results

Positive control samples were taken to monitor for possible degradation during the sample storage. Water samples were collected from Plot 2 on 23 May 2017 before planting rice seedlings. Ten sub-samples were separately filtered using the cartridge filter (ϕ 0.22- μ m and ϕ 0.45- μ m), and stored at -20°C in the same way as the other samples. DNAs of the positive controls were extracted subsequently and analyzed in the same way as the other samples. DNA extraction dates for the positive controls were as follows: 23 May 2017, 27 June 2017, 12 July 2017, 9 August 2017, 22 August 2017, 11 September 2017, 29 September 2017, 11 October 2017, 25 October 2017, and 8 November 2017 (i.e., from an immediate DNA extraction to 5.5 months storage at -20°C).

The 10 most abundant ASVs were extracted and their abundances were visualized in Fig. S2i-l. The most abundant ASVs were robustly detected from most of the positive samples and there were no dramatic reductions in the DNA copy numbers if DNAs were extracted before 3 months after water sampling (red points in the figures). DNAs of the other samples were extracted within 1-2 months after water sampling in the present study, suggesting that there is no serious DNA degradation during the sample storage.

7. Total DNA quantifications

Total DNAs were quantified using a Quant-iT assay kit (Promega, Madison, Wisconsin, USA). Three μ l of each extracted DNA (from ϕ 0.22- μ m Sterivex) was mixed with the fluorescent reagent (fluorescent reagent:buffer solution = 1:399) and DNA concentration was measured following the manufacturer's protocol.

The results were compared with the total DNA copy numbers estimated by the quantitative MiSeq sequencing (qMiSeq) of the four marker regions (i.e., 16S, 18S, ITS and COI). An assumption behind the analysis is that most cellular organisms were captured by sequencing the four marker regions. Fig. S3c shows the correlation between the total DNA concentrations and total DNA copy numbers. The total DNA copy numbers estimated by qMiSeq were correlated with the total DNA concentrations very well, showing the quantitative capacity of qMiSeq.

8. Quantitative PCR (qPCR)

qPCR of the 16S region was performed using the same primer set with qMiSeq (515F-806R primers) (Bates *et al.* 2011; Caporaso *et al.* 2011). Two μl of each extracted DNA (from ϕ 0.22- μm Sterivex) was added to an 8- μl qPCR reaction containing 1 μl of 5 μM 515F primer, 1 μl of 5 μM 806R primer, 5 μl of Platinum SuperFi II PCR Master Mix (ThermoFisher Scientific, Waltham, Massachusetts, USA), 0.5 μl of 20 \times EvaGreen (Biotium, San Francisco, California, USA), and 0.5 μl of MilliQ water. The thermal cycle profile after an initial 30 s denaturation at 98°C was as follows (60 cycles): denaturation at 98°C for 10 s; annealing at 60°C for 10 s; and extension at 72°C for 15 s (fluorescent measured at this step). The measurements were performed in duplicate to mitigate technical errors. The standard dilution series were prepared using STD_pro1 used in qMiSeq.

Fig. S3d shows the relationship between the total 16S copy numbers estimated by qPCR and qMiSeq. Similarly to Fig. S3c, Fig. S3d shows that there is a highly significant positive correlation between the results of qPCR and qMiSeq. The absolute values estimated by qPCR were generally higher than those estimated by qMiSeq. This could be due to non-target amplifications during the qPCR: although 515F-806R primers predominantly amplify prokaryote 16S, the quantification by EvaGreen qPCR may include non-target DNA. A similar trend was also found in the previous study that used a universal fish primer set (Ushio *et al.* 2018). On the other hand, during the library preparation for qMiSeq, only the targeted amplicons were excised using E-Gel SizeSelect. Thus, non-target amplicons which might be included in qPCR quantifications were not included in qMiSeq quantifications. This could be a cause of the difference in the estimated DNA copy numbers between qPCR and qMiSeq. Nonetheless, both methods in general showed consistent results, again showing the quantitative capacity of qMiSeq.

9. Shotgun metagenomic analysis

Shotgun metagenomic analysis was performed for a subset of the samples to check whether and how PCR-based assessments of community composition biased the results. Only four samples, of which diversity was high, were analyzed because much deeper sequencing is necessary for the shotgun metagenomic analysis (e.g., $> \times 100$ deeper sequencing). Briefly, approximately 10–30 ng of total DNA were used as inputs, and Illumina DNA Prep (Illumina, San Diego, CA, USA) was used to prepare the library for the shotgun metagenome. The library was prepared by following the manufacturer's protocol. The double-stranded DNA concentration of the library was then adjusted to 4 nM and the DNA was sequenced on the MiSeq using a MiSeq V2 Reagent kit for 2 \times 250 bp PE (Illumina, San Diego, CA, USA).

In total, 20,601,323 reads (10.3 Gb for 4 samples) were generated (Table S10) and the quality of sequencing was generally high ($>Q30 = 80.4\%$). The low quality reads and adapter sequences were removed using fastp (<https://github.com/OpenGene/fastp>) (Chen *et al.* 2018). After the quality filtering, 14,315,030 sequence reads remained and the filtered sequences were analyzed

using phyloFlash (<http://hrgv.github.io/phyloFlash/>) (Gruber-Vodicka *et al.* 2020). phyloFlash extracted and assembled 16S sequences from the sequence reads. Then, taxonomy and the number of sequence reads assigned to each taxa were summarized in the same pipeline.

Table S10. Sequence results of the shotgun metagenomic analysis

Sample ID	Sampling date	Plot	Sequence reads	Data size
S115	2017/7/14	1	5,197,043	2.6 Gb
S116	2017/7/15	1	4,993,487	2.5 Gb
S565	2017/7/14	5	4,764,166	2.4 Gb
S566	2017/7/15	5	5,646,627	2.8 Gb
total			20,601,323	10.3 Gb

Fig. 3f-i show the comparison between qMiSeq and the shotgun metagenomic analysis. In summary, the numbers of detected taxa were comparable between the two methods: 82-136 by qMiSeq and 99-118 by the shotgun metagenomic analysis. In addition, the community compositions detected by the two methods were also similar. For both methods, Proteobacteria and Bacteroidetes were the two most dominant phyla. Although qMiSeq detected Verrucomicrobia slightly more frequently than the shotgun metagenomic analysis did, the difference was not large. Thus, the shotgun metagenomic analysis also showed that qMiSeq reasonably captured the diversity and composition of the ecological communities in the present study.

10. A meta-analysis

To verify the expectation that community diversity should be explained reasonably well by a nonlinear regression using temperature and total abundance, I compiled published data from various ecosystems. The collected data include two global datasets and four local datasets collected in Japan: (i) global ocean microbes (Sunagawa *et al.* 2015), (ii) global soil microbes (Bahram *et al.* 2018), (iii) fish from a coastal ecosystem (Masuda 2008), (iv) prokaryotes from freshwater lake ecosystems (Okazaki *et al.* 2017), (v) zooplankton from a freshwater lake ecosystem (Sakamoto *et al.* 2018) and (vi) benthic macroinvertebrates from freshwater tributary lagoon ecosystems (Okano *et al.* 2018). Because the influences of temperature and total species abundance/biomass on community diversity (or interaction capacity and connectance) are likely to be nonlinear, I adopted a general additive model (Wood 2004) as follows: $\log(S) \sim s(\log(T)) + s(\log(A))$, where S , T , A and $s()$ indicate species diversity (or OTU diversity), temperature, an index of total species abundance (or biomass) and a smoothing term, respectively. The simple, nonlinear regression using temperature and abundance (or biomass) were separately applied to each data set, which explained biodiversity surprisingly well for the five aquatic data sets (Fig. S10; adjusted $R^2 = 0.453\text{--}0.792$) and somewhat worse for global soil data (Fig. S10; adjusted $R^2 = 0.16$). Connectance may be difficult to predict from abundance in spatially heterogeneous ecosystems such as soil, which may account for the lower predictability of the regression model of the soil data. Also, interaction capacity may be influenced by species identity (e.g., Fig. S8b), and in that case interaction capacity may not be simply a function of temperature and/or abundance. Thus, the accuracy with which diversity can be predicted from temperature and abundance may differ for target organisms and ecosystems.

Nonetheless, biodiversity is surprisingly well explained only by a nonlinear regression using temperature and abundance, suggesting that the interaction capacity hypothesis might be applicable to a wide range of taxa and ecosystems.

In addition, while one piece of evidence supporting the interaction capacity hypothesis is the reasonably accurate predictions of community diversity by nonlinear regression using temperature and abundance, another supporting piece of evidence is the negative relationship between interaction strength and community diversity. To further support the interaction capacity hypothesis, I introduce two empirical results that show the negative relationship between interaction strength and community diversity (Fig. S10g, h). Ratzke et al. (2020) manipulated the interaction strength of microbial community members by modifying medium conditions, and found that the community diversity decreased with an increase of the interaction strength (Fig. S10g). Yamawo et al. (2021) investigated the interaction strength between ant species and a pioneer tree species, and found that increasing the local ant species diversity decreased the ant-plant interaction strength (Fig. S10h). An assumption that a species has a certain interaction capacity is one of the possible explanations for the empirical observations. Nonetheless, evidence presented here are still necessary conditions of the interaction capacity hypothesis, and further testing using independent methods and data is required.

11. Potential limitations of the present study

The interaction capacity hypothesis proposed in the present study seems to have reasonable supports according to the field monitoring, quantitative MiSeq sequencing, EDM, simple mathematical considerations, and meta-analysis. Nonetheless, there may be several potential limitations and I explicitly state them in this section.

First, the number of community types examined in the present study is still small. In particular, comprehensive, quantitative, and multispecies time series were taken from only one freshwater system (the artificial rice plots). Ecological communities with different community members, interaction types, and external driving forces may generate different patterns in community dynamics and interactions. For example, vertebrate species are virtually absent in my system. Because their life history and generation time are substantially different from those of species which were examined in the present study, and because detecting and quantifying interactions between organisms with different body size and time scales may be challenging, including such species may change the pattern. Interaction types (mutualism, competition, and so on) and its diversity (i.e., the ratio of positive and negative interactions in a community) might also matter (Mougi & Kondoh 2012). Differences in external driving forces (e.g., temperature and disturbance) may be another important factor that should be incorporated. Though potential outcomes of the hypothesis were confirmed by several previous studies, further studies, especially empirical ones, are needed to verify the hypothesis.

Second, there are potential experimental limitations in the present study. Because the detections of species relied on DNA sequencing, an implicit assumption of the present study is that the DNA copy number corresponds to each species abundance (note that the conversion from the DNA copy number to species abundance may be species-specific due to differences in amplification efficiencies and/or the copy number of 16S rRNA genes). Although this assumption would be reasonable for microbial species, how (environmental) DNA concentrations are related to the abundance/biomass of macro-organisms is still actively debated. Some studies demonstrated a linear relationship between DNA concentrations and species abundance for some macro-organisms (e.g.,

Takahara *et al.* 2012; Thomsen *et al.* 2012), but how general the positive linear relationship is still unclear. Another potential experimental limitation of the present study is the interval of the time series. As the sampling interval was daily, interactions occurring on much shorter than the daily time scale seems to be elusive in the present study. An automated field sampling system may solve this issue (Yamahara *et al.* 2019; Hansen *et al.* 2020), but at present, such a system is not easily affordable for many field ecologists.

Lastly, there are potential statistical limitations in the present study. The number of data points influences the prediction accuracy of EDM especially when the number of variables included in the S-map becomes large (i.e., the curse of dimensionality). Although this would not cause serious problems in the present study as the number of data points (= 610) is sufficiently large compared to the dimensionality (= optimal E and the number of causal species was smaller than 14 for most species; Figs. S5a, S8a), it is still challenging to reconstruct much higher dimensional interaction networks with fewer data points (Chang *et al.* 2021). We have recently addressed this issue and developments of such statistical tools will help to understand natural, high dimensional ecosystem dynamics (Chang *et al.* 2021). Another important issue is how we can experimentally validate the results of EDM in a natural ecosystem. As described in previous studies (e.g., Sugihara *et al.* 2012), EDM was initially developed for near future predictions and causality detections in a complex, natural system where manipulative experiments are not feasible. However, building an experimental framework to validate results of EDM would be beneficial especially when a target system has a moderate size and the primary object of the analysis is the management and control of the system.

12. Software package versions

Versions of main software packages used in this study are as follows: bcl2fastq v2.18.0.12, Claident v0.2.2019.05.10, cutadapt v2.5, dada2 v1.11.5, phyloseq v1.28.0, rEDM v0.7.5, tidyverse v1.3.0, Rcpp v1.0.3, pforeach v1.3, and glmnet v3.0.1. Detailed information for the packages and analysis environments are available in “00_SessionInfo” folders in Github (<https://github.com/ong8181/interaction-capacity>) and Zenodo (<https://doi.org/10.5281/zenodo.5867264>).

13. Supplementary video

Supplementary Video is available at <https://doi.org/10.6084/m9.figshare.16456179> (citable, figshare version). YouTube version is also available at <https://youtu.be/YzDVPxDxoTM>. The video shows time-varying interaction network of Plot 1. Blue and red arrow indicates positive and negative causal interactions, respectively.

14. Core dataset

Core dataset such as sample metadata, ASV table, and taxa assigned to ASVs is available as CSV file and R objects at https://github.com/ong8181/interaction-capacity/tree/main/data_compiled.

References in the electronic supplementary material, Text

- Amaral-Zettler, L.A., McCliment, E.A., Ducklow, H.W. & Huse, S.M. (2009). A Method for Studying Protistan Diversity Using Massively Parallel Sequencing of V9 Hypervariable Regions of Small-Subunit Ribosomal RNA Genes. *PLOS ONE*, 4, e6372.
- Bahram, M., Hildebrand, F., Forslund, S.K., Anderson, J.L., Soudzilovskaia, N.A., Bodegom, P.M., *et al.* (2018). Structure and function of the global topsoil microbiome. *Nature*, 560, 233–237.
- Bates, S.T., Berg-Lyons, D., Caporaso, J.G., Walters, W.A., Knight, R. & Fierer, N. (2011). Examining the global distribution of dominant archaeal populations in soil. *ISME Journal*, 5, 908–917.
- Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J., *et al.* (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences*, 108, 4516–4522.
- Chang, C.-W., Miki, T., Ushio, M., Ke, P.-J., Lu, H.-P., Shiah, F.-K., *et al.* (2021). Reconstructing large interaction networks from empirical time series data. *Ecology Letters*, 24, 2763–2774.
- Chen, S., Zhou, Y., Chen, Y. & Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34, i884–i890.
- Folmer, O., Black, M., Hoeh, W., Lutz, R. & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, 3, 294–299.
- Gruber-Vodicka, H.R., Seah, B.K.B. & Pruesse, E. (2020). phyloFlash: Rapid Small-Subunit rRNA Profiling and Targeted Assembly from Metagenomes. *mSystems*, 5, e00920–20.
- Hansen, B.K., Jacobsen, M.W., Middelboe, A.L., Preston, C.M., Marin, R., Bekkevold, D., *et al.* (2020). Remote, autonomous real-time monitoring of environmental DNA from commercial fish. *Scientific Reports*, 10, 13272.
- Leray, M., Yang, J.Y., Meyer, C.P., Mills, S.C., Agudelo, N., Ranwez, V., *et al.* (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: Application for characterizing coral reef fish gut contents. *Frontiers in Zoology*, 10, 34.
- Masuda, R. (2008). Seasonal and interannual variation of subtidal fish assemblages in Wakasa Bay with reference to the warming trend in the Sea of Japan. *Environmental Biology of Fishes*, 82, 387–399.
- McMurdie, P.J. & Holmes, S. (2014). Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLOS Computational Biology*, 10, e1003531.
- Mougi, A. & Kondoh, M. (2012). Diversity of interaction types and ecological community stability. *Science*, 337, 349–351.
- Okano, J., Shibata, J., Sakai, Y., Yamaguchi, M., Ohishi, M., Goda, Y., *et al.* (2018). The effect of human activities on benthic macroinvertebrate diversity in tributary lagoons surrounding Lake Biwa. *Limnology*, 19, 199–207.
- Okazaki, Y., Fujinaga, S., Tanaka, A., Kohzu, A., Oyagi, H. & Nakano, S. (2017). Ubiquity and quantitative significance of bacterioplankton lineages inhabiting the oxygenated hypolimnion of deep freshwater lakes. *The ISME Journal*, 11, 2279–2293.

- Ratzke, C., Barrere, J. & Gore, J. (2020). Strength of species interactions determines biodiversity and stability in microbial communities. *Nature Ecology & Evolution*.
- Sakamoto, M., Nagata, T., Hanazato, T., Miyabara, Y., Ha, J.-Y., Park, H.-D., *et al.* (2018). Long-term zooplankton community records (1996–2017) for Lake Suwa (Japan). *Ecological Research*, 33, 1–1.
- Sugihara, G., May, R., Ye, H., Hsieh, C., Deyle, E., Fogarty, M., *et al.* (2012). Detecting causality in complex ecosystems. *Science*, 338, 496–500.
- Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.R., Labadie, K., Salazar, G., *et al.* (2015). Structure and function of the global ocean microbiome. *Science*, 348, 1261359.
- Takahara, T., Minamoto, T., Yamanaka, H., Doi, H. & Kawabata, Z. (2012). Estimation of fish biomass using environmental DNA. *PloS one*, 7, e35868.
- Thomsen, P.F., Kielgast, J.O.S., Iversen, L.L., Wiuf, C., Rasmussen, M., Gilbert, M.T.P., *et al.* (2012). Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology*, 21, 2565–2573.
- Toju, H., Tanabe, A.S., Yamamoto, S. & Sato, H. (2012). High-coverage ITS primers for the DNA-based identification of ascomycetes and basidiomycetes in environmental samples. *PloS one*, 7, e40863.
- Ushio, M. (2019). Use of a filter cartridge combined with intra-cartridge bead-beating improves detection of microbial DNA from water samples. *Methods in Ecology and Evolution*, 10, 1142–1156.
- Ushio, M., Murakami, H., Masuda, R., Sado, T., Miya, M., Sakurai, S., *et al.* (2018). Quantitative monitoring of multispecies fish environmental DNA using high-throughput sequencing. *Metabarcoding and Metagenomics*, 2, e23297.
- Wood, S.N. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association*, 99, 673–686.
- Yamahara, K.M., Preston, C.M., Birch, J., Walz, K., Marin, R.I., Jensen, S., *et al.* (2019). In situ Autonomous Acquisition and Preservation of Marine Environmental DNA Using an Autonomous Underwater Vehicle. *Frontiers in Marine Science*, doi:10.3389/fmars.2019.00373.
- Yamawo, A., Suzuki, N. & Tagawa, J. (2021). Species diversity and biological trait function: Effectiveness of ant–plant mutualism decreases as ant species diversity increases. *Functional Ecology*, 35, 2012–2025.