Electronic Supplementary Material, Methods for "Interaction capacity as a potential driver of community diversity"

Masayuki Ushio

Contents:

Experimental setting

Five artificial rice plots were established using small plastic containers (90 × 90 × 34.5 cm; 216 L total volume; Risu Kogyo, Kagamigahara, Japan) in an experimental field at the Center for Ecological Research, Kyoto University, in Otsu, Japan (34˚ 58′ 18″ N, 135˚ 57′ 33″ E). Sixteen Wagner pots (φ174.6 × φ160.4 × 197.5 mm; AsOne, Osaka, Japan) were filled with commercial soil, and three rice seedlings (var. Hinohikari) were planted in each pot on 23 May 2017 and then harvested on 22 September 2017 (122 days). The rice growth data are being analyzed for different purposes and thus are not shown in this report. The containers (hereafter, "plots") were filled with well water, and the ecological community was monitored by analyzing DNA in the well water (see following subsections).

Field monitoring of the ecological community

To monitor the ecological community, water samples were collected daily from the five rice plots. Approximately 200 ml of water in each rice plot was collected from each of the four corners of the plot using a 500-ml plastic bottle and taken to the laboratory within 30 minutes. Water samples were kept at 4˚C during transport. The water was filtered using Sterivex filter cartridges (Merck Millipore, Darmstadt, Germany). Two types of filter cartridges were used to filter water samples: to detect microorganisms, φ0.22-μm Sterivex (SVGV010RS) filter cartridges that included zirconia beads inside (for degradation of the microbial cell wall) were used [1], and to detect macroorganisms, φ0.45-μm Sterivex (SVHV010RS) filter cartridges were used. Water in each plastic bottle was thoroughly mixed before filtration, and 30 ml and 100 ml aliquots of the water were filtered using φ0.22-μm and φ0.45-μm Sterivex, respectively (slightly adjusted when the filters were clogged). After filtration, 2 ml of RNAlater solution (ThermoFisher Scientific, Waltham, Massachusetts, USA) were added to each filter cartridge to prevent DNA degradation during storage. In total, 1220 water samples (122 days × 2 filter types × 5 plots) were collected during the census term. In addition, 30 field-level negative controls, 32 PCR-level negative controls with or without the internal standard DNAs and 10 positive controls to monitor the potential DNA cross-contamination and degradation during the sample storage, transport, DNA extraction and library preparations were used. Visual inspections of the negative and positive control results indicated no serious DNA contaminations or degradation during analyses (Fig. S2). Detailed information on the negative/positive controls are provided in the electronic supplementary material, Text.

DNA extractions

DNA was extracted using a DNeasy Blood & Tissue kit following a protocol described in my previous study [1]. First, the 2 ml of RNAlater solution in each filter cartridge were removed

from the outlet under vacuum using the QIAvac system (Qiagen, Hilden, Germany), followed by a further wash using 1 ml of MilliQ water. The MilliQ water was also removed from the outlet using the QIAvac. Then, Proteinase K solution (20 µl), PBS (220 µl) and buffer AL (200 µl) were mixed, and 440 µl of the mixture was added to each filter cartridge. The materials on the cartridge filters were subjected to cell lysis by incubating the filters on a rotary shaker (15 rpm; DNA oven HI380R, Kurabo, Osaka, Japan) at 56˚C for 10 min. After cell lysis, filter cartridges were vigorously shaken (with zirconia beads inside the filter cartridges for 0.22-µm cartridge filters) for 180 sec (3200 rpm; VM-96A, AS ONE, Osaka, Japan). The bead-beating process was omitted for 0.45-µm cartridge filters. The incubated and lysed mixture was transferred into a new 2-ml tube from the inlet (not the outlet) of the filter cartridge by centrifugation (3,500 *g* for 1 min). Zirconia beads were removed by collecting the supernatant of the incubated mixture after the centrifugation. The collected DNA was purified using a DNeasy Blood & Tissue kit following the manufacturer's protocol. After the purification, DNA was eluted using 100 µl of the supplied elution buffer. Eluted DNA samples were stored at −20°C until further processing.

Library preparation for metabarcording

Prior to the library preparation, work spaces and equipment were sterilized. Filtered pipet tips were used, and pre-PCR and post-PCR samples were separated to safeguard against cross-contamination. PCR-level negative controls (i.e., with and without internal standard DNAs) were employed for each MiSeq run to monitor contamination during the experiments.

Details of the library preparation process are described in the electronic supplementary material, Text. Briefly, the first-round PCR (first PCR) was carried out with the internal standard DNAs to amplify metabarcoding regions using primers specific to prokaryotes (515F and 806R) [2], eukaryotes (Euk_1391f and EukBr) [3], fungi (ITS1-F-KYO1 and ITS2-KYO2) [4] and animals (mostly invertebrates in the present study) (mlCOIintF and HCO2198) [5,6] (Note that internal standard DNAs were included for each sample). After the purifications of the triplicate 1st PCR products, the second-round PCR (second PCR) was carried out to append indices for different templates (samples) for massively parallel sequencing with MiSeq. Twenty microliters of the indexed second PCR products were mixed, the combined library was purified, and target-sized DNA of the purified library was excised and quantified. 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).

Sequence processing: Amplicon sequence variant (ASV) approach

The raw MiSeq data were converted into FASTQ files using the bcl2fastq program provided by Illumina (bcl2fastq v2.18). The FASTQ files were then demultiplexed using the command implemented in Claident (http://www.claident.org) [7]. I adopted this process rather than using FASTQ files demultiplexed by the Illumina MiSeq default program in order to remove

sequences whose 8-mer index positions included nucleotides with low quality scores (i.e., Q-score < 30).

Demultiplexed FASTQ files were analyzed using the Amplicon Sequence Variant (ASV) method implemented in the DADA2 (v1.11.5) [8] package of R. First, the primers were removed using the external software cutadapt v2.6 [9]. Next, sequences were filtered for quality using the `DADA2::filterAndTrim()` function, and rates were learned using `DADA2::learnErrors()` function (`MAX_CONSIST` option was set as 20). Then, sequences were dereplicated, error-corrected, and merged to produce an ASV-sample matrix. Chimeric sequences were removed using the `DADA2::removeBimeraDenove()` function.

Taxonomic identification was performed for ASVs inferred using DADA2 based on the query-centric auto-k-nearest-neighbor (QCauto) method [7] and subsequent taxonomic assignment with the lowest common ancestor algorithm [10] using "overall_class" and "overall_genus" database and clidentseq, classigntax and clmergeassign commands implemented in Claident v0.2.2019.05.10. I chose this approach because the QCauto method assigns taxa in a more conservative way (i.e., low possibility of false taxa assignment) than other methods. Because the QCauto method requires at least two sequences from a single microbial taxon, only internal standard DNAs were separately identified using BLAST [11].

After the taxa assignment, sequence performance was carefully examined using rarefaction curves, detected reads from PCR and field negative controls (Fig. S2 and the electronic supplementary material, Text). In addition, whether the PCR-based assessments of community diversity were biased was tested by performing shotgun metagenomic analysis of a few representative samples (Fig. S3f–i). Although the lower DNA concentrations and shallow sequencing depth might reduce the detection rate of rare taxa by the quantitative MiSeq sequencing, saturated rarefaction curves (Fig. S2a–d) and the results of shotgun metagenome analysis (Fig. S3f–i) suggested that the quantitative MiSeq sequencing captured most of the diversity in the water samples. Also, the results of PCR and field negative controls suggested that there were low levels of contamination during the monitoring, DNA extractions, library preparations and sequencing (Fig. S2).

Estimations of DNA copy numbers

For all analyses in this subsection, the free statistical environment R 3.6.1 was used [12]. The procedure used to estimate DNA copy numbers consisted of two parts, following previous studies [1,13]: (i) linear regression analysis to examine the relationship between sequence reads and the copy numbers of the internal standard DNAs for each sample (Fig. S3a, b), and (ii) the conversion of sequence reads of non-standard DNAs to estimate the copy numbers using the result of the linear regression for each sample. Linear regressions were used to examine how many sequence reads were generated from one DNA copy through the library preparation process. Note that a linear regression analysis between sequence reads and standard DNAs was performed for each sample and the intercept was set as zero. The regression equation was: MiSeq sequence reads = regression slope × the number of standard DNA copies [/μl]. Most samples show highly significant linear relationship between the copy

numbers and sequence reads of the standard DNAs (Fig. S3a, b), suggesting that the number of sequence reads produced is proportional to the copy number of DNAs within a single sample.

The sequence reads of non-standard DNAs were converted to copy numbers using sample-specific regression slopes estimated using the above regression analysis. The number of non-standard DNA copies was estimated by dividing the number of MiSeq sequence reads by the value of a sample-specific regression slope (i.e., the number of DNA copies = MiSeq sequence reads/regression slope). A previous study demonstrated that these procedures provide a reasonable estimate of DNA copy numbers using high-throughput sequencing [13].

After the conversion to DNA copy number, ASVs with low DNA copy numbers were excluded because their copy numbers are not sufficiently reliable. Also, ASVs with low entropy (information contained in the time series) were excluded because reliable analyses of EDM require a sufficient amount of temporal information in the time series.

Independent validations of the MiSeq sequencing with standard DNAs

The quantitative capacity of the MiSeq sequencing with internal standard DNAs (i.e., the quantitative MiSeq sequencing) is one of important factors that could influence subsequent data analyses. To check the reliability of the quantitative capacity of the method, I performed two independent DNA measurements (fluorescent-based total DNA quantifications and quantitative PCR [qPCR] of the 16S region) and compared the results with those of the quantitative MiSeq sequencing. Brief protocols of the experiments are described below. Details of the total DNA quantification, qPCR and shotgun metagenomic analysis and discussion of the results are provided in the electronic supplementary material, Text.

Total DNAs were quantified using the Quant-iT assay kit (Promega, Madison, Wisconsin, USA). Three µl of each extracted DNA (from φ0.22-µm Sterivex) was mixed with the fluorescent reagent and DNA concentration was measured following the manufactuer's protocol. The results were compared with the total DNA copy numbers estimated by quantitative MiSeq sequencing of four marker regions (i.e., 16S, 18S, ITS and COI) (Fig. S3c). An assumption behind the analysis is that most cellular organisms were captured by sequencing the four marker regions.

qPCR of the 16S region was performed using the same primer set used for the quantitative MiSeq sequencing (515F-806R primers) [2]. Briefly, 2 µl of each extracted DNA (from φ0.22-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 PCR Master Mix (ThermoFisher Scientific, Waltham, Massachusetts, USA), 0.5 µl of 20 × EvaGreen (Biotium, San Francisco, California, USA), and 0.5 µl of H_2O. Sixty cycles of PCR were performed and the fluorescence was measured by LightCycler 480 (Roche, Basel, Switzerland). The total 16S copy numbers measured by qPCR correlated well with those measured by the quantitative MiSeq sequencing (Fig. S3d). In contrast, those measured by qPCR did not show a linear relationship with sequence reads (Fig. S3e).

Furthermore, to check whether and how PCR-based assessments of community

composition biased the results, shotgun-metagenomic analysis was performed for a subset of the samples. Only four samples, of which the community diversity was high, were analyzed because much deeper sequencing is necessary for the shotgun metagenomic analysis. 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 manufacture's protocol. The double-stranded DNA concentration of the library was then adjusted to 4 nM and the DNA was sequenced on MiSeq using a MiSeq V2 Reagent kit for 2 × 250 bp PE (Illumina, San Diego, CA, USA). In total, 20,601,323 reads (10 Gb for 4 samples) were generated. The low quality reads and adapter sequences were removed using fastp [14], and the filtered sequences were analyzed using phyloFlash [15].

Empirical dynamic modeling: Convergent cross mapping (CCM)

The reconstruction of the original dynamics using time-lagged coordinates is known as State Space Reconstruction (SSR) [16,17] and is useful when one wants to understand complex dynamics. Recently developed tools for nonlinear time series analysis called "Empirical Dynamic Modeling (EDM)", which were specifically designed to analyze state-dependent behavior of dynamic systems, are rooted in SSR [18–21]. These methods do not assume any set of equations governing the system, and thus are suitable for analyzing complex systems, for which it is often difficult to make reasonable *a priori* assumptions about their underlying mechanisms. Instead of assuming a set of specific equations, EDM recovers the dynamics directly from time series data, and is thus particularly useful for forecasting ecological time series, which are otherwise often difficult to forecast.

To detect causation between species detected by the DNA analysis, I used convergent cross mapping (CCM) [19,22]. An important consequence of the SSR theorems is that if two variables are part of the same dynamical system, then the reconstructed state spaces of the two variables will topologically represent the same attractor (with a one-to-one mapping between reconstructed attractors). Therefore, it is possible to predict the current state of a variable using time lags of another variable. We can look for the signature of a causal variable in the time series of an effect variable by testing whether there is a correspondence between their reconstructed state spaces (i.e., cross mapping). This cross-map technique can be used to detect causation between variables. Cross-map skill can be evaluated by either a correlation coefficient (ρ), or mean absolute error (MAE) or root mean square error (RMSE) between observed values and predictions by cross mapping.

In the present study, cross mapping from one variable to another was performed using simplex projection [23]. How many time lags are taken in SSR (i.e., optimal embedding dimension; E) is determined by simplex projection using RMSE as an index of forecasting skill. More detailed algorithms about simplex projection and cross mapping can be found in previous reports [19,23].

When the causal relationships between network properties were examined, I considered the interaction time lag between the network properties. This can be done by using "lagged

CCM" [24]. For normal CCM, correspondence between reconstructed state space (i.e., cross-mapping) is checked using the same time point. In other words, information embedded in an effect time series at time t may be used to predict the state of a potential causal time series at time t. This idea can easily be extended to examine time-delayed influence between time series by asking the following question: is it possible to predict the state of a potential causal time series at time $t–tp$ (tp is a time delay) by using information embedded in an effect time series at time t? Ye et al. [24] showed that lagged CCM is effective for determining the effective time delay between variables. In the present study, I examined the time delay of the effects from 0 to 14 days. When examining species interaction in the rice plots, the time delay of the interactions was fixed as –1 in order to avoid extremely large computational costs (i.e., 1197 × 1197 CCMs must be performed for each tp).

The significance of CCM is judged by comparing convergence in the cross-map skill of Fourier surrogates and original time series. More specifically, first, 1000 surrogate time series for one original time series are generated. Surrogate time series were generated so that they conserve seasonality (i.e., `rEDM::make_surrogate_data(method = "seasonal")`; details are in the scripts deposited). Five rice plot replicates were treated as if they were taken in five different years. Second, the convergence of the cross-map skill is calculated for these 1000 surrogate time series and the original time series. Specifically, the convergence of the cross-map skill (measured by ΔRMSE in the present study) is calculated as the cross-map skill at the maximum library length minus that at the minimum library length [19]. Based on consideration of a large number of CCMs among 1197 DNA species, I used $P = 0.005$ as threshold. For CCMs among the network properties, I used $P = 0.05$, a more commonly used threshold.

Empirical dynamic modeling: Multivariate, regularized S-map method

The multivariate S-map (sequential locally weighted global linear map) method allows quantifications of dynamic (i.e., time-varying) interactions [20,25]. Consider a system that has E different interacting variables, and assume that the state space at time t is given by $x(t) = \{x_1(t), x_2(t), \dots, x_E(t)\}$. For each target time point t^*, the S-map method produces a local linear model IS that predicts the future value $x_1(t^*+p)$ from the multivariate reconstructed state space vector $x(t^*)$. That is,

$$\hat{x}_1(t^* + p) = IS_0 + \sum_{j=1}^{E} \boldsymbol{IS}_j \, x_j(t^*)$$

where $\hat{x}_1(t^* + p)$ is a predicted value of x_1 at time t^*+p, and IS_0 is an intercept of the linear model. The linear model is fit to the other vectors in the state space. However, points that are close to the target point, $x(t^*)$, are given greater weighting (i.e., locally weighted linear regression). Note that the model is calculated separately for each time point, t. As recently shown, $\boldsymbol{IS}_j$, the coefficients of the local linear model, are a proxy for the interaction strength between variables[20].

In the present study, all ASVs that have causal influences on a focal species

(determined by CCM) and have non-zero abundance at a target time point were included in the multivariate, regularized S-map. In some cases, the number of causal species exceed the optimal *E*. In that case, I simply added all the detected causal species in the S-map. Nonetheless, in the present study, the number of variables (= the number of causal species for a target species at each time point) included in each S-map model was fewer than 14 for most target species (over 99% of all cases; see Fig. S5a), which allowed rigorous estimations of interaction strengths between species (i.e., the number of data point, 610, is greater than the square of the embedding dimension). In the same way as with simplex projection and CCM, the performance of the multivariate S-map was also measured by RMSE (or a correlation coefficient, ρ) between observed and predicted values by the S-map (i.e., leave-one-out cross validation). In the present study, to reduce the possibility of overestimation and to improve forecasting skill, a regularized version of multivariate S-map was used [26].

Calculations of properties of the interaction network

Properties of the reconstructed interaction network calculated include: ASV diversity, the number of interactions, connectance, mean interaction strength (*IS*) per link, mean interaction capacity, dynamic stability and coefficient of variation (C.V.) in population dynamics. Next, I give the definitions of the properties.

ASV diversity and the number of interactions are the number of ASVs present in a community and the number of interactions among ASVs present in a community, respectively. Connectance, *C*, is defined as $C = N_{link}/S^2$, where *S* and N_{link} indicate the number of species and the number of interactions (links) in a community, respectively. The existence of interactions was defined by significant CCM results. In addition, even if CCM detected significant interactions between two species, the interaction at a certain time point was judged absent if either or both of the species was/were absent at the time point. Mean interaction strength per link, IS_{link}, was calculated as follows:

$$IS_{link} = \sum_{j=1}^{S} \sum_{\substack{i=1 \\ i \neq j}}^{S} \left| IS_{i \to \mathrm{j}} \right| \Big/ N_{link},$$

where $IS_{i \to j}$ indicates an S-map coefficient from *i*th species to *j*th species. Note that I took the absolute value of the S-map coefficient when calculating IS_{link}. Mean interaction capacity, *IC*, was calculated as follows:

$$IC = \left(\sum_{j=1}^{S} \sum_{\substack{i=1 \\ i \neq j}}^{S} \left| IS_{i \to \mathrm{j}} \right| + \sum_{j=1}^{S} \sum_{\substack{i=1 \\ i \neq j}}^{S} \left| IS_{j \to \mathrm{i}} \right| \right) \Big/ S = 2 \times \sum_{j=1}^{S} \sum_{\substack{i=1 \\ i \neq j}}^{S} \left| IS_{i \to j} \right| \Big/ S.$$

Species interaction capacity is defined as the sum of interactions that a single species gives and receives, and mean interaction capacity of a community is the averaged species interaction capacity. Dynamic stability of the community dynamics was calculated as the absolute value of the dominant eigenvalue of the interaction matrix (i.e., local Lyapunov exponent) as described in a previous study [18]. C.V. of the community dynamics at time *t*

was calculated as follows:

$$C.V.(t) = \sum_{i=1}^{S} \frac{\sigma^{i}_{t-3,t+3}}{\mu^{i}_{t-3,t+3}} \Big/ S(t),$$

where $\sigma^{i}_{t-3,t+3}$ and $\mu^{i}_{t-3,t+3}$ indicate the standard deviation and mean value of the abundance of species *i* from time *t*–3 to *t*+3, respectively (i.e., one-week time window). *S*(*t*) is the number of ASVs at time *t*.

Random shuffle surrogate test

To test whether the patterns generated (e.g., in Fig. 3) are statistical artifacts, I did a random shuffle surrogate test. In the test, the original time series were randomly shuffled within a plot using the `rEDM::make_surrogate_shuffle()` function in the rEDM package [21,27] of R. Then, the same number of causal pairs was randomly assigned in a randomly shuffled ecological community. The regularized, multivariate S-map and subsequent analyses of the network properties (all identical to the original analyses) were applied to the randomly shuffled time series.

Meta-analysis of biodiversity, temperature, and abundance

To validate my hypothesis that the diversity is determined by interaction capacity and connectance, and that they are influenced by temperature and total organism 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 [28], (ii) global soil microbes [29], (iii) fish from a coastal ecosystem [30], (iv) prokaryotes from freshwater lake ecosystems [31], (v) zooplankton from a freshwater lake ecosystem [32] and (vi) benthic macroinvertebrates from freshwater tributary lagoon ecosystems [33]. 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 [34] 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 relationships between diversity, temperature and total abundance were analyzed using the model described in the main text. GAM was performed using the "mgcv" package of R [34].

Data analyzed in the meta-analysis were collected from the publications or official websites [28,32], or provided by the authors of the original publications [29–31,33]. Therefore, raw data for the meta-analysis are available from the original publications, or upon reasonable requests to corresponding authors of the original publications. Scripts for the meta-analysis are available in Github (https://github.com/ong8181/interaction-capacity) and Zenodo (https://doi.org/10.5281/zenodo.5867264).

References in the electronic supplementary material, Methods

1. Ushio M. Use of a filter cartridge combined with intra-cartridge bead-beating improves detection of microbial DNA from water samples. Methods Ecol Evol. 2019. doi:10.1111/2041-210X.13204

2. Bates ST, Berg-Lyons D, Caporaso JG, Walters WA, Knight R, Fierer N. Examining the global distribution of dominant archaeal populations in soil. ISME J. 2011;5: 908–917.

3. Amaral-Zettler LA, McCliment EA, Ducklow HW, Huse SM. A Method for Studying Protistan Diversity Using Massively Parallel Sequencing of V9 Hypervariable Regions of Small-Subunit Ribosomal RNA Genes. PLOS ONE. 2009;4: e6372.

4. Toju H, Tanabe AS, Yamamoto S, Sato H. High-coverage ITS primers for the DNA-based identification of ascomycetes and basidiomycetes in environmental samples. PloS One. 2012;7: e40863. doi:10.1371/journal.pone.0040863

5. Leray M, Yang JY, Meyer CP, Mills SC, Agudelo N, Ranwez V, et al. 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. Front Zool. 2013;10: 34. doi:10.1186/1742-9994-10-34

6. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol Mar Biol Biotechnol. 1994;3: 294–299.

7. Tanabe AS, Toju H. Two new computational methods for universal DNA barcoding: a benchmark using barcode sequences of bacteria, archaea, animals, fungi, and land plants. PloS One. 2013;8: e76910. doi:10.1371/journal.pone.0076910

8. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Meth. 2016;13: 581–583.

9. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011;17: 10–12. doi:10.14806/ej.17.1.200

10. Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. Genome Res. 2007;17: 377–86. doi:10.1101/gr.5969107

11. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10: 421. doi:10.1186/1471-2105-10-421

12. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2019. https://www.r-project.org/

13. Ushio M, Murakami H, Masuda R, Sado T, Miya M, Sakurai S, et al. Quantitative monitoring of multispecies fish environmental DNA using high-throughput sequencing. Metabarcoding Metagenomics. 2018;2: e23297.

14. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34: i884–i890. doi:10.1093/bioinformatics/bty560

15. Gruber-Vodicka HR, Seah BKB, Pruesse E. phyloFlash: Rapid Small-Subunit rRNA Profiling and Targeted Assembly from Metagenomes. mSystems. 2020;5: e00920-20. doi:10.1128/mSystems.00920-20

16. Takens F. Detecting strange attractors in turbulence. In: Rand D, Young L-S, editors. Dynamical Systems and Turbulence. 1981. pp. 366–381. doi:10.1007/BFb0091924

17. Deyle ER, Sugihara G. Generalized theorems for nonlinear state space reconstruction. PLOS ONE. 2011;6: e18295. doi:10.1371/journal.pone.0018295

18. Ushio M, Hsieh C, Masuda R, Deyle ER, Ye H, Chang C-W, et al. Fluctuating interaction network and time-varying stability of a natural fish community. Nature. 2018;554: 360–363.

19. Sugihara G, May R, Ye H, Hsieh C, Deyle E, Fogarty M, et al. Detecting causality in complex ecosystems. Science. 2012;338: 496–500. doi:10.1126/science.1227079

20. Deyle ER, May RM, Munch SB, Sugihara G. Tracking and forecasting ecosystem interactions in real time. Proc R Soc B Biol Sci. 2016;283: 20152258. doi:10.1098/rspb.2015.2258

21. Ye H, Beamish RJ, Glaser SM, Grant SCH, Hsieh C-H, Richards LJ, et al. Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling. Proc Natl Acad Sci U S A. 2015;112: E1569–E1576. doi:10.1073/pnas.1417063112

22. Osada Y, Ushio M. R package for Unified Information-theoretic Causality. Zenodo; 2021. doi:10.5281/zenodo.5163234

23. Sugihara G, May RM. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. Nature. 1990;344: 734–41. doi:10.1038/344734a0

24. Ye H, Deyle ER, Gilarranz LJ, Sugihara G. Distinguishing time-delayed causal interactions using convergent cross mapping. Sci Rep. 2015;5: 14750.

25. Sugihara G. Nonlinear Forecasting for the Classification of Natural Time Series. Philos Trans R Soc Math Phys Eng Sci. 1994;348: 477–495. doi:10.1098/rsta.1994.0106

26. Cenci S, Sugihara G, Saavedra S. Regularized S-map for inference and forecasting with

noisy ecological time series. Methods Ecol Evol. 2019;10: 650–660. doi:10.1111/2041-210X.13150

27. Ye H, Clark A, Deyle E, Munch S, Keyes O, Cai J, et al. rEDM: Applications of Empirical Dynamic Modeling from Time Series. 2018. doi:10.5281/zenodo.1294063

28. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, et al. Structure and function of the global ocean microbiome. Science. 2015;348: 1261359. doi:10.1126/science.1261359

29. Bahram M, Hildebrand F, Forslund SK, Anderson JL, Soudzilovskaia NA, Bodegom PM, et al. Structure and function of the global topsoil microbiome. Nature. 2018;560: 233–237. doi:10.1038/s41586-018-0386-6

30. Masuda R. Seasonal and interannual variation of subtidal fish assemblages in Wakasa Bay with reference to the warming trend in the Sea of Japan. Environ Biol Fishes. 2008;82: 387–399.

31. Okazaki Y, Fujinaga S, Tanaka A, Kohzu A, Oyagi H, Nakano S. Ubiquity and quantitative significance of bacterioplankton lineages inhabiting the oxygenated hypolimnion of deep freshwater lakes. ISME J. 2017;11: 2279–2293. doi:10.1038/ismej.2017.89

32. Sakamoto M, Nagata T, Hanazato T, Miyabara Y, Ha J-Y, Park H-D, et al. Long-term zooplankton community records (1996–2017) for Lake Suwa (Japan). Ecol Res. 2018;33: 1–1. doi:10.1007/s11284-017-1528-2

33. Okano J, Shibata J, Sakai Y, Yamaguchi M, Ohishi M, Goda Y, et al. The effect of human activities on benthic macroinvertebrate diversity in tributary lagoons surrounding Lake Biwa. Limnology. 2018;19: 199–207. doi:10.1007/s10201-017-0530-2

34. Wood SN. Stable and efficient multiple smoothing parameter estimation for generalized additive models. J Am Stat Assoc. 2004;99: 673–686.