Supporting Information for "Detecting and validating influential organisms for rice growth: An ecological network approach"

Masayuki Ushio*, Hiroki Saito, Motoyuki Tojo, Atsushi J. Nagano

*Corresponding authors (ong8181@gmail.com)

Contents

Supplementary Materials and Methods

Field experimental setting and rice monitoring in 2017

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) (Figs. 1a and S1a,b). Sixteen Wagner pots (ϕ 174.6-160.4 mm [top-bottom] × 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 containers (hereafter, "plots") were filled with well water.

Daily rice growth was monitored by measuring rice leaf height of target individuals every day using a ruler (the highest leaf heights were measured). We selected four rice individuals at the center of each plot as the target individuals (indicated by the four red points in the right panel of Fig. 1a), and the average height of the four individuals were used as a representative rice height of each plot. Leaf SPAD was also measured every day using a SPAD meter (SPAD-502Plus, KONICA-MINOLTA, Inc., Tokyo, Japan) for the same leaf whose height was measured. Climate variables (temperature, light intensity, and humidity) were monitored using a portable, automated climate logger (Logbee, Chitose Industries, Osaka, Japan). The Logbee data was corrected and refined using the climate station data provided by the Center for Ecological Research, Kyoto University.

Field monitoring of ecological communities in 2017

Detailed protocols to monitor the ecological communities are described in Ushio (2022). 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. 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, ϕ 0.22-μm Sterivex (SVGV010RS) filter cartridges that included zirconia beads inside were used (Ushio 2019), and to detect macroorganisms, ϕ 0.45-μm Sterivex (SVHV010RS) filter cartridges were used. 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.

Quantitative analysis of environmental DNA: Library preparations, sequencing, and sequence data processing

Detailed protocols of the quantitative eDNA analyses are fully described in Ushio (2022) (see https://ndownloader.figstatic.com/files/34067324 and https://ndownloader.figstatic.com/files/34067327), and the protocols below were adopted for the samples collected in 2017 and 2019. Briefly, DNA was extracted and purified using a DNeasy Blood & Tissue kit (Qiagen, Hilden, Germany). After the purification, DNA was eluted using 100 μl of the elution buffer and stored at -20°C until further processing.

A two-step PCR approach was adopted for the library preparation for quantitative MiSeq sequencing. 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) (Bates *et al.* 2011; Caporaso *et al.* 2011), eukaryotes (Euk_1391f and EukBr) (Stoeck *et al.* 2010), fungi (ITS1-F-KYO1 and ITS2-KYO2) (Toju *et al.* 2012) and animals (mlCOIintF and HCO2198) (Folmer *et al.* 1994; Leray *et al.* 2013). The second-round PCR (second PCR) was carried out to append indices for different samples for sequencing with MiSeq. The DNA library was sequenced on the MiSeq (Illumina, San Diego, CA, USA).

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 v0.2.2019.05.10 (http://www.claident.org) (Tanabe & Toju 2013). Demultiplexed FASTQ files were then analyzed using the Amplicon Sequence Variant (ASV) method implemented in "dada2" package (Callahan *et al.* 2016) of R (Team 2022). Taxonomic identification was performed using Claident (Tanabe & Toju 2013).

Estimations of DNA copy numbers and validation of the quantitative capability of the MiSeq sequencing with internal standard DNAs

For all analyses after this subsection, the free statistical environment R was used (Team 2022). The procedure used to estimate DNA copy numbers consisted of two parts, following previous studies (Ushio *et al.* 2018; Ushio 2019). Briefly, we performed (i) linear regression analysis to examine the relationship between sequence reads and the copy numbers of the internal standard DNAs for each sample, and (ii) conversion of sequence reads of non-standard DNAs to estimate the copy numbers using the result of the linear regression for each sample. The regression equation was: MiSeq sequence reads = sample-specific regression slope $\times$ the number of standard DNA copies [/μl]. Then, the estimated copy numbers per μl extracted DNA (copies/μl) were converted to DNA copy numbers per ml water in the rice plot (copies/ml water). The quantitative capacity of this method was thoroughly evaluated by comparing with quantitative PCR, fluorescence-based DNA measurement, and shotgun metagenomic analysis (Ushio *et al.* 2018; Ushio 2022), and the method was shown to have reasonable capacity to quantify DNA. The method has already been shown to be effective for eDNA-based quantitative assessment of ecological community (Tsuji *et al.* 2020; Sato *et al.* 2021).

Detections of potential causal relationships between rice growth and ecological community dynamics

We detected potentially influential organisms for rich growth nonlinear time series analysis based on the quantitative, 1197-species eDNA time-series (Ushio 2022) and rice growth rate measured in 2017. We quantified information flow from eDNA time series to rice growth rate (i.e., a proxy of interaction between an organism and rice growth) by the "unified information-theoretic causality (UIC)" method implemented in the "rUIC" package (Osada & Ushio 2021) of R. UIC tests the statistical clarity of information flow between variables in terms of transfer entropy (TE) (Schreiber 2000; Frenzel & Pompe 2007) computed by nearest neighbor regression based on time-embedded explanatory variables (i.e., cross-mapping: Sugihara *et al.* 2012). In contrast to the standard method used to measure TE, UIC quantifies information flow as follows:

$$TE = \frac{1}{T}\sum_{t=1}^{T} log(\frac{p(y_{t+tp}|x_t, x_{t-\tau}, ..., x_{t-(E-1)\tau}}{p(y_{t+tp}|x_{t-\tau}, x_{t-2\tau}, ..., x_{t-(E-1)\tau})}) \qquad ... [1]$$

where x, y, and z represent a potential causal variable, effect variable, and conditional variable (in our case, climate variables if they are causal factors), respectively. $p(A|B)$ represents conditional probability: the probability of A conditioned on B. t, tp, τ, and E represent the time index, time step, a unit of time lag, and the optimal embedding dimension, respectively. T is the number of times to evaluate TE. For example, if $tp = -1$ in Eqn. [1], UIC tests the causal effect from y_{t-1} to x_t. Optimal E was selected by measuring TE as follows:

$$TE = \frac{1}{T}\sum_{t=1}^{T} log(\frac{p(x_{t+tp}|y_t, x_t, x_{t-\tau}, ..., x_{t-(E-1)\tau}}{p(x_{t+tp}|y_t, x_t, x_{t-\tau}, ..., x_{t-(E_R-1)\tau})}) \qquad ... [2]$$

where E_R ($< E$) is the optimal embedding dimension of lower dimensional models. Significance tests were conducted by bootstrapping data after embedding (the significance levels were set to 0.05). Eqn. [2] is a TE version of simplex projection (Sugihara & May 1990). TE measured according to Eqn. [1] gains the advantage of previous causality tests, i.e., standard TE methods (Schreiber 2000; Runge *et al.* 2012) and convergent cross mapping (CCM) (Sugihara *et al.* 2012), the algorithm of which is explained and implemented in https://github.com/yutakaos/rUIC.

By using UIC, we quantified TE from eDNA time series to rice growth. We standardized the eDNA time series (copies/ml water) and rice growth rates (cm/day) to have zero means and a unit of variance before the analysis. We tested time-lag up to 14 days (i.e., tp was from 0 to -14; effects from up to 14 days ago were considered). We interpret a finding that TE from eDNA time series to rice growth rate is statistically greater than zero to mean that the ecological community member statistically clearly influences rice growth rate.

Field experimental setting in 2019

Based on the results in 2017, we performed a field manipulative experiment. We focused on two potentially influential organisms and prepared three treatments including the control treatment. Nine artificial rice plots were established using smaller plastic containers than those used in 2017 ($42.2 \times 32.0 \times 30.0$ cm; 40.5 L total volume; Sanko, Tokyo, Japan) in the same experimental field at Kyoto University (i.e., the identical location with Plot 3 in 2017). Three Wagner pots were filled with commercial soil, and three rice seedlings (var. Hinohikari) were planted in each pot on 20 May 2019 and then harvested on 20 September 2019 (124 days) (3 treatments $\times$ 3 replicates $\times$ 3 Wanger pots $\times$ 3 rice seedlings = 81 rice seedlings in total). Daily rice growth was monitored by measuring rice leaf height of target individuals every day using a ruler, as in 2017. We selected one rice individual in the middle of each of the Wagner pots as the target individuals (indicated by the three red points in the inlet panel of Fig. 3a; labelled as locations "1", "2", and "3" from east to west). Leaf SPAD was also measured every day using a SPAD meter as in 2017. In 2019, the rice growth was monitored once a week except during the period of the ecological community manipulations when the daily monitoring was performed. The containers (hereafter, "plots 1-9") were filled with well water, and the ecological community was monitored by analyzing eDNA as in 2017. After we harvested rice, we counted the number of fertile and sterile grains and quantified rice yields.

Target organisms in the field manipulation in 2019: *Globiosporangium nunn* and *Chironomus kiiensis*

Field manipulation experiments were performed using two potentially influential organisms, *Globisporangium nunn* (Oomycete) (Kobayashi *et al.* 2010) and *Chironomus kiiensis* (midge species) The reasons why we focused on these two species in 2019 are multifold. First, they can

be relatively easily manipulated; *Globisporangium nunn* had already been isolated and cultivated and *Chironomus kiiensis* is an insect species and relatively large (several mm to cm). Second, the two species were not reported to be pathogens; using pathogen species under field conditions is not recommended in the experimental field because they might have adverse effects on nearby agricultural systems. *Globisporangium nunn* is a potential biocontrol agent and is reported to have some antagonistic activity against pathogens such as *Pythium ultimum* (Kobayashi *et al.* 2010). *Chironomus kiiensis* is a common midge species around the study region (including in rice paddy fields). The larvae are usually about 10 mm and adults are about 5 mm. This species may influence the water quality (e.g., phosphorus concentration) through their feeding activity (Kawai *et al.* 2003).

Notes on *Globiosporangium nunn*

We analyzed the field monitoring data in 2017 to identify potentially influential organisms and the manipulation experiments in 2019 were designed based on those results. In 2018, we found that the ASV identified as "*Pythium nunn*" (= the synonym of *Globisporangium nunn*) was potentially influential on rice growth and we decided to use the species in the manipulation experiments. Although there are differences in the sequence between the ASV detected in 2017 and registered *G. nunn* sequences (e.g., NBCI accession ID = AY598709.2), we used previously isolated and incubated *G. nunn* for the manipulation experiment in 2019 for the following reasons: (1) Among the BLAST top-hit sequences, *G. nunn* was the only species that can be compared to the type strain (the second top-hit sequence, *G. nunn* CBS808.96 is the type strain and has reliable information such as its morphology, culturability, and physiological characteristics is available) (Robideau *et al.* 2011), (2) there are several *G. nunn* sequences among the top-hit sequences, and (3) intraspecific sequence variations are known for *G. nunn*. In addition, there is no isolated and cultivated species for the first top-hit sequence, and the morphology, culturability, and physiological characteristics are not known. For these reasons, we decided to focus on *G. nunn* and used it in the manipulation experiment in 2019. We referred to the ASV detected in 2017 as "putative *Globisporangium*." Also, we referred to similar ASVs detected in 2019 as "putative *Globisporangium* spp.".

Field manipulation experiments in 2019

We performed the rice plot manipulations three times around noon on 24, 26, and 28 June 2019. There were three treatments, namely, *Globisporangium nunn*-added (GN), *Chironomus kiiensis*-removed (CK), and control (CT) treatments, and there were three replicate plots for each treatment (3 treatments × 3 replicates = 9 plots). Midge larvae were removed instead of adding them because isolation and cultivation of this insect species are difficult and time-consuming. For the GN treatment, incubated *Globisporangium nunn* was mixed in vermiculite and approximately 200 ml of vermiculite was added to each Wagner pot. Vermiculite without *Globisporangium nunn* was added to Wagner pots in the two other treatments. For the CK treatment, midge larvae were removed using a ϕ 1.0 mm net. On average, 283 midge larvae were removed from each plot and the DNA of about 100 removed larvae was sequenced using Sanger sequencing, and all were confirmed as *C. kiiensis*. For the other treatments, the net was also inserted into plots, but midge larvae were not removed.

Rice growth and ecological community monitoring and rice leaf sampling during the manipulation experiments

Before and after the manipulations, the rice plots were intensively monitored. See Table S1 for the detailed sampling design in 2019. Rice performance (growth rates and SPAD) was monitored every day from 11 June to 12 July (32 days). Methods for the rice monitoring was same as in 2017. Water samples for eDNA-based ecological community monitoring were collected every day from 18 June to 12 July (25 days). Methods for the eDNA-based ecological community monitoring were the same as those performed in 2017 (Ushio 2022). Rice leaf samples for RNA-seq were collected before and after the field manipulations. The leaf samples were immediately frozen by dipping into liquid nitrogen under field conditions, and kept frozen until stored in a freezer at -20°C. In total, 108 leaf samples were collected for RNA-seq analysis (see details in Table S2).

Effects of the field manipulation experiment on rice growth and ecological communities

The differences in rice performance among the three treatments and total eDNA concentrations of ecological community members (copies/ml water) before and after the manipulation experiments were tested using linear mixed model (LMM) and general linear mixed model (GLMM), respectively, using "lme4" package of R (Bates *et al.* 2015). For rice growth rates, the effect of the treatment was tested with the random effects of rice individual and rice plot (in R, this is `lme4::lmer(growth_rate ~ treatment + (1|ind/plot), data = data)`), and the effects were separately analyzed for before/after the manipulation. Rice cumulative growth was analyzed similarly, but the random effect of rice individuals was not included because there was only one cumulative value for each rice individual. Also, we tested alternative models including the timing and manipulation treatments (in R, these are `lme4::lmer(growth_rate ~ before_or_after_manipulation*treatment + (1|ind/plot), data = data)` and `lme4::lmer(cumulative_growth ~ before_or_after_manipulation*treatment + (1|plot), data = data)`), and obtained general agreement with the results of the above-mentioned analysis. For ecological community, the effects of the manipulation (i.e., the addition of *G. nunn* or the removal of *C. kiiensis*) were separately tested for each treatment with the random effect of rice plot assuming a gamma error distribution (in R, this is `lme4::glmer(DNA_conc ~ before_or_after_manipulation + (1|plot), data = data, family = Gamma(link="log"))`. Also, as in the rice growth analysis, we tested alternative models including the timing and manipulation treatments (in R, these are `lme4::glmer(DNA_conc ~ before_or_after_manipulation*treatment + (1|plot), data = data, family = Gamma(link="log"))`. Differences in the ecological community compositions were visualized using t-distributed stochastic neighbor embedding (t-SNE) (Van der Maaten & Hinton 2008).

Quantifications of RNA expressions of rice leaves

The leaf samples were ground under cryogenic conditions using a Multi-Beads Shocker (Yasui Kikai, Osaka, Japan). Total RNA was extracted using the Maxwell 16 LEV Plant RNA Kit (Promega, Madison, WI, USA). RNA concentration was measured using the broad-range Quant-iT RNA Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA). RNA (500 ng) was used as the input of each sample for library preparation. Library preparation for RNA-sequencing was conducted using Lasy-Seq (Kamitani *et al.* 2019) version 1.1 (https://sites.google.com/view/lasy-seq/). The library was sequenced using HiSeq X (Illumina, San Diego, CA, USA) with paired-end sequencing lengths of 150 bp. On average, 8,105,902 reads per sample (± 5,494,529 S.D.)

were generated (total reads = 875,534,703 reads; Table S2).

All obtained reads were trimmed using Trimmomatic version 0.33 (Bolger *et al.* 2014) using the following parameters: `TOPHRED33, ILLUMINACLIP:TruSeq3-SE.fa:2:30:10, LEADING:19, TRAILING:19, SLIDINGWINDOW:30:20, AVGQUAL:20, MINLEN:40`, indicating that reads with more than 39 nucleotides and average quality scores over 19 were reported. Then, the trimmed reads were mapped onto the reference sequences of the IRGSP-1.0_transcript (Kawahara *et al.* 2013) and the virus reference sequences, which were composed of complete genome sequences of 7457 viruses obtained from NCBI GenBank (Kashima *et al.* 2021) using RSEM version 1.3.0 (Li & Deway 2011) and Bowtie version 1.1.2 (Langmead *et al.* 2009) with default parameters. The output of the analysis, "expected_counts," was used as inputs to the analysis of differentially expressed genes.

Detections of differentially expressed genes (DEGs)

The expected counts were imported as a phyloseq object using "phyloseq" package (McMurdie & Holmes 2013) of R. Then, the object was converted to a DESeq2 object using `phyloseq::phyloseq_to_deseq2()` function, and differentially expressed genes (DEGs) were detected using "DESeq2" package (Love *et al.* 2014) of R. DESeq2 provides statistical frameworks for determining differential expression using a model based on the negative binomial distribution. Briefly, size factors and dispersions were estimated, generalized linear models (GLMs) assuming a negative binomial error distribution were fitted, and the differences were tested. These analyses were performed using `DESeq2::DESeq()` function implemented in the package. In the DESeq2 analysis, false discovery rate was set as 0.05, and genes with an adjusted $P < 0.05$ found by DESeq2 were assigned as differentially expressed.

Data and analysis code availability

Computer codes used in the present study are available in Github (https://github.com/ong8181/rice-ecolnet-2019). Sequence data were deposited in DDBJ Sequence Read Archives (DRA). The accession numbers are as follows: DRA009658, DRA009659, DRA009660 and DRA009661 for eDNA data of ecological communities in 2017 (Ushio 2022), DRA015682, DRA015683, DRA015685, and DRA015686 for eDNA for eDNA data of ecological communities in 2019, and DRA015706 for rice RNA expression data.

References cited in the Supporting Information

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1–48.

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.

Bolger, A.M., Lohse, M. & Usadel, B. (2014). Trimmomatic: A flexible trimmer for illumina sequence data. *Bioinformatics*, 30, 2114–2120.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A. & Holmes, S.P. (2016). DADA2: High-resolution sample inference from illumina amplicon data. *Nature Methods*, 13, 581–583.

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.

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.

Frenzel, S. & Pompe, B. (2007). Partial mutual information for coupling analysis of multivariate time series. *Physical Review Letters*, 99, 204101.

Kamitani, M., Kashima, M., Tezuka, A. & Nagano, A.J. (2019). Lasy-seq: A high-throughput library preparation method for RNA-seq and its application in the analysis of plant responses to fluctuating temperatures. *Scientific Reports*, 9, 7091.

Kashima, M., Sakamoto, R.L., Saito, H., Ohkubo, S., Tezuka, A., Deguchi, A., *et al.* (2021). Genomic basis of transcriptome dynamics in rice under field conditions. *Plant Cell and Physiology*.

Kawahara, Y., Bastide, M. de la, Hamilton, J.P., Kanamori, H., Mccombie, W.R., Ouyang, S., *et al.* (2013). Improvement of the oryza sativa nipponbare reference genome using next generation sequence and optical map data. *Rice*, 6, 4.

Kawai, K., Kawai, T. & Imabayashi, H. (2003). A comparison of improvemental ability of water quality among five chironomid species of the genus chironomus. *Medical Entomology and Zoology*, 54, 37–42.

Kobayashi, S., Uzuhashi, S., Tojo, M. & Kakishima, M. (2010). Characterization of pythium nunn newly recorded in japan and its antagonistic activity against p. Ultimum var. ultimum. *Journal of General Plant Pathology*, 76, 278–283.

Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10.

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.

Li, B. & Deway, C.N. (2011). RSEM: Accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics*, 12, 323.

Love, M.I., Huber, W. & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15, 550.

McMurdie, P.J. & Holmes, S. (2013). Phyloseq: An r package for reproducible interactive analysis and graphics of microbiome census data. *PloS one*, 8, e61217.

Osada, Y. & Ushio, M. (2021). R package for unified information-theoretic causality.

Robideau, G.P., De COCK, A.W.a.M., Coffey, M.D., Voglmayr, H., Brouwer, H., Bala, K., *et*

al. (2011). DNA barcoding of oomycetes with cytochrome c oxidase subunit i and internal transcribed spacer. *Molecular Ecology Resources*, 11, 1002–1011.

Runge, J., Heitzig, J., Petoukhov, V. & Kurths, J. (2012). Escaping the curse of dimensionality in estimating multivariate transfer entropy. *Physical Review Letters*, 108, 258701.

Sato, M., Inoue, N., Nambu, R., Furuichi, N., Imaizumi, T. & Ushio, M. (2021). Quantitative assessment of multiple fish species around artificial reefs combining environmental DNA metabarcoding and acoustic survey. *Scientific Reports*, 11, 19477.

Schreiber, T. (2000). Measuring information transfer. *Physical Review Letters*, 85, 461–464.

Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M.D.M., Breiner, H.-W., *et al.* (2010). Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Molecular Ecology*, 19 Suppl 1, 21–31.

Sugihara, G. & May, R.M. (1990). Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, 344, 734–41.

Sugihara, G., May, R., Ye, H., Hsieh, C., Deyle, E., Fogarty, M., *et al.* (2012). Detecting causality in complex ecosystems. *Science*, 338, 496–500.

Tanabe, A.S. & Toju, H. (2013). Two new computational methods for universal DNA barcoding: A benchmark using barcode sequences of bacteria, archaea, animals, fungi, and land plants. *PloS one*, 8, e76910.

Team, R.C. (2022). R: A language and environment for statistical computing.

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.

Tsuji, S., Shibata, N., Sawada, H. & Ushio, M. (2020). Quantitative evaluation of intraspecific genetic diversity in a natural fish population using environmental DNA analysis. *Molecular Ecology Resources*, 20, 1323–1332.

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. (2022). Interaction capacity as a potential driver of community diversity. *Proceedings of the Royal Society B: Biological Sciences*, 289, 20212690.

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.

Van der Maaten, L. & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9, 2579–2605.