Comparative environmental RNA and DNA metabarcoding analysis for ecological surveys of river algae and arthropods

Kaede Miyata  
Kao Corporation (Japan)

Yasuaki Inoue  
Kao Corporation (Japan)

Yuto Amano  
Kao Corporation (Japan)

Tohru Nishioka  
Kao Corporation (Japan)

Tomohisa Nagaike  
Bioindicator Co., Ltd.

Takamitsu Kawaguchi  
Bioindicator Co., Ltd.

Osamu Morita  
Kao Corporation (Japan)

Masayuki Yamane  
Kao Corporation (Japan)

Hiroshi Honda  
Kao Corporation (Japan)

Article

Keywords:

Posted Date: May 13th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1603139/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License

Version of Record: A version of this preprint was published at Scientific Reports on November 18th, 2022. See the published version at https://doi.org/10.1038/s41598-022-23888-1.
Abstract

Environmental DNA (eDNA) metabarcoding is widely used for species analysis, while the use of environmental RNA (eRNA) metabarcoding is more limited. We conducted comparative eDNA/eRNA metabarcoding of the algae and arthropods (aquatic insects) in water samples from Naka River, Japan to evaluate their potential for biological monitoring and water quality assessment. Both methods detected various algae and arthropod species, however, their compositions were remarkably different from those in traditional field surveys (TFSs), indicating a low sensitivity. For algae, the species composition derived from eDNA and eRNA metabarcoding was equivalent. While TFSs focus on attached algae, metabarcoding analysis theoretically detects both planktonic and attached algae. A recently expanded genomic database for aquatic insects significantly contributed to the sensitivity and positive predictivity for arthropods. While the sensitivity of eRNA was lower than that of eDNA, the positive predictivity of eRNA was higher. The eRNA of terrestrial arthropods indicated extremely high or low read numbers when compared with eDNA, suggesting that eRNA could be an effective indicator of false positives. Arthropod eDNA/eRNA metabarcoding analysis enabled water quality estimates from TFSs. The eRNA of algae and arthropods could thus be used to evaluate biodiversity and water quality and provide insights from ecological surveys.

Introduction

Ecosystem conservation and resource management are crucial sustainability issues that must be addressed to achieve sustainable development goals (SDGs) \(^1\). Frameworks using ecological surveys (e.g., the water framework directive \(^2\)) play a critical role in ensuring biodiversity. However, traditional field surveys (TFSs) are problematic owing to their low reproducibility, invasiveness, and the deference of species identification by evaluators \(^3\). Difficulty in discrimination due to morphological similarities in algae and arthropods, including aquatic insects in rivers, is an issue. Furthermore, to conserve genetic diversity, it is necessary to monitor cryptic species that are morphologically inseparable \(^4\).

Environmental DNA (eDNA) metabarcoding, which analyzes DNA released by macro-organisms in aquatic environments, is the most promising technology that can effectively overcoming the limitations of TFSs \(^5\)–\(^7\). Compared with TFSs, eDNA can non-invasively evaluate wide areas regardless of evaluator expertise. Furthermore, it can provide information regarding the richness and abundance of species \(^8\)–\(^15\). However, because eDNA is stable and can persist in the environment for a long period, the occurrence of false positives (low positive predictivity) in metabarcoding analysis remains a major issue \(^16\), \(^17\).

Hence, the use of environmental RNA (eRNA) in surveys, which may be less stable than eDNA, has attracted attention as a strategy to reduce false positive rates \(^18\)–\(^20\). However, the usefulness of eRNA in ecological surveys has not yet been fully evaluated. Recently, Miyata et al. \(^21\) demonstrated for the first time that fish eRNA was abundant in rivers and that eRNA metabarcoding analysis was useful for precise ecological surveys with high positive predictivity. In addition, Broman et al. \(^22\) reported the utility of shotgun sequencing using the eRNA of microeukaryote groups, including Arthropoda and Bacillariophyta, to investigate biodiversity. However, the abundance and usefulness of eRNA to discern other algae species and macrofauna (e.g., aquatic insects) in rivers remains unclear. Moreover, the utility of eRNA metabarcoding for ecological surveys of these species has not yet been fully evaluated. In the eco-toxicology of river ecosystems, fish, algae, and arthropods are commonly used as model trophic species, including herbivores and predators (e.g., for risk management or assessments of chemical substances). Ecological surveys of these species in actual environments are extremely important for the identification and management of hazardous factors. Therefore, the usefulness of eDNA/eRNA metabarcoding analysis should be evaluated in the model trophic species.

In this study, we conducted comparative eDNA/eRNA metabarcoding analyses using next-generation sequencing to confirm the abundance of algae and arthropod eRNA and determine the utility of eRNA for river ecological surveys. We also conducted TFSs for algae and arthropods and compared the results with those from eDNA/eRNA metabarcoding for a performance evaluation. Miyata et al. \(^21\) reported the use of eRNA for classifying saltwater and brackish water fish that were considered false positives. Hence, in this study, terrestrial arthropods and brackish and saltwater algae were defined as
potential false positives that might not inhabit the sampling points to elucidate the differences in species detected using eDNA and eRNA. Furthermore, to evaluate arthropods, we examined the effects of using a recently expanded genomic database on species detection performance and calculated the average score per taxon (ASPT) using eDNA, eRNA, and TFS for water quality evaluations.

**Materials & Methods**

**eDNA and eRNA samples**

Water samples were collected and eDNA/eRNA extracted, as previously described by Miyata et al. 21. The water samples (3.0 L, n = 22: 1 sample/site/day × 2 sites × 11 days) were collected twice or thrice a month at Nakagawa-oohashi (N36°32'55", E140°19'34" E) and Shin-nakabashi (N36°45'26", E140°08'30" E) on the Naka River in Japan. Then, Sterivex™ filter units were used to filter the water samples (nominal pore size, 0.45 µm; Millipore, Billerica, MA, USA), which were subsequently filled with ice-cold RNAProtect Tissue Reagent (Qiagen, Hilden, Germany). Total eDNA was extracted using an extraction buffer with RNase A, proteinase K, 15% polyvinylpolypyrrolidone, and buffer AL from a DNeasy blood and tissue kit (Qiagen, Hilden, Germany). Next, it was purified according to Miya et al. 23 using an MPure bacterial DNA extraction kit (MP Biomedicals, Santa Ana, CA, USA) and AMPure XP (Beckman Coulter, Brea, CA, USA). Total eRNA was extracted using a ChargeSwitch total RNA cell kit (Thermo Fisher Scientific, Waltham, MA, USA) with DNase treatment. Sample preparation was as follows. Sterivex™ filter units were filled with 1,000 µL of the prepared Lysis Mix and incubate at 60°C for 1 hr. After incubation, the lysate was collected by centrifugation and transferred into a 1.5 mL microcentrifuge tube. Then, the lysate was mixed briefly by vortexing and cooled the samples for 1 minute on ice. From then on, we followed the manufacture's protocol. Then, cDNA was synthesized using a PrimeScript II 1st strand cDNA synthesis kit (TaKaRa Bio Inc., Shiga, Japan), following the manufacture's protocol.

**Amplicon library preparation and next-generation sequencing**

The amplicon library was prepared, and MiSeq sequencing was conducted as previously described by Miyata et al. 21. Briefly, amplicon libraries of partial 16S rRNA and psbA genes were obtained by PCR amplification using gInsect and psbA (Bioengineering Lab. Co., Ltd., Japan) 24 (Table 1). The first and second PCR (and purification of amplicons) were performed according to Miyata et al. 21. However, for psbA, we modified the conditions of the first PCR. psbA underwent initial denaturation at 94°C for 2 min, followed by 35 cycles of denaturation at 94°C for 30 s, and annealing at 52°C for 30 s. Then, it was extended at 72°C for 30 s with a final extension at 72°C for 5 min. The first PCR amplification was replicated four times, and the second PCR used adaptor sequences and 8 bp index sequences.

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence (5' → 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>gInsect F</td>
<td>ACACTCTTTCCTACACGACCGCTCTTCCGATCTNNNNNNGATAGAAACCAACCTGGCT</td>
</tr>
<tr>
<td>gInsect R</td>
<td>GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNNGACGAGAAGACCTGATA</td>
</tr>
<tr>
<td>psbA_597f</td>
<td>ACACTCTTTCCTACACGACCGCTCTTCCGATCTTNCAYTTCTAYCCNVHTGGAAG</td>
</tr>
<tr>
<td>psbA_927r</td>
<td>GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTRNCAATGGAATGCTGAT</td>
</tr>
</tbody>
</table>

**Bioinformatics analysis of high-throughput sequencing data**

Bioinformatics analysis was conducted according to Miyata et al. 21. First, multiple FASTQ files, matched against the beginning of the specified index reads, were extracted using the fastq_barcode_splitter tool in the program FASTX-Toolkit (ver. 0.0.14). Amplicon sequence variants (ASVs) were produced using DADA2 25. The primer sequences in psbA were...
trimmed 50 bp from the 3-terminal end. Next, chimeras and noise sequences were removed using Qiime 2 plugin wraps DADA2 with default settings (ver. 2020.8). In addition, a sequence table was created as a matrix in which each row corresponds to a processed sample and each column indicates a non-chimeric inferred sample sequence. All processed sequences for algae and arthropods were subjected to BLASTN search (ver. 2.9.0) against the NCBI nonredundant nucleotide sequence database, and those for arthropods were also subjected to Mitochondrial Genome Database of Benthos and Insect (for Environmental Research) in Japan (MBIJ) developed independently by the researchers of the Bioengineering Lab. Co., Ltd. et al. (see Data Availability section). The list of species included in MBIJ is provided in Supplementary Table S1. The Qzv file was exported to another file type.

**Traditional field survey (TFS)**

Arthropod TFSs were conducted using a method adapted from the Manual of the National Census of River Environments. Samples were collected via kick-sampling or wash-sampling using dip nets in each identified instream habitat (e.g., ripples, pools) and a Surber sampler measuring 25 cm × 25 cm (0.5 mm mesh). Samples were collected at three points in ripples at each research site/season using a Surber sampler.

The TFSs for algae were conducted with a focus on epilithic (attached) algae. Three cobbles were collected from ripples at each research site/season, and algae were scraped from 5 × 5 cm sampling areas on each cobble. Samples were also collected from each identified instream habitat by scraping the cobbles (the sampling area was not fixed).

**Performance evaluation of eDNA/eRNA metabarcoding analyses in ecological surveys**

To examine the performance of the eDNA/eRNA metabarcoding analysis, their sensitivity and positive predictivity were calculated as follows:

Sensitivity = (number of true positives)/(number of positives in TFS)

Positive predictivity = (number of true positives)/(number of positives in metabarcoding analysis)

In this analysis, true positives were defined as “identified species using metabarcoding analysis of the observed species in TFS.” Specificity and negative predictivity were not evaluated as there were no true negatives in the TFS.

The average total read numbers at the two sampling points were used to compare the merged TFS data (Nakagawa-oohashi and Shin-nakabashi). Species with read numbers < 10 bp for both eDNA and eRNA were excluded. In the arthropod analysis, sensitivity and positive predictivity were evaluated using two types of TFS data: 1) quantitative data only and 2) quantitative and qualitative data (entire data). The performance was also evaluated for each taxon, focusing on Ephemeroptera, Plecoptera, Trichoptera, Odonata, and Diptera (EPTOD). Furthermore, we confirmed the performance of metabarcoding analysis for arthropod ecological surveys using the TFS database of the Naka River. However, since the sequence data for the 16S rRNA has recently been expanded, performances were compared using a current database and an expanded database (species included in current database plus MBIJ). Finally, using the best analytical conditions (TFS: quantitative data only, Metabarcoding: the expanded database), Venn diagrams were utilized to compare the number of species detected using eDNA, eRNA, and TFS. For arthropods, species that were detected and not detected in the eDNA/eRNA metabarcoding analysis were analyzed in terms of sequence registration in the database and the detected ratio of each taxon. In addition, the top 30 arthropod species with high abundance detected by TFS and read numbers following metabarcoding analysis were also compared using a heatmap. Finally, we analyzed the status of sequence registration in the database.

**Usefulness of arthropod eDNA/eRNA metabarcoding analysis for water quality evaluation.**

We evaluated the usefulness of eDNA/eRNA metabarcoding analysis for deriving water quality indices. The eDNA/eRNA metabarcoding analysis and TFS data were used to calculate Ephemeroptera, Plecoptera, and Trichoptera (EPT) (%) and the
average tolerance scores of all macroinvertebrate families found at the site (ASPT), which were utilized to evaluate water
quality. The ASPT was calculated according to the manual for water quality assessment for aquatic organisms in Japan\textsuperscript{29}. Then, we compared the EPT (%) and ASPT scores.

**Differences in eDNA/eRNA metabarcoding analyses for classifying false positives**

To discuss the capacity to distinguish living biotic assemblages from dead biotic assemblages, terrestrial arthropods and
brackish and saltwater algae were defined as potential false positives that were not likely to inhabit the sampling points. Moreover, the features of attached/planktonic algae were analyzed but not defined as false positives as planktonic algae are not usually evaluated in TFSs. The habitats were examined using information from the MLIT-RE database\textsuperscript{28}, scatter plots between eDNA and eRNA were illustrated for all species detected, and the distributions were compared using box plots.

**Statistical analyses**

One-way ANOVA, followed by the Tukey–Kramer test, was employed to analyze statistical differences in sensitivity and positive predictivity for multiple comparisons under different analysis conditions for arthropods. The sensitivity and positive predictivity for algae were analyzed using Student's t-test. The F-test was used to determine the homogeneity of variance across the EPT (%), ASPT, and read numbers. The Student's t-test was performed when the variance was homogeneous, whereas the Mann–Whitney U test was performed when the variance was heterogeneous. All statistical analyses, except for the Tukey–Kramer test, were performed using Origin 2019b (OriginLab, Northampton, MA, USA). Tukey–Kramer test was performed using the multcomp package in R software. Statistical significance was set at P < 0.05.

**Results**

**Performance of eDNA/eRNA metabarcoding analysis for ecological surveys**

Sensitivity and positive predictivities were compared across different experimental conditions (Fig. 1). The sensitivity and positive predictivity of the eDNA/eRNA metabarcoding analysis in arthropods both drastically improved when an expanded database was used (merged data; current database: 0–3.2%/0–22.2%, expanded database: 7.6–22.1%/43.7–73.1% (sensitivity/positive predictivity)) (Fig. 1A). The sequences in the expanded database improved the performance of the metabarcoding analysis. Furthermore, when only quantitative data were used as TFS data, there were maximum sensitivity values for the eDNA and eRNA. However, positive predictivities using quantitative analysis slightly decreased when compared with those using quantitative and qualitative data (no significant differences were found). Similar results were confirmed for the metabarcoding of arthropod ecological surveys using the MLIT database (Supplementary Fig. S1). Thus, in this study, we mainly analyzed the conditions with high sensitivity (using quantitative TFS data obtained in the present study and an expanded database of metabarcoding analysis) for arthropods.

When comparing eDNA and eRNA, the sensitivity of eDNA was significantly higher than that of eRNA in the expanded database. In contrast, the positive predictivity of eRNA was higher than that of eDNA, although no significant difference was found.

When individual datasets were merged, sensitivity increased, and positive predictivity decreased. The order of the performance of eDNA and eRNA did not change. However, the percentage increase in sensitivity with the merged data was higher for eRNA (eDNA: 12.3–22.1% [×1.79], eRNA: 3.5–11.4% [×3.25]). In contrast, the percentage decrease in positive predictivity was higher for eDNA (eDNA: 67.4–43.7% [×0.83], eRNA: 73.4–61.5% [×0.64]). This indicates that eRNA can easily increase sensitivity, and eDNA may be prone to generating false positives on using merged data.

The performance of each taxon was also evaluated, focusing on the EPTOD (Table 2). Plecoptera was not detected in eDNA or eRNA metabarcoding. In terms of sensitivity, Ephemeroptera had the highest average score for eDNA and Diptera had the
greatest score for eRNA. Odonata had the highest merged score for both eDNA and eRNA. In terms of positive predictivity, the average and merged scores were the highest for Ephemeroptera in eDNA and Trichoptera in eRNA.

In algae, the sensitivity was mostly equivalent. The average score was slightly higher for eRNA (eDNA 9.1%, eRNA 9.4%), and the merged score was equal (no significant difference). Positive predictivities were also similar. The average score was higher for eRNA, whereas the merged score was higher for eDNA. However, no significant differences were found (Fig. 1B).

**Differences in the species detected using eDNA/eRNA vs. TFS analyses**

Venn diagrams illustrate the differences in species detected between the eDNA/eRNA metabarcoding analysis and TFS under multiple analytical conditions (Fig. 2). For the arthropods, eDNA and eRNA analyses detected 71 and 26 species, respectively. The TFS detected 140 species at two sampling points (Nakagawa-oohashi and Shin-nakabashi). The lower species detection rate with eRNA was probably responsible for the low sensitivity and high standard deviation of positive predictivity in eRNA (Fig. 1A). Although 25 species were commonly detected in the eDNA/eRNA analyses, only 16 species were common in the eDNA/eRNA analyses and TFS. The sensitivity of eRNA detection (11.4%) was lower than that of eDNA detection (22.1%). The positive predictivity of eRNA (61.5%) was higher than that of eDNA (43.7%).

For algae, eDNA and eRNA analyses detected 71 and 80 species, respectively, equivalent to the number detected with the TFS (61). However, only 8 species were common with the TFS results. Conversely, almost all species (62) were common in eDNA/eRNA analyses. The sensitivity and positive predictivity of eDNA/eRNA were equivalent (sensitivity, 13.1%; positive predictivity, 10–11.3%). The TFS focuses on algae attached to stones in rivers, whereas eDNA/eRNA was collected from surface water. Hence, eDNA/eRNA analyses could detect species that were identified by TFS when using bottom water.

A total of 109 species (67+42 species in Fig. 3A) were not detected in the metabarcoding analysis, but were detected in the TFS. Moreover, 42 of the 109 species have already been registered in an expanded database. Then, we compared the composition of the species detected with the eDNA/eRNA metabarcoding analysis and those only detected in the TFS (Fig. 3B). The results showed that the EPTO ratio was similar (approximately 60–70%). However, Plecoptera was not detected in eDNA/eRNA metabarcoding, whereas all Odonata species could be detected in that analysis.

**Comparison of the read numbers from the metabarcoding analysis and TFS abundance**

The abundance of the top 30 arthropod species in the TFS and their read numbers in the metabarcoding analysis were illustrated using a heatmap (Fig. 4). No species were detected in the eRNA metabarcoding only, and the eDNA identified all species detected in the eRNA. For several species (e.g., *Drunella ishiyamana* and *Isonychia valida*), the eDNA read numbers were remarkably higher than the eRNA read numbers. Of the top 30 species in TFS, 10 were found in eDNA metabarcoding. Of these 10 species, 7 were found in eDNA and 6 were detected in eRNA. Of the 20 species not found in eDNA/eRNA metabarcoding, 10 have been already registered in the expanded database.

**Comparison of water quality indices in eDNA/eRNA metabarcoding and TFS**

To evaluate the usefulness of metabarcoding analysis as a water quality index, the EPT (%) and ASPT scores derived from TFS were compared with those derived from eDNA/eRNA metabarcoding analyses. For EPT (%), the average score derived from the eDNA was slightly higher than that derived from TFS. In contrast, the average score derived from eRNA was lower than that derived from TFS (Fig 5A). When comparing the merged scores, eRNA had the highest score when compared with eDNA and TFS. The average ASPT score derived from TFS was higher than that from the eDNA and eRNA. When comparing the merged scores, the ASPT scores derived from the eDNA, eRNA, and TFS were equivalent (Fig. 5B). These data suggest that individual surveys using eDNA can allow the stable evaluation of water quality indices derived from the TFS. However, the individual values of the indices derived from eRNA may be slightly lower than those from TFS, while merged values are likely to be equivalent to those from the TFS.

**Effect of ecological traits related to false positive generation in eDNA/eRNA metabarcoding analysis**
Some species of arthropods were detected only in the eDNA analysis and the common functions of these species were explored. Interestingly, terrestrial arthropods were mainly detected in eDNA and not in eRNA. However, two terrestrial species (*Pteronemobius fascipes* and *Nysius plebeius*) had extremely high read numbers in eRNA when compared with those in the eDNA (Fig. 6A). These results suggest that the amount of eRNA in terrestrial arthropods can be extremely high or low. Thus, eRNA could act as an indicator to determine whether a species is actually inhabiting the water environment or if it enters from the outside. Moreover, the eDNA read numbers of aquatic arthropods were significantly higher than those of terrestrial arthropods (Fig. 6B). Hence, the combined analysis of eDNA and eRNA may provide an effective means to detect false positives.

In algal metabarcoding analysis, common ecological traits were not found in species detected only in eDNA or eRNA. Planktonic and attached algae, together with seawater and brackish algae were distributed homogeneously (Fig. 6C).

**Discussion**

Our results indicate that the eRNA from a variety of algal and arthropod species is maintained in rivers, which enables metabarcoding analysis. It has been reported that the eRNA of various aquatic species are available in the marine\(^{30-37}\). Additionally, sufficient amounts of fish eRNA for ecological survey have been detected in rivers \(^{21}\). These results indicate that the eRNA of major trophic species, including herbivores and predators in rivers, is likely maintained in all water environments. This is also consistent with the view that eRNA has a certain level of stability \(^{38-40}\) that allows its maintenance in the environment \(^{20}\).

The performance of eDNA/RNA metabarcoding in ecological surveys did not achieve sufficient levels for either algae or arthropods when existing databases were used. Sensitivity and positive predictivity for the algae and arthropods were low when compared with those previously reported for fish \(^{21}\). For algae, the detected eDNA species were in good agreement with those of the eRNA when compared with arthropods, and significant differences were not observed in sensitivity and positive predictivity. The lack of differences in algae may be associated with the direct capture of algae in the filter units. As the size of algae exceeded the pore size \(^{41}\), our protocol may have captured the eDNA and eRNA of algae, regardless of whether they were dead or alive. Although several studies have reported that eRNA is a sensitive marker of environmental changes \(^{42}\), further analysis is required to investigate the differences between eDNA and eRNA metabarcoding. Nonetheless, a novel aspect identified in this study regarding eDNA/eRNA metabarcoding was that both planktonic and attached algae could be detected. The species detected in metabarcoding were similar to those identified using TFSs. However, half of the species identified were planktonic algae, indicating that they were associated with low performance. Metabarcoding analysis, which can be used to analyze planktonic algae, may be useful for investigating a wide range of species. TFSs or metabarcoding analysis using bulk samples, which target only non-moving attached algae, could be useful for investigating their relationship with environmental factors.

The identification of arthropods was extremely low when using the existing sequence database for metabarcoding; however, it immensely improved when the expanded database was used. In particular, positive predictivity showed a maximum value in eRNA (combined data)—as the value was changed from approximately 0% to 80%, which was considered sufficient. This suggests that the sequence registration of organisms found at the site may be an efficient approach to enhance eDNA/eRNA metabarcoding for ecological surveys. Additionally, this may increase positive predictivity for arthropods. To further improve the sensitivity, two approaches should be considered. The first strategy involves the sequencing and registration of species not identified by metabarcoding. However, of the 109 species that were not detected, 42 have been already registered in the expanded database. Thus, sequence registration alone has limitations in improving the sensitivity. Crustacean eDNA is released in small quantities \(^{43}\) as their bodies are covered by exoskeletons. Therefore, even arthropods covered with exoskeleton may release a small amount of eDNA and eRNA, reflecting the fact that EPTO taxa are difficult to detect in water samples \(^{44}\). Hence, the development of a new approach for collecting eDNA and eRNA is necessary. For example, a bulk analysis may be conceivable, wherein fragments of arthropods may be collected from sediment in rivers, regardless of
whether they are alive or dead. Metabarcoding analysis using bulk samples can reportedly improve performance\textsuperscript{44}. Second, water samples can be collected from various subsites. TFS qualitative analysis was conducted in shallow areas, splash water, pools, and riverbanks, whereas metabarcoding analysis used surface water to prevent contamination with PCR inhibitors. Therefore, matching sampling sites may increase accuracy. This is supported by evidence that sensitivity increased when quantitative data alone were used in this study. If sensitivity is sufficiently improved by these measures, a stable and high level of positive predictivity may be obtained in arthropod eRNA as well as fish eRNA.

Cristescu\textsuperscript{20} discussed the possibility that eRNA enables ecological surveys with high positive predictivity (low false-positive rate) and Miyata et al.\textsuperscript{21} indicated that eRNA contributes to reducing false positives in river ecological surveys. The positive predictability of fish eRNA was higher than that of fish eDNA, and the sensitivity of eRNA was higher than that of eDNA when the results of several analyses were combined. In this study, positive predictivity and sensitivity were equivalent for the eDNA and eRNA of algae. However, the positive predictivity of eRNA was higher than that of eDNA in all surveys and merged data. Although this result is consistent with that of previous fish surveys, the sensitivity of eRNA in the merged data was lower than that of eDNA. In addition, the differences between the sensitivities of eDNA and eRNA in the merged data were smaller than those in the individual surveys. Thus, expanding the sequence database and developing high efficiency eDNA/RNA collection methods may contribute to the recovery of eRNA metabarcoding accuracy in the future.

In a previous study investigating positive predictivity, fish species that were detected only in the eDNA metabarcoding analysis were sea and brackish water fish that did not inhabit the sampling points\textsuperscript{21}. Similarly, although saltwater algae were equally detected in the eDNA and eRNA, terrestrial arthropods tended to be harder to detect using eRNA than when using eDNA. Interestingly, the eRNA abundances of two species (\textit{Pteronemobius fascipes} and \textit{Nysius plebeius}) were remarkably higher than their eDNA abundances. \textit{Pteronemobius fascipes} are found in gravel areas near riverbanks in Japan\textsuperscript{45} and tend to flow into rivers. There are cases where RNA abundances are higher than the DNA abundance in the body and organisms\textsuperscript{46,47}. This suggests that if organisms, which are the source of pollution, flow into a location alive or shortly after death, they may have higher eRNA values in comparison with eDNA values. Thus, the ratio of eDNA to eRNA could be a useful indicator of false-positive results.

This study did not focus on the differences in target sequencing for metabarcoding. Several universal primers have been reported for different gene sites, especially in macro-invertebrates. At the beginning of DNA barcoding analysis, cytochrome c oxidase subunit I (COI) was used\textsuperscript{48,49}. COI eDNA metabarcoding analysis on arthropods has been conducted for a long time, allowing the development of some universal primers. Primers for other regions (12S and 16S rRNA) have also been used\textsuperscript{24,50}. Presently, the number of registered COI sequences is much higher than those for 12S rRNA and 16S rRNA. Hence, arranging universal primers may improve detection efficiency by using different genomic positions in COI\textsuperscript{51} or a combination of PCR primers\textsuperscript{52}. However, comparison studies\textsuperscript{53,54} and a simulation\textsuperscript{55} indicated that 16S and 12S could be superior to COI for identifying several taxa of insects. Therefore, 12S and 16S rRNA may enable precise identification in the future by expanding the sequence database. In this study, glInsect primers (16S rRNA) could detect all \textit{Odonata} species at the sampling points (Supplementary Table S1), which might be difficult to detect using COI metabarcoding.

For the environmental assessment of rivers, benthic macro-invertebrates have been used as variable measures. We found that ASPT values derived from the metabarcoding analysis in arthropods were similar to those derived from TFS. Currently, eRNA analysis is less sensitive than eDNA analysis; however, merged eDNA/eRNA data can be used to estimate water quality values from TFSs. Similarly, Emilson et al.\textsuperscript{56} reported that DNA metabarcoding and morphologically derived macroinvertebrate metrics were positively correlated. Uchida et al.\textsuperscript{57} reported that the EPT index derived from eDNA showed a more positive relationship with the total nitrogen concentration than the indices calculated using the Surber net survey data. More adequate scores of species for the metabarcoding analysis could thus be determined to help assess the water environment. However, since the water quality is typically assessed using information on richness, the features of DNA/eRNA metabarcoding analysis, which are useful for comprehensive species detection, are suitable for deriving the indices. Thus,
current knowledge and measures could be used in eDNA/eRNA metabarcoding analysis to improve the limitations of ecological survey using morphological evaluations.

Conclusions

In conclusion, to our knowledge, this study is the first to conduct a comparative eDNA/eRNA metabarcoding analysis on algae and arthropods to evaluate the performance of ecological surveys in rivers. The results indicated that both eDNA and eRNA detected a variety of species. The species composition in these analyses was remarkably different from that in the TFS, and the sensitivities were low. However, for arthropods the low sensitivity could be improved by using an expanded genomic database. Additionally, although TFSs focus on attached algae, metabarcoding analysis can detect both planktonic and attached algae; however these differences require further investigation. For arthropods, although the sensitivity of eRNA was found to be lower than that of the eDNA, the positive predictability of eRNA was higher. This indicates that eRNA could be useful for determining false positives in eDNA analysis. Moreover, the balance between eDNA and eRNA can vary depending on the habitat of an organism. Further extensive investigations are thus required to determine the value of implementing eRNA analysis in ecological surveys.

Declarations

Acknowledgements

We thank Mr. Masahiko Kakinuma (Asahi Taxi) for his support during the field experiment. We also acknowledge the contributions to the metabarcoding analysis from the researchers at the Bioengineering Lab. Co., Ltd., Japan.

Author contributions


Competing Interest

Data Availability

The sequence data in MBIJ that support the findings of this study belongs to a third party (Bioengineering Lab. Co., Ltd., Plantbio Co., Ltd., Kanagawa Environmental Research Center and Takumi Shimizu (Keio University)). Thus, availability restrictions apply to the data used under license for the current study, and consequently, these are not publicly available. However, species data were listed in Supplementary Table S1. If you need additional information from the party, please contact the form in Bioengineering Lab. Co., Ltd. (https://gikenbio.com/contacts/) or e-mail address (dna@gikenbio.com).

References


### Tables

**Table 1: PCR primers used in the present study**

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence (5′ → 3′)</th>
</tr>
</thead>
<tbody>
<tr>
<td>gInsect F</td>
<td>ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNGATAGAAACCAACCTGGCT</td>
</tr>
<tr>
<td>gInsect R</td>
<td>GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNGACGAGAAGACCCTATA</td>
</tr>
<tr>
<td>psbA_597f</td>
<td>ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNNNGATAGAAACCAACCTGGCT</td>
</tr>
<tr>
<td>psbA_927r</td>
<td>GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNGATAGAAACCAACCTGGCT</td>
</tr>
</tbody>
</table>

**Table 2: Sensitivity and positive predictivity of metabarcoding analysis in each class**

<table>
<thead>
<tr>
<th>Sensitivity</th>
<th>eDNA</th>
<th>eRNA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Ephemeroptera</td>
<td>28.1</td>
<td>6.4</td>
</tr>
<tr>
<td>Plecoptera</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Trichoptera</td>
<td>13.2</td>
<td>6.0</td>
</tr>
<tr>
<td>Odonata</td>
<td>27.3</td>
<td>24.9</td>
</tr>
<tr>
<td>Diptera</td>
<td>21.8</td>
<td>3.9</td>
</tr>
<tr>
<td>Positive predictivity</td>
<td>eDNA</td>
<td>eRNA</td>
</tr>
<tr>
<td>-----------------------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Ephemeroptera</td>
<td>86.9</td>
<td>6.1</td>
</tr>
<tr>
<td>Plecoptera</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Trichoptera</td>
<td>56.1</td>
<td>21.6</td>
</tr>
<tr>
<td>Odonata</td>
<td>26.7</td>
<td>22.6</td>
</tr>
<tr>
<td>Diptera</td>
<td>36.9</td>
<td>7.6</td>
</tr>
</tbody>
</table>

**Figures**

**Figure 1**

**Performance of eDNA/eRNA metabarcoding analysis in ecological surveys of arthropods and algae.** Sensitivity and positive predictivity of metabarcoding analysis for (A) arthropods and (B) algae. eDNA/eRNA: eDNA/eRNA metabarcoding analysis, quant.: quantitative analysis, qualit.: qualitative analysis. Dots indicate the performance when individual data are merged. The values represent the mean + standard deviation (SD). Different letters indicate significant differences (Tukey–Kramer test). The values represent the mean + standard deviation (SD). Arcsine transformation did not change the letters.
Figure 2

Comparison of the species detected using the eDNA/eRNA metabarcoding analyses and traditional field surveys (TFSs). (A–F) Comparison of the species detected using eDNA/eRNA metabarcoding and TFS. Venn diagrams were illustrated for multiple analytical conditions in a matrix (species vs. sites). For example, (A) indicates a result using two-site merged data in arthropods, and (E) uses data only from Nakagawa-oohashi for algae.

Figure 3

Number of arthropod species detected in the traditional field survey and metabarcoding analysis and their sequence data availability in an expanded database. (A) Comparison of the arthropod species detected using eDNA/eRNA metabarcoding and traditional field surveys (TFSs), and the existence of recently registered species in the expanded database. (B) Comparison of the ratios between species detected in environmental nucleic acids (eNA [eDNA and eRNA]) and TFS and species detected in only TFS.
Figure 4

The 30 arthropod species with the greatest abundance detected in traditional field surveys and their read numbers in metabarcoding analysis. MTB: metabarcoding analysis, TFS: traditional field survey, DB: database. Current: current database, New: recently registered species in the expanded database. Read numbers in eDNA/eRNA metabarcoding analysis and abundances in TFS were illustrated.

Figure 5

Comparison of Ephemeroptera, Plecoptera, and Trichoptera (EPT) ratios and the average score per taxon (ASPT) derived from eDNA/eRNA metabarcoding analysis and TFS. (A) Comparison of EPT (%) between eDNA/eRNA metabarcoding analysis and TFS. (B) Comparison of the ASPT scores between the eDNA/eRNA metabarcoding analysis and TFS. Square; merged score in eDNA/eRNA.
Figure 6

Ecological trait evaluation using eDNA/eRNA metabarcoding analyses to identify false positives. (A) Scatter plot between eRNA and eDNA read numbers for arthropods and species’ habitat. (B) Differences in eDNA/eRNA read numbers between aquatic and terrestrial arthropods. (C) Scatter plot between eRNA read numbers and eDNA read numbers in algae and ecological traits related to false positives, (Left) attached or planktonic (Right) fresh water or seawater and brackish water. **p < 0.01.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigureS1.docx
- SupplementaryTableS1.xlsx