Rationale:

Biodiversity assessment is a crucial tool for conserving flora and fauna. Conventional sampling methods, reliant on field conditions and specific skills, limit comprehensive assessments. Despite their potential, many eDNA-based studies tend to focus on a singular or a limited number of taxonomic groups, while multi-trophic eDNA assessments remain scarce. Notable examples of such approaches include the works of Stat et al. (2017) and Zhang et al. (2020), which employed “tree-of-life” eDNA metabarcoding in marine environments. In freshwater systems, the only known multi-trophic eDNA study, conducted by Beentjes et al. (2021), was confined to three groups: bacteria, phytoplankton, and chironomids.

To complement eDNA-based biodiversity assessments, reduced-representation sequencing methods such as double-digest RADseq (ddRAD) provide a powerful means to assess genetic diversity and population structure within key taxa. While eDNA metabarcoding offers a broad snapshot of community composition, ddRAD enables finer-scale resolution of intraspecific diversity, evolutionary history, and potential adaptive responses. Integrating these approaches allows for a more comprehensive understanding of biodiversity, from species presence to genetic variation, which is crucial for effective conservation and management strategies. Here, we explore the potential of combining eDNA metabarcoding and reduced-representation sequencing approach, such as double-digest RADseq (ddRAD), to assess biodiversity at multiple levels from community composition to intraspecific genetic diversity.

Procedure:

One crucial step in using ddRAD for community-level analysis is optimizing library preparation to accommodate diverse taxonomic groups. This optimization is essential for ensuring the successful integration of ddRAD and eDNA metabarcoding in comprehensive community-level assessments.

Restriction Enzyme Selection. We tested eleven enzyme pairs and various incubation times (1 h, 3 h, 6 h, 12 h, and 16 h). The initial enzyme selection was based on published studies and in silico analysis using ddSilico. DNA digest fragments were then assessed via gel electrophoresis and a Bioanalyzer (Agilent 2100) to confirm fragment size distribution. Enzyme pairs and incubation times that produced appropriate fragment sizes (≤300 bp) across different samples were selected for subsequent analyses.

Library Preparation. After the optimization of enzyme digestion in various samples, we then prepared the library following the protocols as described by Peterson et al. (2012), Parchman et al. (2012), and Kess et al. (2016) with slight modifications.

Following library preparation, samples were pooled at equimolar concentrations. SPRI size selection and cleanup were then performed to ensure high-quality samples for sequencing.

Paired-end sequencing was performed on a NovaSeq X10 at the Novogene Tokyo, Japan laboratory. Since ddRADseq generates diverse nucleotide sequences, a low PhiX percentage was used during sequencing to optimize data quality and throughput.

Sequence Quality Analysis and Trimming. Sequence quality per read and per base, repetitive sequences, and adapter contamination were assessed using FastQC and MultiQC. Trimming was then performed with Trimmomatic to remove ambiguous base calls, adapter sequences, and barcode regions.

De Novo Assembly and Parameter Optimization. A major challenge in using ddRAD for community-level analysis is the lack of reference genomes for most non-model organisms. Additionally, assembling paired-end RAD reads is particularly difficult for species with high genetic variability. In this study, we addressed these challenges by constructing pseudo-reference genomes for various non-model organisms using RADassembler (Li et al., 2018). We first optimized the -m, -M, and -N parameters in Stacks, which were then applied during the assembly process. The resulting pseudo-reference genomes were used for reference-based assembly in Stacks, and the assembled data were compared to those obtained through de novo assembly.

Progress and Preliminary Results

***Restriction Enzyme Selection*.** The in-silico performance of restriction enzymes, including the types of overhangs produced, sequencing efficiency, and fragment distribution, is summarized in Table 2. These predictions provide insights into

enzyme efficiency and compatibility for ddRAD library preparation. Additionally, the predicted number of SNPs generated by different enzyme pairs followed this order: NlaIII-MspI > EcoRI-MspI > NlaIII-MluCI > EcoRI-MseI > PstI-MseI (data not shown).

Interestingly, despite its high predicted SNP yield, NlaIII-MspI did not produce an optimal fragment size distribution across different incubation durations (data not shown). Instead, the NlaIII-MluCI enzyme pair generated the desired fragment size distribution at an incubation time of 1 hour (Figure 1).

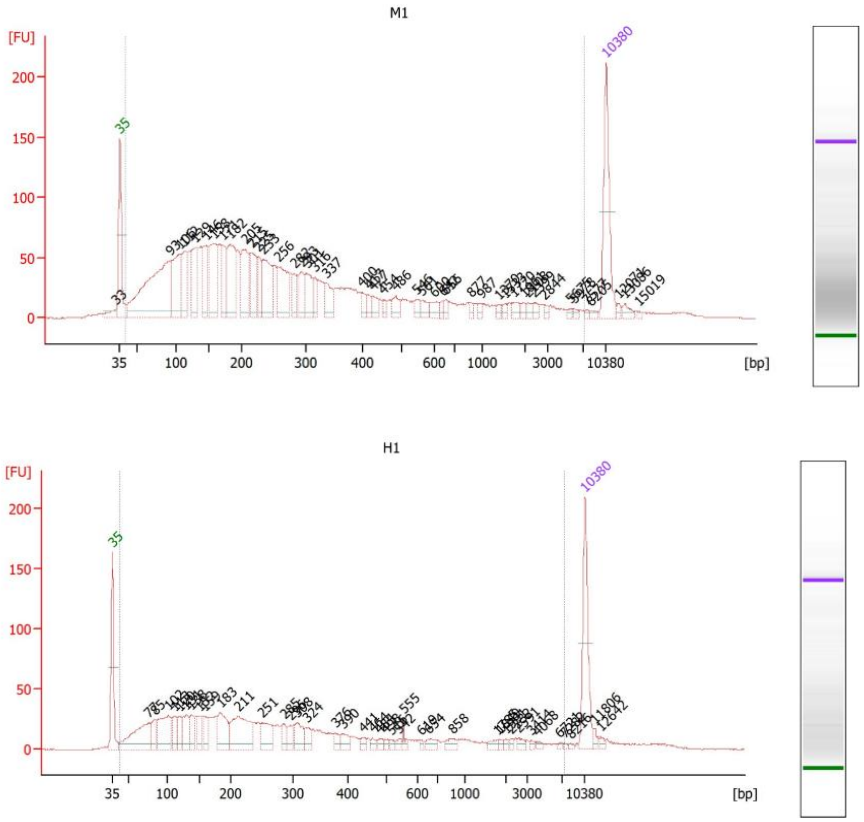
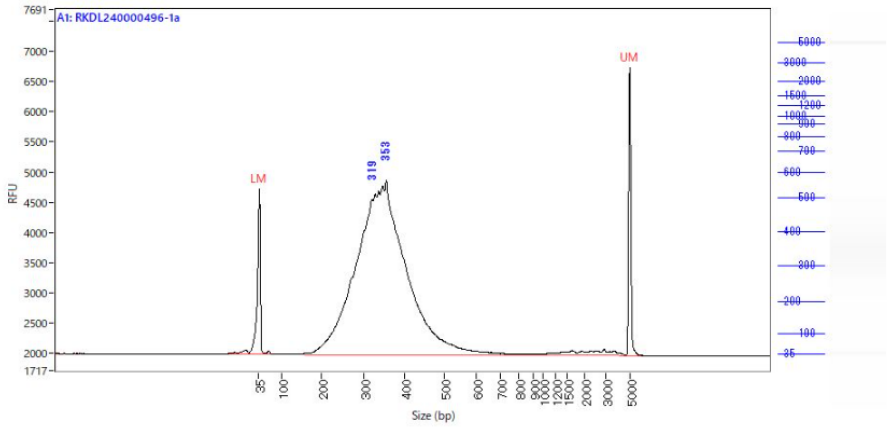


Figure 1. Fragment size distribution of digested genomic DNA after 1-hour incubation. (M1) *Stenopsyche marmorata*, and (H1) *Hydropsyche orientalis*.

After library preparation, a final quality control was performed using bioanalyzer (Figure 3) and qPCR (data not shown).



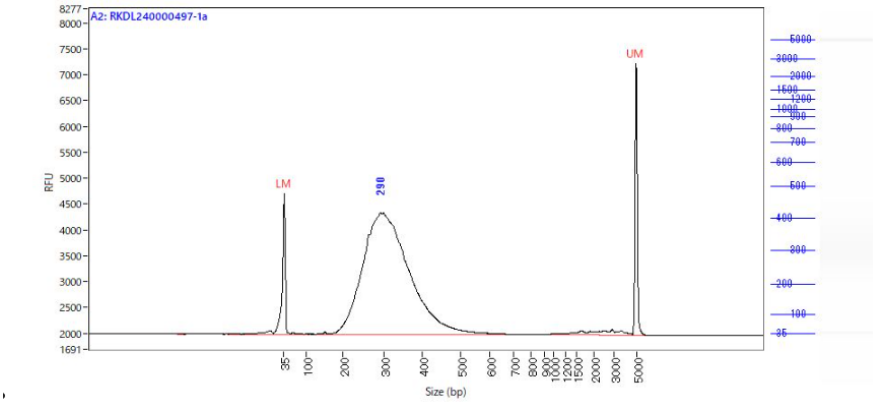


Figure 2. Bioanalyzer quality control analysis after size selection and equimolar pooling. The top and bottom figures represent two different sample pools.

Sequence Quality and Parameter Optimization. We obtained high-quality sequencing data, with mean Phred scores consistently around 40, indicating minimal sequencing errors (Figure 3). Additionally, ambiguous base calls were virtually absent across all sequences, ensuring high confidence in downstream analyses.

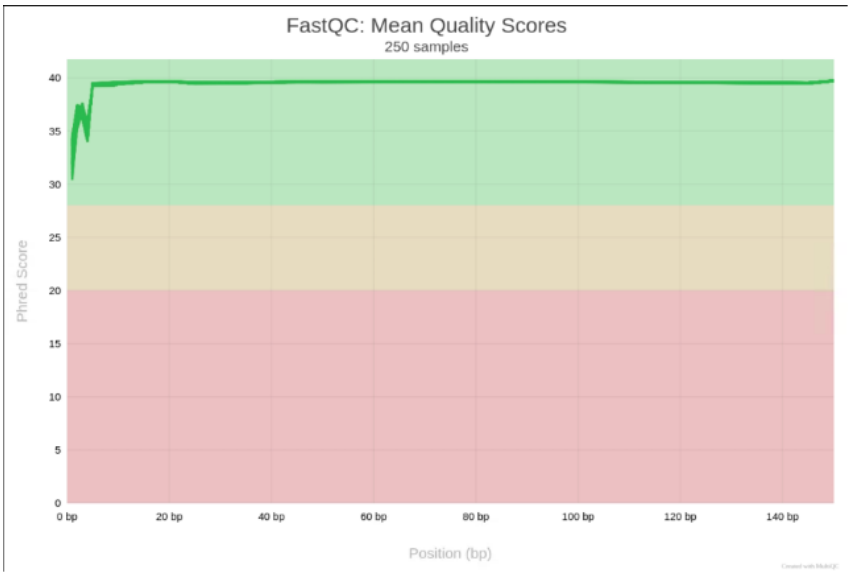


Figure 3. Mean quality scores of sequencing reads assessed using MultiQC.

Future Plans

With the successful optimization of ddRAD library preparation and the completion of sequencing and data analysis, the next step is to proceed with eDNA metabarcoding library preparation. This will enable the integration of both approaches, allowing for a more comprehensive assessment of biodiversity at both the community and population levels. Future work will also focus on optimizing bioinformatics pipelines for efficient data integration and interpretation.