

Microbial metabarcode database of Mekong river

Programme:BIFA

Project ID: BIFA6_030

Project lead organization:National Center for Genetic Engineering and Biotechnology

Project implementation period:2/9/2021 - 28/2/2023

Report approved: 26/9/2022

Narrative Midterm report

Executive Summary

The project "Microbial metabarcode database of Mekong river" has aimed to create a comprehensive microbial biodiversity database and share the biodiversity information of microorganisms in Mekong river derived from environmental-DNA (eDNA) data. To accomplish these goals, the project has been compiling, analysing eDNA metabarcode sequence, conducting data standardization and data sharing with GBIF. So far, the eDNA 16S rRNA metabarcoding sequence data from Mekong river have been analyzed. The initial analyses showed the temporal and taxonomic composition in the Mekong river's bacterial communities. The analysis results will be cleaned and re-translated to occurrence-level data. The eDNA-derived occurrence dataset and sampling metadata with Darwin Core terms were published the first dataset to GBIF.org. Moreover, in order to raise awareness of importance of eDNA data for Mekong biodiversity study and conservation in local communities around Mekong river, series of eDNA data analysis and data mobilization workshops had been conducted and will be conducted at various institutes and universities in Northern and Northeastern Thailand. Since the beginning of 2022, the eDNA data analysis and data mobilization workshops had been held 3 times in Thailand. The lectures about eDNA data analysis and diversity data mobilization had been given to the workshop participants. These two workshops have successfully attracted audience from university students, lecturers, researchers, and research organization, paving the foundation of an eDNA-based diversity study network in Mekong local communities in Thailand.

Progress against milestones

Has your project published at least one dataset through GBIF.org?: Yes

Dataset published:

Dataset	DOI
Lancang-Mekong microbiome project 16S rRNA metabarcode	10.15468/vkpxzv
Lancang-Mekong microbiome project sampling	10.15468/uhd963

Has at least one member of your project team received certification following the BIFA capacity enhancement workshop?: Yes

Name of the workshop participant:Somsak Likhitrattanapisal

Certification obtained: Basic Badge

Report on Activities

Activity progress summary

Metagenomic data analysis and Database construction

The first and second raw eDNA datasets of MekongDNA project's 16S rRNA metabarcoding have been analyzed. The diversity of Mekong river's bacteria in the metabarcoding sequences was analyzed. The taxonomic verification of these bacterial sequences is in progress. The initial analyses showed the temporal and taxonomic compositions in the Mekong river's bacterial communities, which comprised more than 1,000 species-level OTUs in each spatial site and time of collection. The analysis results were cleaned and re-translated to occurrence-level data. The database system has been developed to comprehensively store the microbial DNA metabarcode sequences, OTU taxonomic and abundance data, and occurrence data for efficient data management. The diversity of bacterial OTUs was formatted to the standardized eDNA-derived occurrence data and sampling metadata with Darwin Core terms and publish the first dataset to GBIF.org on 22 August

2022. The full dataset should be complete within the project duration.

1st, 2nd and 3rd eDNA workshops

The 1st, 2nd and 3rd eDNA workshops entitled “eDNA data and their importance in biomonitoring of Mekong river microbial diversity” have been held on 19-20 January 2022 at Chiang Rai Heritage Hotel, on 23-24 March 2022 at Ubon Ratchathani University, and on 16-17 June 2022 at Mahidol University Amnat Chareon Campus, respectively.

The main activity of the workshops consisted of 4 chapters of lecture session; 1) an introduction to eDNA-based biomonitoring of aquatic microbial community, 2) introduction to eDNA data analysis, and 3) eDNA for Mekong river bioindicator and 4) eDNA-based diversity data mobilization. After the lecture session, all workshop participants joined the Q&A session in which they discussed and shared ideas and experiences about diversity studies and conservation monitoring in Mekong river.

Web development and Data sharing API development

The project workshop activity has been publicised in MekongDNA’s Citizen Science program <https://www.mekongdna.org/citizen.php>. The data sharing data API will be developed based on our in-house AmiBase data API <https://amibase.org/api.php>. The API backend and data fields is being updated to support the standardization with Darwin Core terms and extensions for eDNA-derived occurrence data.

Completed activities

Activity name: The 1st and 2nd workshop

Description: The 1st workshop of the project was held on 19-20 January 2022 at Chiang Rai Heritage Hotel, Chiang Rai Province, Thailand. 23 participants, including students, university lecturers, teachers, and officials from Chiangsaen Hydrology Station, attended the first workshop.

The second workshop, which was collaborated with Faculty of Science, Ubon Ratchathani University, took place on 23-24 March 2022 at U PLACE convention hall, Ubon Ratchathani University, Ubon Ratchathani Province, Thailand. 57 participants, including university students and lecturers, attended the second workshop. The third workshop took place on 16-17 June 2022 at Mahidol University, Amnat Chareon Province, Thailand. 59 participants, including university students and lecturers, attended the third workshop

Start Date - End Date: 19/1/2022 - 17/6/2022

Verification Sources: <https://www.mekongdna.org/citizen.php>

Report on Deliverables

Deliverables progress summary

The initial analyses of eDNA from Mekong river have been conducted. The taxonomic assignment have been performed to identify OTUs. The diversity data from the analysis results were cleaned and compiled into the database. The first dataset then were formatted and submitted to GBIF using the cloud IPT tool.

In addition, one of the project’s deliverables and data publication routes is the data sharing API, which is being developed from the existing AmiBase data API. The plan is to create the computerized tools for managing and translating eDNA metabarcode data to occurrence data. The data fields of AmiBase API will also be expanded to support eDNA-based microbial diversity data. The additional metabarcode-related data fields include marker genes, primers, etc. The data fields will be mapped to Darwin Core terms following the GBIF guideline for publishing DNA-derived data. In the future, the data used in the project will be deposited to EMBL European Nucleotide Archive (ENA; <https://www.ebi.ac.uk/metagenomics/>) or NCBI GenBank (<https://www.ncbi.nlm.nih.gov/genbank/metagenome/>) via MGnify platform or other available tools, as proposed.

Progress towards deliverables

Title: Microbiome database of Mekong bacteria and fungi

Type: Dataset

Status update: The first dataset of bacterial occurrence data and sampling metadata (with 82 records) are published to GBIF.org within August 2022.

Dataset scope: Microbial metabarcode database (Bacteria and Fungi)

Expected number of records: 10000

Data holder: BIOTEC

Data host institution: BIOTEC

Sampling method: Metabarcoding of eDNA (primarily 16S rRNA and ITS metabarcoding)

% complete: 15

DOI:

Expected date of publication: 2023-02-06

Title: Web and Data sharing API

Type: Other

Description: The project workshop activity has been publicised in MekongDNA's Citizen Science program <https://www.mekongdna.org/citizen.php> . The data API data API will be developed based on AmiBase data API <https://amibase.org/api.php> . The API data will be updated and standardized with Darwin Core terms.

Sources of verification: <https://www.mekongdna.org/citizen.php> <https://amibase.org/api.php>

Events

The 1st, 2nd and 3rd eDNA data analysis and data mobilization workshops

Dates: 2022-01-19 - 2022-06-17

Organizing institution: BIOTEC

Country: Thailand

Number of participants: 137

Comments:

Website or sources of verification: Report Attachment: Section 1

Events

Workshop coordination meeting at Chiangkhan Hydrology Station

Dates: 2022-03-15 - 2022-03-15

Organizing institution: BIOTEC

Country: Thailand

Number of participants: 5

Comments:

Website or sources of verification: Report Attachment: Section 2

Communications and visibility

The principal investigator and partners regularly discussed the status of the project. The discussion topics included the metabarcoding analysis and data management tool development. The project activity updates have been published in the website www.MekongDNA.org in order to increase the project visibility. The project news and information has also been posted to the social network such as TBRCNetwork Facebook page (example:

<https://www.facebook.com/tbrcnetwork/posts/pfbid0gh71WwXM1DHBMcvCsz82bDuB271BKSSAFaf4USQxgXEmJvFRnCmAVU5K3>

Monitoring and evaluation

Monitoring and evaluation findings

The manual and computerized-script monitoring have been performed in order to ensure that the to-be-imported cleaned data are accurately matched and in sync with datasets of the data source, mainly the eDNA data of MekongDNA project. The monitoring will also check taxonomic assignment and status of the scientific name, geolocation metadata, and time of collection given to each records. The first stage of data API development has been implemented and tested with the data example so far. Moreover, the workshop satisfaction level are evaluated by responses and feedbacks from workshop participants and collaborators. The past two workshops had already been reviewed as "satisfactory".

Impact of COVID-19 pandemic on project implementation

Due to COVID-19 outbreaks in Thailand, the domestic travel between different regions was heavily restricted, causing some delays in the workshop events. In the workshop venues, COVID-19 protection measures, ATK test kit, face masks, and alcohol sprays, also had to be strictly implemented, thus causing the additional costs to the project's budget. The numbers of workshop participants in each workshop were smaller than expected because many local students, lecturers, teachers and researchers were being quarantined or infected with COVID-19 at the time of the workshops. Many participants requested for the expansion of the workshop to live online events.

GBIF leads the Biodiversity Information Fund for Asia (BIFA), a programme funded by the Ministry of the Environment, Government of Japan. The programme provides supplementary support for activities addressing the needs of regional researchers and policymakers through mobilization and use of biodiversity data.



