

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: 22/3/2023

Narrative Final report

Executive Summary

The project "Microbial metabarcode database of Mekong river" has analysed environmental-DNA (eDNA) metabarcode sequence data from the Mekong River's water. The bacterial diversity derived from 16S rRNA metabarcode sequence data showed more than 10,000 occurrences of identified bacteria in the Mekong River samples collected during March - December 2021. This occurrence dataset were compiled, standardized, and submitted to GBIF occurrence dataset accompanied by the sampling event records, which can be accessed at <https://doi.org/10.15468/8bfgdv> . In order to achieve the goal of sharing the eDNA-derived biodiversity information of microorganisms in Mekong river, the metabarcode sequence data were also deposited to NCBI Bioproject database. The diversity and occurrence datasets had been integrated to AmiBase system and the data sharing API was developed.

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. Throughout the project timeline, four workshops of the eDNA data analysis and data mobilization workshops had been held, attracting a total of 250 participants including a variety of audience from university students, lecturers, researchers, and R&D personnels in private sectors. The success of these workshop activities are hopefully paving the foundation of an eDNA-based diversity study network in Thailand.

Progress against milestones

Has your project completed all planned activities?: Yes

Has your project produced all deliverables?: Yes

Report on Activities

Activity implementation summary

Metagenomic data analysis and Database construction

The eDNA datasets of MekongDNA project's 16S rRNA metabarcoding have been analyzed in order to assess the diversity of Mekong river's bacteria in the water samples collected during March - December 2021. The analyses showed the 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, filtered and re-formatted to GBIF occurrence dataset. 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 final eDNA-derived occurrence dataset and sampling event metadata were formatted with Darwin Core terms and published GBIF database on 28 February 2023.

The eDNA data analysis and management workshops

Throughout the project duration, four DNA workshops had been held. The first 3 workshops titled "eDNA data and their importance in biomonitoring of Mekong river microbial diversity" took place on;

1. 19-20 January 2022 at Chiang Rai Heritage Hotel
2. 23-24 March 2022 at Ubon Ratchathani University
3. 16-17 June 2022 at Mahidol University Amnat Chareon Campus,.

The main activity of these 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.

The final workshop was conducted in conjunction with the WDCM global training course: ASEAN satellite training course titled "eDNA for Biomonitoring in Environments and Mycobiome: Importance and its Applications". It was held as a hybrid seminar including onsite and online sessions during 28-29 November 2022 at Swissotel Hotel in Bangkok, Thailand.

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 was developed based on our in-house AmiBase data API <https://amibase.org/api.php>.

Completed activities

Activity name: Metagenomic data analysis

Description: Microbiome diversity analysis from metabarcoding sequence data

Start Date - End Date: 2/9/2021 - 24/10/2022

Verification Sources: SRA data were submitted to NCBI Bioproject database

<https://www.ncbi.nlm.nih.gov/bioproject/848859>

Activity name: Database construction

Description: Integrating metabarcoding and analysed results into a comprehensive database

Start Date - End Date: 2/9/2021 - 30/11/2022

Verification Sources: The database has been constructed and integrated into AmiBase database.

The data in BIOM format will be released publicly after the manuscript is accepted to be published in the journal https://www.amibase.org/microbiome_download.php

Activity name: 1st training workshop

Description: 1st training workshop on metabarcoding sample techniques and eDNA data management

Start Date - End Date: 19/1/2022 - 20/1/2022

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

Activity name: 2nd training workshop

Description: 2nd training workshop on metabarcoding sample techniques and eDNA data management

Start Date - End Date: 22/3/2022 - 24/3/2022

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

Activity name: 3rd training workshop

Description: 3rd training workshop on metabarcoding sample techniques and eDNA data management

Start Date - End Date: 15/6/2022 - 17/6/2022

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

Activity name: 4th training workshop

Description: 4th training workshop and seminar on metabarcoding sample techniques and eDNA data management

Start Date - End Date: 27/11/2022 - 29/11/2022

Verification Sources: https://www.mekongdna.org/detail_workshop-2022-11-29.php

Activity name: Web development and Data sharing API development

Description: Web-based data publishing and data sharing API development for microbial eDNA and diversity data

Start Date - End Date: 15/1/2022 - 30/1/2023

Verification Sources: <https://www.amibase.org/api.php>

Activity name: Publication and reports

Description: Data publication to GBIF and report submission

Start Date - End Date: 10/8/2022 - 28/2/2023

Verification Sources: Occurrence data are published under the title

Lancang-Mekong microbiome project 16S rRNA metabarcode (March - December 2021) to GBIF sampling dataset UUID: 2df7ae37-2ff6-4328-92bb-e17b7028dd75 (<https://doi.org/10.15468/8bfgdv>).

Report on Deliverables

Production of Deliverables - Summary

The analyses of eDNA from the Mekong river samples collected during March - December 2021 were

conducted. The taxonomic assignment have been performed to identify OTUs. The diversity data from the analysis results were cleaned and compiled into the AmiBase database.

The occurrence dataset comprising 12,614 records as well as the accompanying sampling event dataset were formatted with Darwin Core terms for DNA-derived data and submitted to GBIF using the cloud IPT tool. The occurrence data are published under the title "Lancang-Mekong microbiome project 16S rRNA metabarcoding (March - December 2021)" to GBIF dataset UUID: 2df7ae37-2ff6-4328-92bb-e17b7028dd75 (<https://doi.org/10.15468/8bfgdv>). And the sampling events of the Lancang-Mekong microbiome project are published to GBIF sampling dataset UUID: 77c5d29c-2c6e-4d01-bc66-eb23c274c73f (<https://doi.org/10.15468/8bfgdv>).

In addition, one of the project's deliverables and data publication routes is the data sharing API, which was developed from the existing AmiBase data API. The 16S rRNA metabarcoding sequence data used in the project were deposited to NCBI Bioproject database (<https://www.ncbi.nlm.nih.gov/bioproject/848859>). The manuscript of these bacterial diversity data was prepared and submitted to the scientific journal. The manuscript is currently being under a peer-review process. The deposited bioproject will be published when the manuscript is accepted.

Production of deliverables

Title: Data sharing API

Type: Other

Description: The data API has been developed to retrieve microbial diversity data from AmiBase system

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

Title: Final report

Type: Other

Description: The final report was submitted to GBIF.

Sources of verification: The final report was submitted to GBIF.

Title: Mekong bacterial diversity database and occurrence dataset

Type: Dataset

Status update: The diversity data from the analysis results were cleaned and compiled into the AmiBase database. The occurrence dataset comprising 12,614 records as well as the accompanying sampling event dataset were formatted with Darwin Core terms for DNA-derived data and submitted to GBIF using the cloud IPT tool. The occurrence data are published under the title "Lancang-Mekong microbiome project 16S rRNA metabarcoding (March - December 2021)" to GBIF dataset UUID: 2df7ae37-2ff6-4328-92bb-e17b7028dd75 (<https://doi.org/10.15468/8bfgdv>). And the sampling events of the Lancang-Mekong microbiome project are published to GBIF sampling dataset UUID: 77c5d29c-2c6e-4d01-bc66-eb23c274c73f (<https://doi.org/10.15468/8bfgdv>).

Dataset scope: Bacterial occurrence data derived from eDNA samples collected March - December 2021}

Expected number of records: 12614

Data holder: Thailand Bioresource Research Center

Data host institution: Thailand Bioresource Research Center

Sampling method:

% complete: 100

DOI: <https://doi.org/10.15468/8bfgdv>

Expected date of publication:

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. At the 4th workshop, the workshop and seminar were conducted in a hybrid fashion including both onsite and online sessions.

Events

The 1st eDNA data analysis and data mobilization workshop

Dates: 2022-01-19 - 2022-01-20

Organizing institution: Thailand Bioresource Research Center

Country: Thailand

Number of participants: 23

Comments:

Events

The 2nd eDNA data analysis and data mobilization workshop

Dates: 2022-03-23 - 2022-03-24

Organizing institution: Thailand Bioresource Research Center

Country: Thailand

Number of participants: 57

Comments:

Website or sources of verification: Report Attachment: Section 1

Events

The 3rd eDNA data analysis and data mobilization workshop

Dates: 2022-06-17 - 2022-06-17

Organizing institution: Thailand Bioresource Research Center

Country: Thailand

Number of participants: 58

Comments:

Website or sources of verification: Report Attachment: Section 1

Events

The 4th workshop and the seminar “eDNA for Biomonitoring in Environments and Mycobiome: Importance and its Applications”

Dates: 2022-11-28 - 2022-11-29

Organizing institution: Thailand Bioresource Research Center

Country: Thailand

Number of participants: 112

Comments: This workshop and seminar was conducted in conjunction with the WDCM global training course: ASEAN satellite training course. The number of participants include both onsite and online sessions.

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, especially the eDNA workshop series, have been published in the MekongDNA project website (www.MekongDNA.org) in order to increase the project visibility. The workshop events were shown at the Citizen Science page of MekongDNA.org website <https://www.mekongdna.org/citizen.php>. The project news and information has also been posted to the social network such as TBRCNetwork Facebook page.

For example;

The news post announcing the start of the project

<https://www.facebook.com/tbrcnetwork/posts/pfbid0gh71WwXM1DHBMcVcsz82bDuB271BKsAFaf4USQxgXEmJvFRnCmAVU5K3>

The news post summarizing the success of the 4th workshop and seminar

<https://www.facebook.com/tbrcnetwork/posts/pfbid0EBaaGMzc1LocJpVTkCgN57285XLQ5TrLBYn2hAvCsFJL9b3x4ekhv9AFySSzI>

Monitoring and evaluation

Final Evaluation

The computerized scripts and tools have been employed to detect any discrepancies in the resulted eDNA-derived diversity dataset. The taxonomic assignment and status of the scientific name, geolocation metadata, and time of collection given to each records were standardized with the widely-accepted taxonomy databases and conventions. Besides, after the data have been formatted into the occurrence dataset, they were manually checked again to ensure accuracy and consistency. The project has successfully reached the goal of making Mekong microbial occurrence dataset with more than 10,000 records.

The workshop satisfaction level were evaluated by responses and feedbacks from workshop participants and collaborators. The majority of participants had rated the workshops to be "useful", "relevant", and "concise". The 4th workshop and seminar was indeed successful, so that one of the participants from Mahasarakam University contacted the workshop team to plan to develop eDNA-

related research project and graduate curriculum in the university.

Best Practices and Lessons Learned

At the beginning of the project, GBIF was held the online "Data Mobilization Course". The course was not only informative about diversity data analysis, formatting, management which are very critical to the project's success, but also very helpful in project implementation and management. The hands-on experience in after-course exam gave the opportunity to explore and try the GBIF IPT tools with real examples.

The GBIF coordinators was also warm and kind. Thanks to Mr. Chihjen Ko's technical support and advice, the occurrence dataset had been properly submitted. The skill learned from these experiences helped in developing the computer-programmed scripts to format the occurrence dataset, which is essential for the large dataset with more than 10,000 records.

Post Project Activity(ies)

Post project activities include;

1. The project is going to update on the published Mekong river occurrence datasets (if any).
2. After the manuscript is accepted by a scientific journal, the diversity dataset of Mekong river bacteria in the AmiBase and the deposited NCBI Bioproject will be opened to public access.

Sustainability plans

The project partners in Thailand, Dr. Supawadee Ingsriswang and Dr. Suprapap Patshinghasanee, have been collaborated in data acquirement, analysis, and project coordination. As the project has successfully interested many university lecturers and researchers in governmental organizations in Thailand's Mekong region. These partners are the key to establish the network of eDNA study in river and freshwater in Thailand. Mr. Christian Elloran from ACB has collaborated and helped in developing data sharing API. The resulted data sharing protocol is going to be useful in connecting eDNA-derived microbial diversity in the Mekong river to ACB's ASEAN biodiversity database and Clearing House Mechanism.

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.

