

DNA barcoding Asian Pteridophytes (1): 1.5KP DNA-referenced Collections from Taiwan and Southeast Asian Countries

Programme:BIFA

Project ID: BIFA6_010

Project lead organization:Institute of Molecular & Cellular Biology, National Tsing Hua University

Project implementation period:1/9/2021 - 28/2/2023

Report approved: 21/3/2023

Narrative Final report

Executive Summary

In this project, we reevaluated promising DNA barcode regions in ferns, and selected *rbcl* and *trnL-L-F* for the downstream works as mentioned in the Midterm report. Finally, we sequenced > 1500 samples of Asian fern collections. Most of them have been generated with both barcode regions. In the latest uploaded datasets, a total of 1536 vouchered occurrences includes 828 from Taiwan, 302 from Vietnam, 224 from Malaysia, 184 from the Philippines, 7 from China, one from India. They are sequenced with *rbcl*. We also demonstrated/introduced our newly developed NGS pipeline, PowerBarcoder, in the workshop in Taiwan. Our study provides one of the best practices to "high throughput" DNA barcode sequences for fern collections, that will be very useful for DNA-based identifications for these organisms. In addition, we published three new fern species in two systematic papers based on the new findings from these DNA sequences. These certainly broaden our understanding of fern biodiversity from the DNA to species level.

There remains two challenges in our project. First, even aided with the sequence information, taxonomy of certain fern groups is still difficult. We need additional efforts to make their identification reliable to a species level. Second, the NGS pipeline for the DNA barcode region, *trnL-L-F*, needs to be verified moderately. To fix above issues, we are inviting more fern systematists to work on poorly identified collections, and testing new pipelines to process NGS reads of *trnL-L-F*. By such improvements, we plan to publish the data paper with polished identification results, and also prepare systematic manuscripts for any novel taxa.

Progress against milestones

Has your project completed all planned activities?: No

Rationale: Because of the COVID situation, the Asian fern symposium didn't take neither in a physical nor online manner at 2022. As a result, we didn't hold the international workshop to demonstrate our barcoding works. *TrnL-L-F* region was already sequenced by NGS platforms, but these current outputs contained unreliable results. We therefore not yet included these sequences in the latest datasets. It is likely due to tandem repeats in ~25% sequences of our collections, and our current NGS pipeline seems sensitive to such PCR amplicons. The manuscript of data paper is not ready yet, and we hope to cover better identification results of our collections as well as *trnL-L-F* sequences.

Has your project produced all deliverables?: No

Rationale: Currently, the sequences are open and accessible through our uploaded datasets on GBIF, but not in GenBank and BOL. After obtaining reliable identification results, these sequences will be submitted to these platforms, and we can then complete the data paper too.

Report on Activities

Activity implementation summary

"Selecting 1.5KP Asian collections and their DNA work" and "Reevaluating previously proposed DNA barcodes" are completed as already mentioned in the Midterm report. The important details of "Designing NGS protocol for DNA barcoding" are provided in the PowerBarcoder github,

our method paper published in "Application in Plant Science", and the online slides in the Midterm report. We had "Within country workshop to broadcast the workflow & DNA barcode information". "Compiling 1.5KP dataset & data mobilization" is almost completed, excepting that some identification may need reevaluation and all occurrences/sequences data are currently available in GBIF.

We published additional two systematic papers with three newly described fern species using these materials and DNA barcode sequences. The process in our DNA work was delayed due to COVID situation, the downstream analyses for dataset compiling were delayed too. Similarly, the within country workshop was postponed to the October 2022. After widely sequencing fern collections, we found that our pipeline is not perfect for certain barcoding region, such as those with tandem repeats in trnL-L-F. We now try to fix bugs, and prepare new pipelines in our PowerBarcoder. We therefore slightly reordered our schedule, and the data paper is still in preparation.

Completed activities

Activity name: Selecting 1.5KP Asian collections and their DNA work

Description: To fill in taxonomic and geographic gaps currently in GBIF, we first select ~600 of collections from Taiwan, and another ~900 from other countries in Southeast Asia. Importantly, DNA data are still scarce for pteridophytes in these regions. We then extract DNA, PCR, and DNA-barcode these 1.5KP collections from the TPG DNA bank.

Start Date - End Date: 1/9/2021 - 28/2/2022

Verification Sources: <https://doi.org/10.15468/dcv5gh>
<https://doi.org/10.15468/cqucep>
<https://doi.org/10.15468/nhsnn2>

Activity name: Reevaluating previously proposed DNA barcodes

Description: Since DNA barcodes for pteridophytes are still inconclusive, we need to reevaluate performances of these barcodes (i.e. sequencing success rate, quality and discrimination power) based on a subset of 200 collections, and finally choose two of them for subsequent works.

Start Date - End Date: 1/10/2021 - 1/11/2021

Verification Sources: https://drive.google.com/file/d/1Z_RKMBHb9xRaLd_cHtW-tbLNGCFqHvRe/view?usp=share_link

Activity name: Compiling 1.5KP dataset & data mobilization

Description: With collaboration with partners from TAIF herbarium and Southeast Asian countries, we will gather voucher information, images, and DNA barcodes for each of the 1.5KP collections, and compile all into datasets for GBIF and other biodiversity platforms.

Start Date - End Date: 1/11/2021 - 23/2/2023

Verification Sources: <https://doi.org/10.15468/dcv5gh> <https://doi.org/10.15468/cqucep>
<https://doi.org/10.15468/nhsnn2>

Activity name: Designing NGS protocol for DNA barcoding

Description: Considering a cost-effective way to PCR/sequence a great number of highly diverse pteridophytes in Asia, it's essential to design universal primer sets and index these primers for runs of multiplexed, amplicon NGS sequencing.

Start Date - End Date: 1/11/2021 - 1/12/2021

Verification Sources: https://drive.google.com/file/d/1Z_RKMBHb9xRaLd_cHtW-tbLNGCFqHvRe/view?usp=share_link
<https://doi.org/10.1002/aps3.11462>
<https://github.com/PowerBarcoder>.

Activity name: Within country workshop to broadcast the workflow & DNA barcode information

Description: Communicating with research community with similar aims in Taiwan. Sharing experience of our standardized workflow/protocol. The workshop was co-organized with "National Museum of Natural Science", "Taiwan Society of Plant Systematics" and TaiBIF.

Start Date - End Date: 1/10/2022 - 1/10/2022

Verification Sources: https://fb.watch/iS4l_qlomb/

Activity name: Complete missing data & result double confirmation

Description: For any missing data, we will complete them during this stage. We will also perform some analyses to exclude any contaminated results or misidentification.

Start Date - End Date: 1/4/2022 - 15/2/2023

Verification Sources:

Report on Deliverables

Production of Deliverables - Summary

We submitted three occurrence datasets with a total of 1536 vouchered specimens, including 828 from Taiwan, 302 from Vietnam, 224 from Malaysia, 184 from the Philippines, 7 from China, one from India. They all have rbcL sequences.

In addition, we generated a NGS pipeline which is available in github (<https://github.com/PowerBarcoder>). We also published two systematic papers with three newly described fern species in Asia (DOI: 10.3897/phytokeys.187.77035, 10.3897/phytokeys.204.86045), and one method paper (DOI: 10.1002/aps3.11462). By incorporating these new datasets and workflow/pipeline, we also made a good process to carry out DNA identification works of fern gametophytes in fields (e.g. <https://2022.botanyconference.org/engine/search/index.php?func=detail&aid=436>).

Production of deliverables

Title: Improving knowledge of Asian pteridophytes through DNA sampling of specimens in regional collections

Type: Dataset

Status update: The ferns, lycophytes and seed-free vascular plants commonly described as pteridophytes exhibit hyperdiversity in the insular vegetation that often characterizes Asian floras. Despite harboring biodiversity hotspots, these plants and their georegions have been poorly surveyed, particularly in Southeast Asia, where one third of the world's pteridophyte species are concentrated. More than 60 per cent of the approximately 4,500 species lack georeferenced records in GBIF and only 6 per cent have been DNA barcoded. This project aims to increase the available knowledge on Asian pteridophytes by compiling a georeferenced occurrence dataset that includes images, DNA barcodes and other vouchering information from thousands of recent collections, building on the efforts of the Taiwan Pteridophyte Research Group and its Southeast Asian collaborators. The project team will set up a workflow incorporating next-generation sequencing for 1,500 Asian pteridophyte specimens from selected collections in Taiwan, Vietnam, the Philippines, Malaysia and other Southeast Asian countries that can fill in taxonomic and geographic gaps and represent Asian pteridophyte diversity. Mobilization of and access to these vouchered and georeferenced DNA-derived records will advance further research into the biogeography of pteridophytes and other terrestrial vegetation and support the development of novel approaches to monitor biodiversity along the spatiotemporal scale, including metabarcoding of the invisible diversity held in soil and spore banks.

Dataset scope: Certain areas among South East Asia}

Expected number of records: 1494

Data holder: Li-Yaung Kuo & Chien-Wen Chen

Data host institution: Institute of Molecular & Cellular Biology, National Tsing Hua University; Herbarium of Taiwan Forestry Research Institute

Sampling method: Floristic survey.

% complete: 100

DOI: 10.15468/nhsnn2

Expected date of publication:

Title: two_new_asian_lomariopsis

Type: Dataset

Status update: Two East Asian Lomariopsis (Lomariopsidaceae, Polypodiales) species, *Lomariopsis moorei* and *Lomariopsis longini*, which were previously misidentified as *L. spectabilis*, are here described as new species based on evidence from morphological characters and a molecular phylogeny. The two species differ from the three other described species in East Asia by their venation, pinna shapes, and perine morphology. A phylogeny based on a combined dataset of three chloroplast regions (rbcL+ rps4-trnS + trnL-L-F) showed that *L. moorei* and *L. longini* each formed a well-supported monophyletic group which was distantly related to both *L. spectabilis* and the other morphologically similar East Asian species, *L. boninensis*.

Dataset scope: providing occurrence records of two newly described ferns and their DNA barcode sequences }

Expected number of records: 16

Data holder: Li-Yaung Kuo
Data host institution: Taiwan Biodiversity Information Facility (TaiBIF)
Sampling method: Specimen collections
% complete: 100
DOI: 10.15468/cqucep
Expected date of publication:

Title: new_hymenophyllum_from_taiwan

Type: Dataset

Status update: Hymenophyllum chamaecyparicola T.C.Hsu & Z.X.Chang, a new filmy fern species (Hymenophyllaceae) has been described from Taiwan and illustrated based on morphological and phylogenetic evidence. Although the new species resembles members in the subgenus Mecodium, namely H. wrightii, our plastid phylogeny has revealed that it is genetically distant from H. wrightii and forms a clade nested within subg. Hymenophyllum. The most notable characteristic to differentiate H. chamaecyparicola from related species is the presence of minute spathulate hairs on the surface of the rachis and veins. Hymenophyllum chamaecyparicola is currently only known from a small area in northern Taiwan, and endemic to that country.

Dataset scope: providing occurrence records of the newly described fern and their relatives; and their DNA barcode sequences}

Expected number of records: 25

Data holder: Li-Yaung Kuo

Data host institution: Taiwan Biodiversity Information Facility (TaiBIF)

Sampling method: Specimen collections

% complete: 100

DOI: 10.15468/dcv5gh

Expected date of publication:

Impact of COVID-19 pandemic on project implementation

Because of COVID-19, the DNA bench work in the lab and re-identification of voucher specimens in the herbarium were delayed. Because we hope to improve specimen identification results before publishing in a peer-reviewed journal, the timeline of the data paper is delayed too.

Events

Next generation DNA barcode

Dates: 2022-10-01 - 2022-10-01

Organizing institution: TaiBIF, National Museum of Natural Science, Taiwan Society of Plant Systematics,

Country: Taiwan

Number of participants: 26

Comments: Because of COVID situation in Taiwan, 26 participants attended in person. More joined the talks of workshop online.

Website or sources of verification: https://fb.watch/iS4l_qlomb/

Communications and visibility

All occurrences in our datasets contain DNA barcode sequence information, and these datasets can be freely accessible via GBIF. Most importantly, if any GBIF user notifies any sequenced collection needing a taxonomic reevaluation, he or she can feed back to GBIF as well as to the research community. In our plan for the data paper, these GBIF datasets will be highlighted as the main resource of DNA barcodes for Asian ferns. We hope, in the near future, more users, not only for systematists but also ecologists, can explore this DNA resource through GBIF, and use these reliable databases for their researches needing DNA-based identification. Regarding our workflow and pipeline, they are freely accessible from our method paper and the scripts on github. Their links are informed in the GBIF website too (e.g. in the midterm and final reports). Hopefully, this cost-effect approach for DNA-barcoding can accelerate DNA-referenced works during floristic surveys in record-poor areas (e.g. Southeastern Asia and South Pacific regions). We are looking forward to new GBIF projects providing contributors to initiate such works.

Monitoring and evaluation

Final Evaluation

The current outputs of this project, including the pipeline, datasets, and DNA collections, are indeed attractive to researchers from different fields. Not only fern researchers, but also museumists and systematists working on non-fern organisms attended our workshop, and provided their positive feedback to our high-through put NGS workflow. For fern researchers, particularly taxonomists, they are most interested in our datasets associated with DNA sequence information, which is important to provide further systematic insights into Asian fern diversity. Moreover, based on these outputs, we are looking forward to filling in gaps of Asian fern biodiversity, which has been poorly recorded and sequenced.

Best Practices and Lessons Learned

Traditionally, DNA barcoding works are labor intensive and costly, but remain important. Nonetheless, the updated NGS workflow provides a much cost-effective approach to complete this mission. Although our current pipeline is not perfect to deal with NGS sequences of certain genetic regions (e.g. trnL-L-F), further modifications and tests shall solve such difficulties. While, even with sequence information, expertise from systematists remain very important in order to provide reliable identification linking to correct scientific names.

Post Project Activity(ies)

Based on the uploaded dataset, we are timely reviewing the identification of our collections. In the near future, the improved identification results will be updated in our GBIF datasets, and incorporated into our data paper too. Afterward, we plan to submit the data paper to "Biodiversity Data Journal". At the same time, we are now testing new pipelines to solve issues of trnL-L-F NGS sequencing. After fixing it, we are going to prepare another methodology paper to introduce our full workflow/pipeline. In addition to these, we are also preparing systematic papers to describe new fern taxa found in our dataset, that will finally provide formal taxonomic names for these previously unaware Asian species.

Sustainability plans

We hope that these fruitful results, including promising undescribed species, can facilitate international collaborations to survey fern diversity in different Asian regions. Our partners from different countries can simply follow our approach to continue the regional investigation of fern floristics. Further NGS workflow and pipeline can be carried out if they collaborate with a molecular lab at their own countries or internationally with my lab. We also initiated further collaborations with different fern systematics in order to solve taxonomic issues found among these fern collections. By efficiently and continually collecting DNA barcodes of ferns, we also hope to setup new collaborative researches with ecologists, who can apply a molecular approach into their advanced ecological research.

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.

