Implementation of Systematic Access to Isolation Sources of Microbes in a Culture Collection System

Ezra Earl P. Villanueva and Mylah Rystie U. Anacleto

Abstract— **Limited accessible microbial databases and information on microbial communities in cave environments in the Philippines have created a need for a solution. Caves isolation source web application, was developed to provide a database for the microbial data and metadata collected from caves in CALABARZON. It highlights the usage of an internationally accepted Microbial Isolation Source Ontology (MISO) 3-level categories to describe microbial origins and environmental information. Additionally, it features an interactive map to display strain count information from specific areas or sampling sites. The application's usability was well-received, achieving an above-average System Usability Scale (SUS) score of 82.1875, indicating an acceptable level of usability.**

Index Terms— **bacterial metadata, bat, bat guano, culture collection, database management system, metadatabase**

I. INTRODUCTION

A. Background of the Study

Of more than 2500 recorded caves in the Philippines, only 18% are assessed and classified as of today [1]. Caves are important biodiversity and natural heritage assets for they are dwelling places for many diverse and ecologically valuable organisms like bats and microbial communities. Bats play many important ecological and economic roles in material and nutrient distribution, biological pest control through predation, bioindicators, and sources of guano fertilizers [2], [3]. The bat guano has an interesting interplay as the main resource for the development of various bacteria, fungi, protists, and arthropods [3]. Studies show that bats and bat guano are reservoirs of functional pathogenic and zoonotic agents that are potentially harmful to the bat population, other animals, and humans [4], [5]. Microorganisms are ubiquitous and may cause various diseases however in comparison with their great species variety and research domains, it is considered underexplored research [6], [7]. The collection of microbial data and metadata is essential to provide easy and reliable access for future research and development. In addition, studies on Philippine cave microbiology are necessary to comprehensively address safety and security protocols in bat guano mining, as well as sustainable cave conservation.

Over the years, the range of information on microbial data and metadata is usually scattered and confined in various scientific journals, books, culture collections, and private databases around the world [8]. To mobilize and integrate data from diverse sources and make it accessible into a meta-database, some present contributions developed are BacDive, GOLD, and NCBI databases. There are no or very limited accessible microbial databases available in the Philippines as well as information on microbial communities in cave environments. Hence, this study will present a local microbial database with the same approach. Furthermore, this study will adopt the controlled vocabulary *Microbial Isolation Source Ontology* (MISO) of BacDive and will provide visualizations through a map and Krona-plot to provide systematic access and analysis to the origin of microbial strains.

B. Significance of the Study

The study will provide a database of standardized and digitized data and metadata of the collected microbial strains from the 5 caves in CALABARZON, focusing on isolation source and environmental information. The database will mobilize and harmonize the scattered information to constitute structured database content with easy and reliable user access. This will encourage researchers and institutions to contribute data and metadata of their own microbial collections. Furthermore, the output database will promote collaboration among researchers, serve as a valuable resource for the academic community, and inspire further research in microbiology and related fields.

C. Objectives

The general objective of this study is to develop a database management system for the microbial data and metadata collected from 5 caves in CALABARZON. Specifically, the study aims:

• To provide a systematic access to the origin and environmental information of microbial strains through a search feature focused on isolation sources;

• To design a location-based microbial database for storage of bacterial metadata; and

• To utilize internationally-accepted controlled vocabularies such as Microbial Isolation Source Ontology (MISO) to categorize data in the database;

II. REVIEW OF RELATED LITERATURE

Microbial data and metadata are scattered in various sources, from databases to scientific literature. Large aggregations of microbial metadata are not published but stored in laboratory notes or private bacterial culture collections [8], [9]. Over the past years, dramatic progress has been made in the research on the diversity and function of microbial communities in the environment as a consequence of the advancement in high-throughput metagenomic sequencing technologies and methodological approaches [10]. Despite the increasing information from various sources, it remains underutilized for many reasons: data are found in a wide range of sources; data and metadata have various formats that prevent interoperability; much information is expressed in free text even in databases, and therefore difficult to access [9].

Present contributions have been made to centralize and standardize microbial metadata with structured databases. Bac-Dive– the Bacterial Diversity Metadatabase of Leibniz Institute DSMZ is regarded as one of the most extensive databases for standardized prokaryotic information on strain level in the world with over 80,000 strains at present. BacDive advanced web service offers RESTful API for large-scale and easy retrieval of data [7]. The Genomes OnLine Database (GOLD) of the Department of Energy Joint Genome Institute (DOE-JGI) is a curated collection of genome projects and metadata, featuring over 1.17 million entries across studies, organisms, biosamples, sequencing projects, and analysis projects [11]. Aside from maintaining comprehensive databases like GenBank and PubMed, National Center for Biotechnology Information (NCBI) also manages minor databases like the NCBI Taxonomy and NCBI Genome containing a vast amount of data and metadata covering a wide range of organisms including bacteria, viruses, and fungi [12]. The majority of the databases of NCBI are interlinked with each other and are publicly available.

The standardization efforts on the databases extend to microbe isolation classification to describe the geographic location and environment context from which the biological sample originated. The Genomic Standards Consortium (GSC) introduced minimum information about any (x) sequence (MIxS) and the 'environmental packages', offering a framework for microbial metadata providers and resource centers. This includes the geographic location and environmental information: biome, environmental feature, and environmental material [13]. BacDive developed Microbial Isolation Source Ontology (MISO), a hierarchical controlled vocabulary of 387 terms in three levels (e.g. #Environmental — #Terrestial — #Soil). BacDive isolation sources search implemented these terms into a tagging system to manually index strains [7]. Another, GOLD utilizes a five-level ecosystem classification system of nearly 800 distinct combinations, composing of microbial ecosystem, ecosystem category, ecosystem type, ecosystem subtype, and specific ecosystem (e.g. Host-associated — Mammals — Digestive System — Large Intestines — Fecal) [11].

Visualization is a vital aspect in the field of understanding microbial communities. Maps are the popular way of visualizing quantitative and geospatial data. Web mapping and the use of geospatial information have evolved over the years creating many tools, libraries, and applications for visualization. Leaflet.js is a widely adopted open-source JavaScript library used for creating interactive maps on web pages. It demonstrates efficient cross-platform compatibility that functions seamlessly across major desktop and mobile platforms, providing various mapping functionalities, including zooming, panning, adding markers, polygons, and overlays, as well as supporting tile layers from different map providers like OpenStreetMap [14]. Both GOLD and BacDive implement Leaflet.js for the display of distribution maps.

III. MATERIALS AND METHODS

A. Development Tools

The web application was created using MongoDB, Express, React, and NodeJS (MERN) development stack, following a Server-Client architecture. MongoDB serves as the relational database system while Express handles the web framework. React, a JavaScript library for building user interfaces, and NodeJS, an asynchronous event-driven JavaScript runtime environment, were utilized to create a scalable web application. The system was developed on a laptop computer with the following specifications:

- Operating System: Ubuntu 22.04.4 LTS 64-bit
- Processor: Intel® Core™ i3-4005U CPU @ 1.70GHz × 4
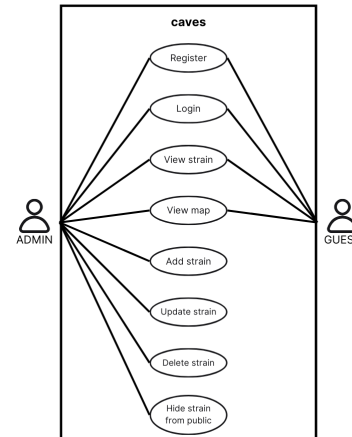- Memory: 4.0 GiB

Fig. 1: User case diagram

B. System Features

1) Isolation Source Modules: The system has the following isolation sources search module required to achieve the objectives of the study.

Source Map: displays strains' count and isolation sources based on location information.

Isolation Source Table: displays strains' isolation source information and its corresponding Microbial Isolation Source Ontology tags.

Isolation Source Statistics: displays charts based on strains' isolation information field.

2) User Level Modules: The system has a dedicated user interface and privileges for each user level.

Guest User: can only view and search strain/s information and may not need an account.

Admin User: can add, edit, and delete strain/s information as well as all the privileges of a Guest user.

C. Non-functional Requirements

Non-functional requirements (NFR) describe the capabilities and limitations of the services and features of the system. The system has the following NFRs:

• The system is web-based and should operate only in browsers like Chrome and Firefox.

• The system should be available at any time except for system maintenance or an unavailable Internet connection.

• The system will have a minimal user interface that is easy to learn and use by users of any technical background.

• Personnel knowledgeable of the MERN stack is necessary to fix, update, and maintain the system.

• The system should be highly maintainable for future development.

D. Data Population

NICER CAVES Project 4 provided the data that populated the web application. Isolation source information was patterned for each isolate produced on a single source or sample. For additional demo data to be displayed on public or guest view, some strain data from BacDive was inputted into the system.

IV. RESULTS AND DISCUSSION

A. Web Application

The Caves web application is currently hosted on free web hosting platforms, with both the backend and frontend accessible at https://isolationsources-caves.netlify.app/. As of the moment, the system contains 167 strain entries. Users can perform search functionalities and view data statistics. Admin users can add, update, and delete from their collection, as well as view strains that are not visible to the public.

The application was created to serve as an isolation source-based metadatabase. With this objective, the web application was designed to emphasize isolation sources of strains in its interface that can be easily understood and used by the users. The interface features a minimalist and modern design in contrast with foreign metadatabases. Tailwind CSS and shadcn/ui was used as CSS framework libraries and components integrated with React. The interactive map layer built on Leaflet added an avenue to search strains based on geographic or map location and display a visualization of strain count data.

The application's search and filtering was designed utilizing minimal search input and select buttons that can implement search features on any field of isolation source information. The MISO categories adds an easier filtering if its controlled vocabulary on isolation source information was utilized.

	ODD ITEMS	EVEN ITEMS	SCORE	GRADE
TESTER 1	22	5	92.5	A
TESTER 2	17	10	67.5	C
TESTER 3	21	7	85	A
TESTER 4	25	9	90	A
TESTER 5	24	10	85	A
TESTER 6	20	8	80	B
TESTER 7	19	9	75	B
TESTER 8	21	8	82.5	A
MEAN SCORE:			82.1875	A

Fig. 2: Caves SUS Score

B. System Testing

The testing was conducted in a remote setup, and the testers were provided with links to the web application, user guide, and evaluation form. The user guide contains details about the system's key features and functionalities, aloSg with user-level instructions. The evaluation form utilized John Brooke's System Usability Scale (SUS), a widely used method for measuring website usability, and was administered via Google Forms. It consists of a 10-item rating scale with 5 options ranging from Strongly Disagree to Strongly Agree, each with corresponding values of 1 to 5. The web application underwent testing by 8 respondents, comprising of NICER CAVES members and UPLB BS Biology undergraduates and graduates. As shown in Figure 2, the web application obtained a score of 82.1875, qualifying for an A grade, indicating an acceptable score and excellent usability.

In addition to the SUS, the evaluation form also asked users for feedback and comments. Some notable comments include the web application's modern and fresh look that differs from traditional online metadatabases. Some feedback stated to include authors and citations, as well as simple descriptions and discussions on the statistical results of the data.

V. CONCLUSION AND FUTURE WORK

The Caves Isolation source web application was developed using the MERN stack and is currently deployed on Netlify. The application provides a metadatabase on the isolation source information of the microbial collected from bats and caves in CALABARZON. It emphasizes the usage of MISO 3-level categories, an internationally accepted controlled vocabulary, to describe microbial origins and environmental information and the usage of an interactive map to display strain count information from a specific area or sampling site. These features allow users to search and view strains based on isolated source information. The application received an above-average SUS score of 82.1875. Overall, the web application has accomplished its objectives of developing an isolation source focus database.

The testers' feedback may be adopted for future improvements to the web application and its functionalities. To complete the project it is recommended to add the strains' taxonomic information, citations and references, pictures or graphics, and other relevant information.

APPENDIX I
SCREENSHOTS OF THE WEB APPLICATION

A. Admin and Guest Views

Fig. 3: Home page

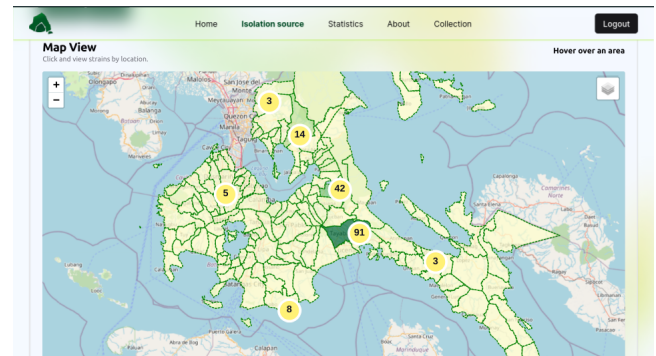Fig. 6: Isolation source table

Fig. 4: Signup page

Fig. 7: Isolation source map

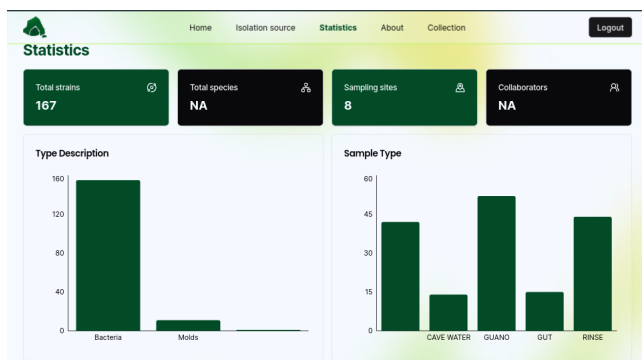Fig. 5: Log in page
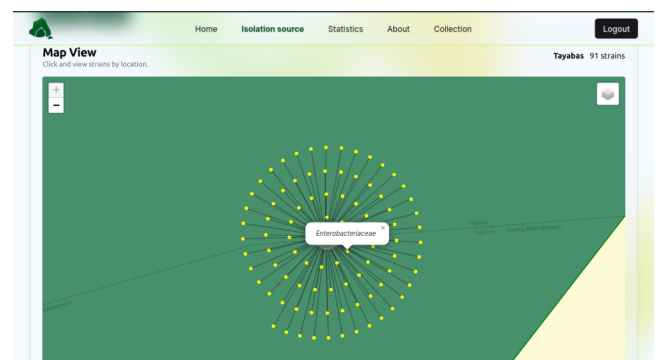
Fig. 8: Isolation source map strains

Fig. 9:Statistics page

Fig. 12: Update strain form

B. Admin Views

Fig. 10: Admin collection page

Fig. 12: Delete strain prompt

Fig. 11: Add strain form

REFERENCES

[1] "Philippine Caves: Beneath the Earth's surface: Conservation and management," elibrary, https://elibrary.bmb.gov.ph/elibrary/books/philippine-caves-beneath-the-earths-surface-conservation-and-management-3/

[2] M. Kasso and M. Balakrishnan, "Ecological and economic importance of bats (order Chiroptera)," ISRN Biodiviersity, vol. 2013, pp. 1–9, 2013. doi:10.1155/2013/187415

[3] S. Sakoui et al., "The life hidden inside caves: Ecological and economic importance of Bat Guano," International Journal of Ecology, vol. 2020, pp. 1–7, 2020. doi:10.1155/2020/9872532

[4] M. P. De Leon, A. D. Montecillo, D. S. Pinili, M. A. Siringan, and D.-S. Park, "Bacterial diversity of bat guano from Cabalyorisa Cave, Mabini, Pangasinan, Philippines: A first report on the metagenome of Philippine bat guano," PLOS ONE, vol. 13, no. 7, 2018. doi:10.1371/journal.pone.0200095

[5] S. Banskar, D. T. Mourya, and Y. S. Shouche, "Bacterial diversity indicates dietary overlap among bats of different feeding habits," Microbiological Research, vol. 182, pp. 99–108, 2016. doi:10.1016/j.micres.2015.10.006

[6] L. C. Reimer et al., "Bacdivein 2019: Bacterial phenotypic data for high-throughput biodiversity analysis," Nucleic Acids Research, vol. 47, no. D1, 2018. doi:10.1093/nar/gky879

[7] L. C. Reimer et al., "Bacdive in 2022: The knowledge base for standardized bacterial and Archaeal Data," Nucleic Acids Research, vol. 50, no. D1, 2021. doi:10.1093/nar/gkab961

[8] Reimer,L.C., Söhngen,C., Vetcininova,A. and Overmann,J. (2017) Mobilization and integration of bacterial phenotypic data—Enabling next generation biodiversity analysis through the BacDive metadatabase. J. Biotechnol., 261, 187–193

[9] S. Dérozier et al., Omnicrobe, an open-access database of microbial habitats and phenotypes using a comprehensive text mining and data fusion approach, 2022. doi:10.1101/2022.07.21.500958

[10] J. Imhoff, "New Dimensions in Microbial Ecology—Functional Genes in Studies to Unravel the Biodiversity and Role of Functional Microbial Groups in the Environment," Microorganisms, vol. 4, no. 2, p. 19, May 2016, doi: 10.3390/microorganisms4020019

[11] S. Mukherjee et al., "Genomes online database (GOLD) v.8: Overview and updates," Nucleic Acids Research, vol. 49, no. D1, 2020. doi:10.1093/nar/gkaa983

[12] National Center for Biotechnology Information, https://www.ncbi.nlm.nih.gov/ (accessed May. 5, 2023)

[13] P. Yilmaz et al., "Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications," Nature Biotechnology, vol. 29, no. 5, pp. 415–420, 2011. doi:10.1038/nbt.1823

[14] V. Agafonkin, "An open-source JavaScript library for interactive maps," Leaflet, https://leafletjs.com/ (accessed Apr. 6, 2023)