

BATSeq: A Web-Based System for CAVES's Nucleotide Sequences with Integrated Localized Basic Local Alignment Search Tool

Ramnick Francis P. Ramos and Arian J. Jacildo

Abstract—In 2021, the University of the Philippines Los Baños Museum of Natural History (UPLB-MNH), through the DOST-NICER Center for Cave Ecosystem Research (CAVES) Program, generated a comprehensive dataset containing microbial metadata and corresponding 16S rRNA sequences from cave ecosystems in CALABARZON. Despite the volume and scientific value of these outputs, no existing platform consolidated, visualized, or communicated these nucleotide findings for researchers or the general public.

This study developed *BATSeq*, a web-based database management system that centralizes the CAVES Program's nucleotide datasets and integrates core bioinformatics features, including FASTA visualization, phylogenetic tree generation, and a localized implementation of the Basic Local Alignment Search Tool (BLAST) for sequence similarity analysis. The system was evaluated through the System Usability Scale (SUS) and a task-based usability test was administered to 11 respondents with backgrounds in microbial research.

BATSeq obtained a SUS score of 75.45, placing it within the “Good” usability category. The task-based evaluation produced an overall mean score of 4.27 out of 5, indicating strong approval of the system’s dashboard, CRUD functionalities, BLAST module, sequence visualizer, and phylogenetic tree. These results suggest that BATSeq is both usable and effective in supporting data curation workflows for nucleotide sequences within the CAVES Program. Future improvements may focus on enhancing navigation for novice users, expanding visualization features, and extending integration to related MNH bioinformatics systems.

bat guano, bioinformatics, BLAST, cave ecosystems, database systems, metagenomics, RNA sequences, sequence alignment

I. INTRODUCTION

A. Background of the Study

In 2021, the Museum of Natural History of the University of the Philippines—Los Baños (UPLB-MNH), together with the Department of Science and Technology—Niche Centers in the Regions for R&D (DOST-NICER) Program, launched the Center for Cave Ecosystem Research (CAVES) [1]. CAVES, which is a three-year venture program, seeks to implement research-based protection of cave biodiversity within the region of CALABARZON through conservation management.

The project produced numerous observational field research that contained pieces of information regarding the caves within CALABARZON. The researchers for this program, with the

Presented to the Faculty of the Institute of Computer Science, University of the Philippines Los Baños in partial fulfillment of the requirements for CMSC 170: Introduction to Artificial Intelligence © 2024 ICS University of the Philippines Los Baños

aim of biologically profiling these caves on aspects specific to the microbial culture within the bat guano isolates in there, have created datasets that provide useful insights regarding the biodiversity of the caves within the said region – aiding in the ecosystem surveillance and protection by the said research organizations.

In depthly, this program was established under the mission of conserving cave ecosystems through a thorough understanding of the ecological systems that exists therein through analyzing the various specimens from cave-dwelling species – particularly bats. Of these research-based ventures, one key output was the paper “Antibiotic-resistant Gram-negative Bacteria from the Fecal Pellets of Bats Collected from the Cavinti Underground River and Cave Complex, Cavinti, Laguna, Philippines” [2]. This observational biology paper provided insights, particularly on the bacterial characteristics and identity, through the 16S ribosomal ribonucleic acid (16S rRNA) of isolates. Nucleotide sequences, particularly rRNA sequences, provide a vital role in the exposition of the “changing environmental conditions” of an ecosystem where a bacteria had developed [3].

This sequence, through identification and phylogenetic analysis, will then provide an explanation on the evolutionary development of bacteria within it– which for this specific research project, the caves ecosystem in the CALABARZON Region.

To wit, the provision of the bacterial identity of bat guano samples or isolates through these rRNA sequences of the CAVES Program is crucial, for it provides an additional layer of understanding of the ecological role of cave organisms, particularly bats, in these caves. Adjunct to this bacterial identity identification, there exists current efficiency issues regarding leading bioinformatics tools that provide the bacterial identity of a given 16 rRNA sequence; this then implicatively hinders the research within the NICER CAVES Program. Specifically, the National Center for Biotechnology Information (NCBI), particularly with their commonly used BLAST+ web service for identifying bacteria, had been known to have recurring unavailability tendencies (Error 505) due to web traffic [4].

Having said, this developmental study aims to provide an efficient means on understanding the dataset of UPLB-MNH on the NICER-CAVES program – particularly on the data containing RNA Sequences – through creating a web-based database management system. This database management system, to be called “BATSeq”, will also incorporate localized/in-client bioinformatics tool such as basic local alignment search

tool to ease and optimize the identification and searching of specific information regarding nucleotides.

B. Statement of the Problem

Given the presence of outputted dataset that contains nucleotides deoxy- and ribonucleic acid sequence information of cave ecosystems, there presents a need to develop an interface that allows the researchers of the Center for Cave Ecosystems Research (CAVES) of the University of the Philippines Los Baños Museum of Natural History (UPLB-MNH) to communicate their research findings regarding the nucleotide of microbes they have gathered. The lack of means for researchers to share their outputs leaves insightful data about the phylogeny of microbial culture from these caves off the knowledge of the general public. Such lack of a centralized sequence database entails a gap in the science communication of this research for the primary mission of Department of Science and Technology Niche Centers in the Regions for R&D (DOST NICER) Program, which is to promote biodiversity protection for cave ecosystems through knowledge dissemination. Furthermore, the present leading bioinformatics tool used by researchers to provide the characteristics and identity of nucleotides gathered, particularly the Basic Local Alignment Search Tool (BLAST), also presents an opportunity to be incorporated and tailor-fitted to the needs of the CAVES program. In relation to this, specifically with the researchers of the UPLB-MNH under the CAVES Program, there is a pressing need for bioinformatics tools that are engineered to address the needs for caves ecosystem analysis in the region of CALABARZON.

C. Significance of the Study

The creation of a Nucleotide Database for the NICER CAVES program addresses the issue of communicating research findings to the greater public that the outputs within this program face. Having a dedicated database system will enable other researchers from different institutions, may it be external, to view and provide their own analysis of the nucleotide data of the NICER CAVES Program. Furthermore, having such a dedicated database also provides a convenient and efficient workspace environment for present researchers within the UPLB-MNH who are working or have worked under the NICER CAVES program. This will allow them, via a web-based interface, to create, update, edit, and analyze the RNA data that they have sequenced.

Furthermore, through localizing the basic local alignment search tool in a dedicated server space, the integrated BLAST within the system will then be able to optimally manage web traffic as it will niche down the users coming, tandem with specifying down the nucleotide mapping database to a sub-database specific for the NICER CAVES dataset.

D. Objectives of the Study

The objective of the study is to provide a database management system that enables researchers to create a web-portal exhibit of the sequence findings of the Center for

Cave Ecosystems Research (CAVES) of the University of the Philippines Los Baños Museum of Natural History (UPLB-MNH) for the general public. This website's administrators will be able to perform the fundamental database operations of creation, updating, removing, editing, and viewing insight reports regarding the biological data information of cave ecosystems and its corresponding metadata.

Specifically, the study aims to address the following objectives:

- 1) To develop a web-based deoxy- and ribonucleic acid sequence database system tailored to addressing the functional needs of researchers of the Center for Cave Ecosystems Research (CAVES) of the University of the Philippines Los Baños Museum of Natural History (UPLB MNH);
- 2) To streamline the data curation process of researchers under the CAVES Program by developing an interface that will enable them to view a report dashboard, perform basic creation, reading, updating, and deletion operations on deoxy- and ribonucleic acid sequence, to view the phylogenetic tree of the current dataset, and to perform Basic Local Alignment Search Tool (BLAST) searching within the database; and
- 3) To assess the effective performance of said web-based database management system vis-a-vis its unit-specific functions.

E. Scope and Limitations

The study developed a web-based database management system for the nucleotide sequences of the Center for Cave Ecosystems Research (CAVES) of the University of the Philippines Los Baños Museum of Natural History (UPLB MNH).

The web-based database management system incorporated the fundamental functionality of creation, updating, deletion, and editing of the dataset presented by the CAVES Program. Specifically, this database management system will be limited to incorporating details regarding the following: 16S rRNA (FASTA File); Isolate Code; Type of Sample; Bat Source; Sampling Site; Gram reaction; Cell shape; Oxygen requirement; Presence of cytochrome c oxidase; Endospore-forming capability; Antibiotic resistance profile; Identity; and Pathogenicity.

Other microbial characteristics that are cave-specific may not be present within the database system if not within the present dataset of the aforementioned research center.

Additionally, this database system contained a sequence-specific viewer for ribonucleic acid sequence. This will also contain an in-dashboard view on the genetic similarities of the dataset through a phylogenetic tree. Lastly on the functions, the web-based database management system had its localized implementation of Basic Local Alignment Search Tool.

The developed web-based database management system was user-tested by a sample of 11 student-researchers from the Institute of Biological Sciences of the University of the Philippines Los Banos (UPLB) and Institute of Computer Science. Purposively, these respondents have prior knowledge on microbial research.

The respondents for the user testing of the developed website was gathered through purposive sampling in a typical case sampling method. Said sampling method was done to ensure that the initial users of the website will contain student researchers who are studying or are currently working for microbial microbiological research, the domain of CAVES' research.

II. REVIEW OF RELATED LITERATURE

A. Bat Guano 16S rRNA Sequences

The use of 16S rRNA sequences from bat guano or bat pellets to further understand an ecosystem is a cornerstone application of metagenomics – where the understanding of the environment is centered on the acquisition of environmental microbiomes through recovered materials directly from the ecosystem [5]. This approach includes the "holistic characterization" of microbiomes such as soil, lakes, and, specific to the NICER CAVES Project, animal gut – particularly bat guano.

In the utilization of this, the characterization of the environment allows the elaboration of the taxonomic changes and characteristics within a microbial community, giving helpful insights regarding the proactive conservation of an ecosystem. Through the exposition of biodiversity, 16S rRNA sequences that are used to derive the identity of a bacteria – as it holds details for bacteria's ribosomal information regarding its protein synthesis – are used to profile the population of microbes in an environment.

Immediate and relevant studies regarding the use of such metagenomics are the totality of RNA research of the NICER CAVES Program [2]. Here, they utilized 16S rRNA sequences to give identity and specify characteristics in bat guano specimens. A shortened example of this in an isolate from studies within the NICER CAVES Program is:

```
>H231124-R03_C13_B1I1_27F.fasta 1207
CGCAGTGGCGGCAGCTACACATGCAGTCGAA
...CCGGGA
```

Typically composed of 1500 base pairs, this sequence in FASTA format holds codons that researchers could leverage to get the identification of bacteria in order to supplement the analysis of the microbial population profile of a specific ecosystem.

B. NICER CAVES Program Environmental Profiling Studies on RNA Sequences

The NICER CAVES Program, as mentioned, aims to provide a deeper understanding of the biodiversity profile of caves within CALABARZON. Under its primary directive, this program believes "caves will have a role in the future and understanding these environments will be essential in humans' survival in extremely adverse environments" [1] [6].

In relation to this, a considerable amount of studies under the NICER CAVES Program conduct metagenomic environmental profiling that assesses the environmental policies protecting cave ecosystems [2]. For instance, "Antibiotic-resistant Gram-negative Bacteria from the Fecal Pellets of

Bats Collected from the Cavinti Underground River and Cave Complex, Cavinti, Laguna, Philippines" analyzed the importance of giving policies or regulations regarding using bat fecal materials from caves in Cavinti, Laguna. This, through both morphological and 16S rRNA sequence identification, allowed the analysis of the environmental hazard of using bat fecal pellets as fertilizers as they may cause bat-borne bacterial zoonotic diseases.

Isolates, or specific data within the datasets, of these researches include the following attributes:

- Isolate Code
- Type of Sample
- Bat Source
- Sampling Site
- Gram reaction
- Cell shape
- Oxygen requirement
- Presence of cytochrome c oxidase
- Endospore-forming capability
- Antibiotic resistance profile
- Identity
- Pathogenicity
- 16S rRNA (FASTA File)

C. Basic Local Alignment Search Tool (BLAST)

Basic Local Alignment Search Tool, more commonly called BLAST, utilizes string comparisons on the sequence of nucleotides by finding regions of similarity from a query sequence, the sequence being identified, to a mapping database of sequences, the numerous sequences that the query sequence will be compared to [7]. This comparison exposes similarities in characteristics amongst sequences that enable researchers to identify the supposedly unknown identity of a sequence at hand.

Incorporated with statistical significance analysis for comparison, this alignment tool allows researchers to make informed inferences about the function, evolutionary relationship, and identity of sequences to help identify gene families. These facets of the characteristics of an isolate provide an understanding of the environmental status in metagenomics.

There are several types of BLAST searchers, accordingly categorized to their query sequence type and mapping database type. For comparing protein sequences to a database of nucleotides, BLASTP is used; for comparing nucleotide sequences such as DNA and RNA to other nucleotide sequences in a database, BLASTN is commonly used. However, if a more in-depth comparison is needed, query sequences can be translated (the process of transcription of nucleotide to proteins) and then string compared to their corresponding databases; examples of this include BLASTX, TBLASTN, and TBLASTX [8].

$$S(i, j) = \max \begin{cases} S(i - 1, j - 1) + mat & \text{if } i > 0, j > 0 \\ \quad \text{and } a_i = b_j \\ S(i - 1, j - 1) + mis & \text{if } i > 0, j > 0 \\ \quad \text{and } a_i \neq b_j \\ S(i, j - 1) + ind & \text{if } j > 0 \\ S(i - 1, j) + ind & \text{if } i > 0. \end{cases} \quad (\text{Eq 1})$$

Eq 1 provides an overview of the greedy algorithm that is used by BLAST when aligning nucleotide sequences [9]. The primary characteristic of this algorithm revolves around the aspects of implementing local maxima, where the region of highest similarity serves as the starting point for extending alignment. This matching algorithm makes the alignment of nucleotides more optimal.

D. Previous Web Systems Catering to NICER CAVES Dataset

In recent years, there have been a couple of projects that were implemented under research clusters within the Computational Interdisciplinary Laboratory of the University of the Philippines Los Banos (UPLB-CINTERLABS) that catered to managing the datasets outputted by the NICER CAVES Program. One example of this is the developmental study “Implementation of Systematics Access to Isolation Sources of Microbes in a Culture Collection System” in which the geological locations of each sample were visualized [10]. As a primary feature, said highlighted location of the caves within CALABARZON where the specific isolates were collected were presented using interactive maps. However, no attempts have been made to visualize and present data regarding nucleotides as of yet.

E. Computation Tools and Technologies

Visualization of data is a vital aspect of understanding RNA sequences. Several pre-existing JavaScript libraries assist in the presentation of complex sequences for easier and more user-friendly interpretation. For visualizing sequences, the npm library `seqviz` can be used [11]. This library allows users to view DNA, RNA, and protein sequences either in a linear or circular format, incorporating streamlined color-coding of codons.

Furthermore, for incorporating BLAST into JavaScript applications, `blastjs`, a BLAST+ wrapper for Node applications, can be implemented [12]. This library conveniently incorporates BLAST+ suite functionalities—such as the greedy algorithm application of local maxima for aligning nucleotides—into Node.js applications with ease of use.

Lastly, to represent a larger-scaled database of already-identified RNA sequences in a visualized tree, `phylotree` can be used [13]. This library creates a phylogenetic tree of identified organisms’ identities, obtained via BLAST, to highlight their evolutionary relationships.

F. Evaluation Tools

Evaluation tools and other testing procedures and techniques are requisite for a systematic means of identifying the success of developmental project implementation. Of this testing, three primary tests are relevant to BATSeq’s implementation: Usability Testing, User Testing, and Unit Testing.

User Testing provides insights into a more personalized and realistic usage scenario of a web service, as it requires the web application to be used by real users [14]. Guided by a set of instructions, user testing aims to evaluate the specific objectives of a project.

On the other hand, for more standardized testing of the usability of a web application, the System Usability Scale (SUS) is commonly used. The System Usability Scale is used to quantify the usability of a service through a 10-item statement-based questionnaire, composed of Likert-scaled questions [15].

Lastly, Unit Testing is used to determine the effectiveness of a system (i.e., if each function works as expected) by testing each endpoint of a database [16]. For JavaScript applications, Jest is used to ensure that the codebase is bug-free via a test automation framework.

III. MATERIALS AND METHODS

This section is subdivided into the following discussion sections: Development Tools, System Architecture, Types of Users, Database Design and Models, Functional Requirements, Evaluation, as well as some Initial Development.

A. Development Tools

Throughout the development of this study, a machine with the following hardware specifications was used: a processor of 11th Gen Intel(R) Core(TM) i5-11400H @ 2.70GHz (2.69 GHz), 24.0 GB of RAM (23.7 GB usable), and a system type of 64-bit operating system, x64-based processor. Furthermore, the system operates on Windows 11.

Moreover, the following software development tools and technologies will be used throughout the development process:

- **Visual Studio Code** – An Integrated Development Environment (IDE) for creating software applications.
- **Git** – A Version Control Management System for development.
- **MySQL** – A Relational Database Management System for Structured Query Languages.
- **Express.js** – A Web Framework used to build APIs for Node.js applications.
- **React.js** – A JavaScript library for creating native user interfaces.
- **Material UI** – An open-source ReactJS-based frontend framework that provides readily available components for web applications.
- **Node.js** – An open-source JavaScript runtime environment.
- **Phylotree.js** – A taxonomic tree visualizer for JavaScript applications.
- **blastjs** – A wrapper library for Node.js applications.
- **seqviz** – An npm library used to visualize nucleotide sequences.

B. System Design and Development

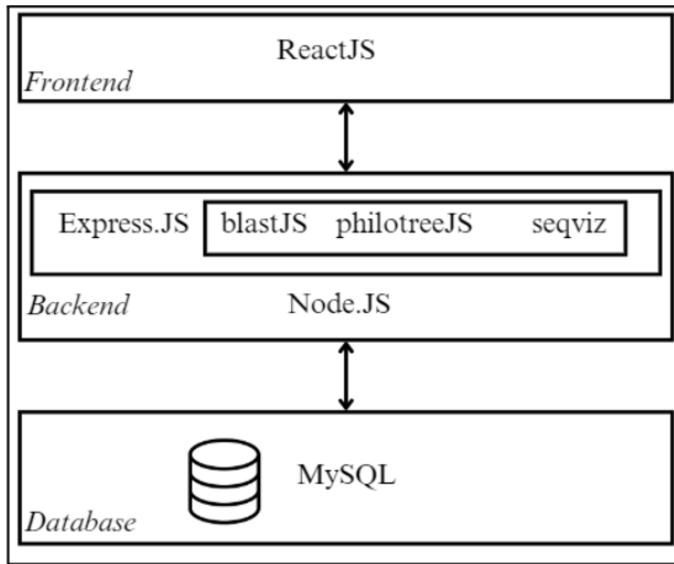


Fig. 1: System Architecture Diagram of the MERN Stack Environment with the bioinformatics/visualization libraries

The system architecture of the BATSeq is patterned after the standard interaction of a three-sectioned model for MERN (MySQL instead of MongoDB) Applications to ensure seamless data processing. These three main sections are comprised

of frontend, backend, and database. Furthermore, specific for BATSeq, a submodule under the backend section where the bioinformatics-specific functions are performed is also separated.

For the frontend portion of the system, ReactJS together with the components under Material UI was used to handle the Graphical User Interface representation of the Database Management System. This includes aspects of uploading FASTA Files, having input fields for creating new isolates, and the like. Furthermore, this end of the system will also be receiving some visualizations to be thrown by some node packages under the backend tier of the system.

The backend ran mainly using Node JS and framed using ExpressJS, on the other hand, will be responsible for the tasks under aligning sequence and running node packages for visualization and representation of nucleotides – aided by npm packages blastjs, seqviz, philotree.js. The last three mentioned packages are under the module of bioinformatics tools as they are specific for visualizing the RNA Sequences at hand.

Lastly, a MySQL database stored the RNA sequences, metadata for each isolate, details regarding the users, and image link files. The database schema will include the CRUD operations for the isolates, including the updating of an RNA sequence via a FASTA File uploader.

In general, the flow of data will always be frontend-backend-database for setting the present database or database-backend-fronted for retrieval of information from the database. These sections will be connected via API endpoints.

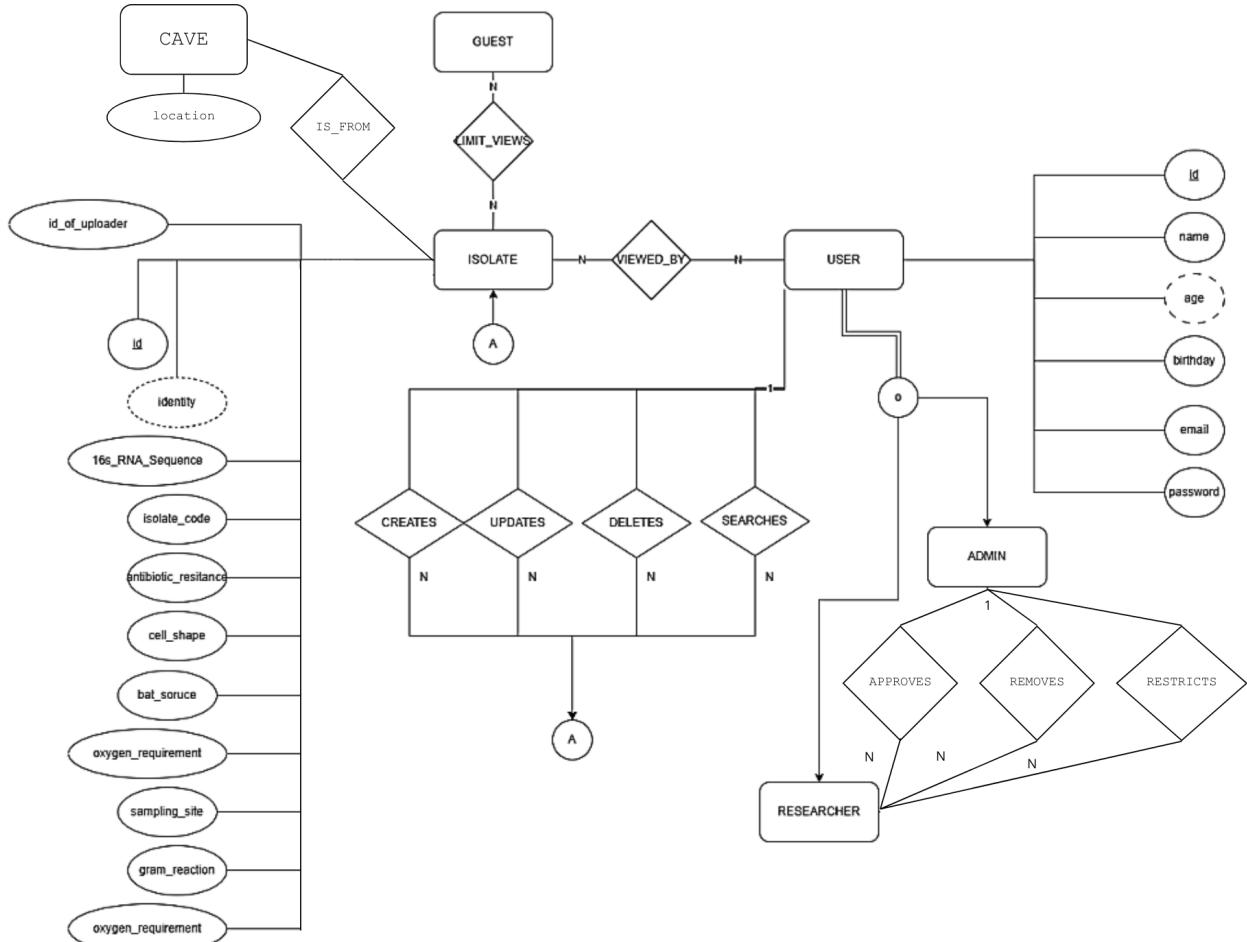


Fig. 2: Entity-Relationship Diagram of BATSeq

C. User-Centric Features and Operations

The following are the categories of users that will be interacting with the database management system:

1) **Administrator:** The administrator will be in charge of inputting new user-researchers into the database. The administrator is capable of creating, reading, updating, and deleting isolates within the database. The administrator has the privilege to manage the current pool of researchers in the database (delete, update, and remove). There will only be one administrator for the entire database (assigned to the data curator of the Bioinformatics Lab of the Institute of Biological Sciences).

2) **Researchers:** The researcher-users are capable of creating, reading, updating, and deleting isolates that they created within the database. They can perform 16S rRNA identification and searching via the BLAST functionality of the system.

3) **General Users:** General users can perform 16S rRNA identification and searching via the BLAST functionality of the system. They can read all the information of the isolates within the database system.

With these aforementioned user type outlines, their corresponding primary functional requirements are as follows:

4) Administrator:

- **Log in:** Administrators will be required to log in to the application via their respective account.
- **View Researcher-Users Details:** The Administrator is capable of viewing the tables and the entirety of its values on the researcher-users.
- **Manage Researcher-Users:** The administrator is allowed to create, update, and delete researchers within the database system.
- **Manage Isolates:** The administrator is allowed to create, update, and delete all Isolate data within the database system.
- **View Dashboard:** The Administrator is capable of viewing the dashboard containing the phylogenetic tree of all the Isolates, including a breakdown snapshot of the database.
- **View Isolates:** Through sequence visualizers, the administrator can view the details of the Isolates together with their 16S rRNA sequence.

5) Researchers:

- **CRUD their own Isolates:** Once within the system, the researcher-users are allowed to add, remove, update, and delete their own respective isolates.
- **View Dashboard:** Just like the Administrator, the researchers are capable of viewing the dashboard containing the phylogenetic tree of all the Isolates, including a breakdown snapshot of the database.
- **View Isolates:** Similarly with the Administrator, through sequence visualizers, the researchers can view the details of the Isolates together with their 16S rRNA sequence.

6) General Users:

- **Read Function:** The general users (unlogged-in users) may continue to visit the website and view the dashboard containing a snapshot of the database.
- **View Isolates:** Similarly to the Administrator and Researchers, through sequence visualizers, the guest users can view the details of the Isolates together with their 16S rRNA sequence.

On the other hand, database design of the BATSeq utilized a Structure Query Language (SQL) approach by using MySQL as the management database system to ensure standardized scalability of the database, populated with the dataset from the NICER CAVES Program. Illustrated in Figure 2 is the Entity Relationship Diagram (ERD) for the system.

The system will mainly consist of three entities: Admin Users, Researcher Users, and the Isolates themselves. A supplementary entity “Guest” is also added.

The User entity represents all interacting individuals in the system including the researchers, admin, and regular user (categorized as guests if not an admin or a researcher). Moreover if categorized, the user entity will have different privileges in mutating the database (i.e., an admin can manage the researchers within the system). The admin “managing” the researchers encompasses both the addition, deletion, and updating of particular researchers in the database. The “sign-up” privilege being solely for the administrator was decided to centralize the control, in order to better track the input and deletion of the editors (i.e. researchers) within the system.

The *Isolate Entity*, on the other hand, represents the microbial data within the dataset from the NICER CAVES Program. It will contain attributes corresponding to the preexisting metadata fields for each isolate in the dataset. These attributes include: identity, 16S rRNA sequence, isolate code, antibiotic resistance, cell shape, bat source, oxygen requirement, sampling site, gram reaction, and endospore-forming capabilities.

Administrator-User oversees all isolate data within the system; they can update and edit all the isolate data. However, for Researcher-Users, only those Isolates they uploaded will be available for update, edit, or delete for them.

Furthermore, the identity of the Isolate is a derived attribute since, via the implementation of BLAST Identification, said identity can be discerned through Nucleotide Alignment.

For easier representation, See Figure 2 for the Use Case Diagram.

D. Evaluation

In order to determine the quantifiable success of the implementation, this paper conducted a pronged evaluation of the BATSeq. Namely, the tests: System Usability Scale (SUS) and User Testing.

The System Usability Scale (SUS) was used to determine the general usability of the web application, as a quantifiable measure for the objective of developing a web-based RNA

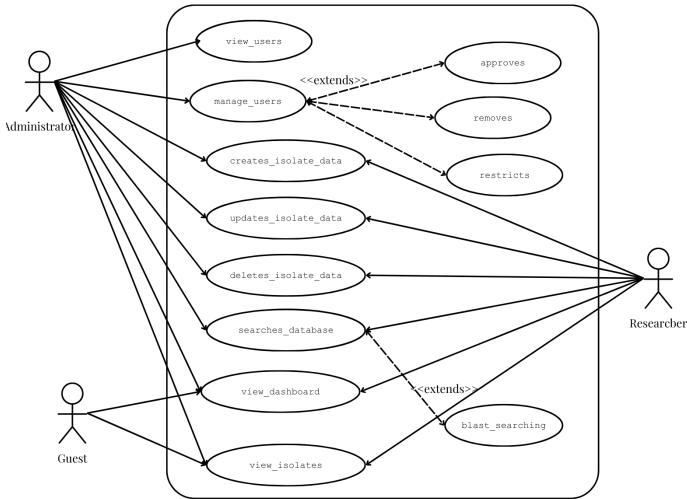


Fig. 3: Use Case Diagram of BatSeq

database system tailored to the details (attributes) specified under the NICER CAVES Program. It consisted of 10 statements with a five-point Likert Scale from “Strongly Disagree” to “Strongly Agree” to be answered by 10 present or past involved biologists with the said research program. The following statements was given (See Table 1).

TABLE I: Statements for the System Usability Scale

No.	Questions
1	I think that I would like to use this system frequently.
2	I found the system unnecessarily complex.
3	I thought the system was easy to use.
4	I think that I would need the support of a technical person to be able to use this system.
5	I found the various functions in this system were well-integrated.
6	I thought there was too much inconsistency in this system.
7	I would imagine that most people would learn to use this system very quickly.
8	I found the system very cumbersome to use.
9	I felt very confident using the system.
10	I needed to learn a lot of things before I could get going with this system.

Furthermore, to adjunct and complement the findings of SUS, User testing will be used to give specific insights into the usefulness of the system. This is in response to determining the success of the objective of enhancing the data curation process of researchers under the NICER CAVES Program by providing an interface that will enable them to view a report-bearing dashboard, perform basic CRUD operations on RNA Sequences, and perform BLAST searching within the database.

To be analyzed using a 5-point Likert scale, the following statements will be identified by the same 10 respondents:

General Usability and Interface:

- Ease of Use:** The dashboard and the interface can be easily navigated.
- Effectiveness of the Dashboard:** The report of the dashboard contains the necessary information and visualization for providing insights on the data of the NICER

CAVES Program.

- Intuitiveness:** it was easy to understand how to perform the creation, deletion, and updating of RNA Sequences and their metainformation in the web application.
- User Interface Design:** The visual layout of the website was appealing.

Data Curation Functionalities:

- Basic CRUD operations:** I was able to add, edit, and delete meta information and RNA sequences without any difficulty.
- Search Functionality:** The searching using the nucleotide sequence features made it easy to explore the database.

BLAST Module:

- Integration of BLAST:** The BLAST tool was integrated seamlessly with the website.
- BLAST Hits Result Presentation:** The results of using the BLAST, particularly its user interface are easy to understand.
- Sequence Visualizer:** The sequence visualizer was helpful in understanding the data.
- Phylogenetic Tree:** The phylogenetic tree was helpful in adding insights to the system.

Overall Performance:

- Speed:** the web application responded quickly to my queries and interactions.
- Satisfaction:** Overall I am satisfied with the system's feature on curating meta information and sequence data for the NICER CAVES Program.

In the ending section of this User Testing, an open-ended Recommendation/Comments short-response field will be given to the respondents.

IV. RESULTS AND DISCUSSION

For development of the web application BATSeq, a web-based database management website was successfully developed to perform the basic CRUD functionality on the meta-information of Isolates. (See Figure 4 and Figure 5).

Furthermore, it was also able to incorporate the bioinformatics module features including the BLAST (Figure 6), Sequence Visualizer (Figure 7), and Phylogenetic Tree (Figure 8).

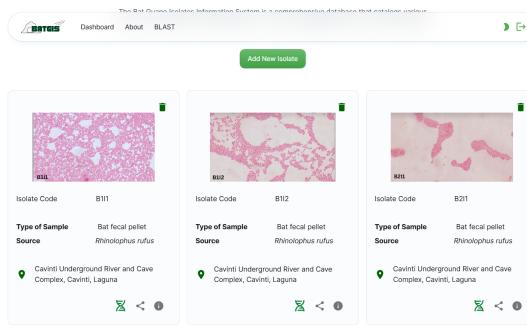


Fig. 4: Grid-view Dashboard of all the Isolates in the Website (Light Mode)

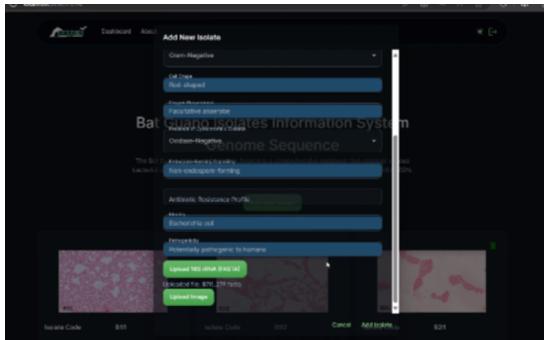


Fig. 5: Input Field for the Creation of an Isolate Data

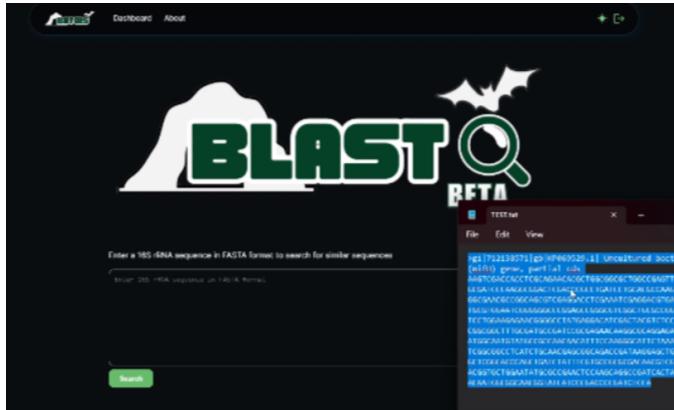


Fig. 6: Inputting 16S RNA sequence for Identification

A. Respondents

Sample. The respondents that was gathered for the evaluation of the website was 11 participants. All respondents voluntarily agreed to the data privacy statement and confirmed their participation. These participants were purposively gathered and was selected to have prior knowledge in sequencing. The testing was conducted on November 20, 2025.

Affiliations. Majority of participants the participants came from the College of Arts and Sciences(CAS) of the University of the Philippines Los Banos, specifically the Institute of Biological Sciences. One participant came from the UPLB Museum of Natural History and the Institute of Computer Sciences. Most of the participants are BS Biology students, while two were studying Zoology under the Mathematics and Science Teaching program and one graduated from BS Computer Science. Expertise. Familiarity of the respondents on RNA Sequences ranged from unfamiliar(scale) to having advanced familiarity () (See Fig 9 and Fig 10). Two of the participants were involved in the NICER Caves Program.

B. System Usability Scale(SUS)

The BATSeq web-based management system scored a mean of 75.45 which places it into “Good” usability at about 80th percentile in the standard of System Usability Scale. The median was 77.5 and the results range was 48.5 to 100 (See Table II). This suggests that the development was successful in outputting a website that users find usable, although it still

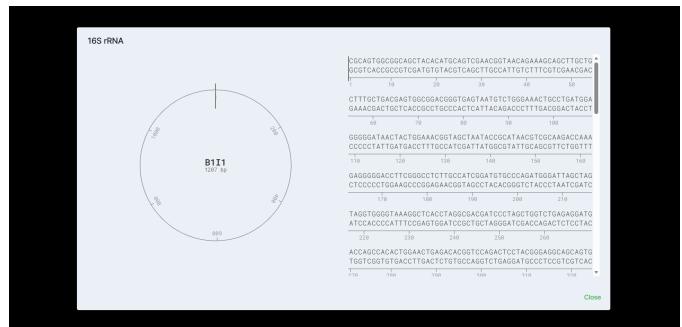


Fig. 7: Sequence Visualizer

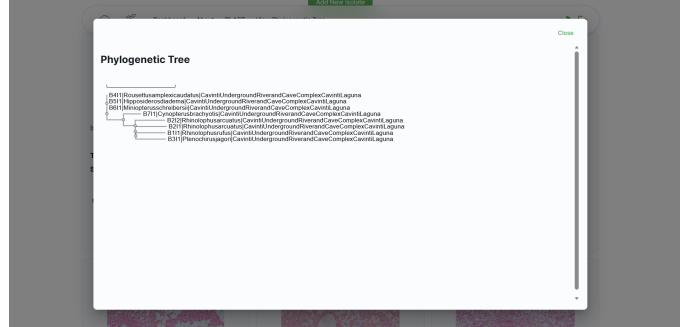


Fig. 8: Phylogenetic Tree

has areas of improvements with the evident lower individual scores.

TABLE II: Summary of the Results of SUS

Mean	74.1
Median	77.5
Minimum	47.5
Maximum	100
Standard Deviation	16.2

C. Usability Testing

For evaluating the specific components of the features of the website, the usability testing was conducted. It tested the website on the specific fields of: General Usability & Interface, Data Curation Functionalities, BLAST Module, and Overall Performance. Shown in Table III is the overview of the results scores obtained.

Overall, the website scored the lowest on the Ease of Use focused question, showing that the system has major areas of improvement on being easily understandable. However, with an overall mean of 4.27 (median: 4.27, mode: 4.18, 4.36), it could be deemed that the performance of the website, specific to what is asked on the features of the application, leans towards the satisfactory level of the respondents.

D. Comments and Suggestions by the Respondents

The respondents of the survey had indicated that the visualization tools for the sequences was one of the most helpful features for them. Majority had also included that the search

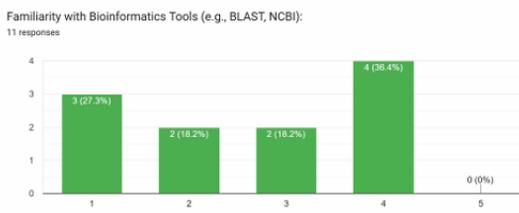


Fig. 9: Familiarity with RNA Sequences

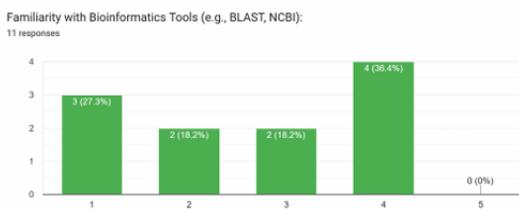


Fig. 10: Familiarity with Bioinformatics Tools

tool using RNA Sequences was one of the major strengths of the web application.

Some, on the other hand, had indicated that the navigation difficulty for beginners, those who are new to bioinformatics, could be a potential challenge. Labels and headings were also confusing.

For improvement, respondents had mentioned that visualization could be better enhanced with aesthetic elements. Experts from the Museum of Natural History had mentioned that a feature that allows the editing of FASTA file within the system would really be helpful for researchers like them. More sample variety on the attributes could also be an area of improvement for the data entries. More notably, the phylogenetic tree could also be incorporated with more information-bearing structure.

V. CONCLUSION AND FUTURE WORK

A. Conclusion

In conclusion, the SUS had demonstrated that BATSeq could be regarded as intuitive and usable. It was effective in hosting RNA sequences and their corresponding metadata as scored by the SUS output of 75.45. The web application system was also deemed effective in curating the dataset of RNA Sequences with the Usability Score's mean of 4.27. It was able to streamline the uploading, deletion, editing, and viewing of the dataset of the NICER CAVES program.

In general, it can be inferred from the results of the evaluation that respondents view the system as something that is manageable to learn. The bioinformatics feature had reinforced the role of the system on being a database for nucleotide sequences.

Individual low scores, on the other hand, was still present in the survey conducted. However, these issues did not reprimand the system from performing its fundamental operations on creation, reading, updating, and deletion, as well as viewing the nucleotides, searching using RNA sequences, and viewing the phylogenetic tree.

TABLE III: Summary of the Results of Usability Testing

Area	Question Focus	Mean
General Usability & Interface	Ease of Use	3.18
	Effectiveness of Dashboard	4.36
	Intuitiveness	4.18
	User Interface Design	4.27
Data Curation Functionalities	CRUD Operations	4.18
	Search Functionality	4.27
Bioinformatics Module	Integration of BLAST	4.36
	BLAST Hits Result Presentation	4.18
	Sequence Visualizer	4.45
	Phylogenetic Tree	4.36
Overall Performance	Speed	4.18
	Satisfaction	4.36

B. Future Work

For future features of the website, it is recommended to incorporate the features of BatSEQ on presently deployed web projects of the NICER Caves program as an imported component to their applications; implementations maybe thorough routing. Moreover, exporting the JavaScript component of the “view sequence button” and including this into the other database systems would elevate the experience of brother applications of this database. Lastly, since the system currently only allows viewing the sequences of the isolates, it could also be of better implementation to add features that allows the researchers to edit them in the website, instead of reuploading a FASTA file.

ACKNOWLEDGMENT

I would like to express my greatest gratitude to my adviser Sir Arian J. Jacildo. His guidance, tandem with unmatched patience, had made this research topic proposal possible. Furthermore, I would also like to acknowledge my co-interns from the Computational Interdisciplinary Laboratory of UPLB (UPLB-CINTERLABS), Midyear 2024 Cohort, who helped me in ideating this project. Moreover, I would also like to express my deepest gratefulness to the Institute of Biological Science of UPLB, especially Mr. Lou Gene Sibal and Sir Andrew Montecillo, for being of great guidance in my traversing the intricacies of biology for this endeavor.

REFERENCES

- [1] F. Cruz, “Regional center for cave research launched by mn, dost and partners,” <https://mnh.uplb.edu.ph/news/regional-center-for-cave-research-launched-by-mnh-dost-and-partners/>, accessed: Nov. 20, 2024.
- [2] L. G. B. Sibal, G. A. T. Barredo, N. G. Sabino, and M. P. D. Leon, “Antibiotic-resistant gram-negative bacteria from the fecal pellets of bats collected from the cavinti underground river and cave complex, cavinti, laguna, philippines,” Dissertation, University of the Philippines Los Banos, 2024.
- [3] A. Chełkowska-Pauszek *et al.*, “The role of rna secondary structure in regulation of gene expression in bacteria,” *International Journal of Molecular Sciences*, 2021, accessed: Nov. 20, 2024.
- [4] (2024) Slow ncbi blast searches (from may 2024). Geneious. [Online]. Available: <https://help.geneious.com/hc/en-us/articles/26244155736852-Slow-NCBI-BLAST-searches-from-May-2024>
- [5] A. Mukherjee, E. D'Ugo, R. Giuseppetti, F. Magurano, and P. D. Cotter, *Metagenomic approaches for understanding microbial communities in contaminated environments: Bioinformatic tools, Case Studies and Future Outlook*. Elsevier, 2023, pp. 103–156.
- [6] (2024) Mn nicer caves program. UPLB Museum of Natural History. [Online]. Available: <https://agora.uplb.edu.ph/mnh-nicer-caves-program/>

- [7] B. Library. (2019) Library guides: Ncbi bioinformatics resources: An introduction: Blast: Compare & identify sequences. [Online]. Available: <https://guides.lib.berkeley.edu/ncbi/blast>
- [8] S. F. Altschul, "Blast algorithm," *eLS*, Sep. 2005.
- [9] Z. Zhang, S. Schwartz, L. Wagner, and W. Miller, "A greedy algorithm for aligning dna sequences," *Journal of Computational Biology*, vol. 7, no. 1–2, pp. 203–214, Feb. 2000.
- [10] E. E. Villanueva and M. R. Anacleto, "Implementation of systematic access to isolation sources of microbes in a culture collection system," Special Problem Thesis, University of the Philippines Los Baños, 2024.
- [11] (2024, October) seqviz. [Online]. Available: <https://www.npmjs.com/package/seqviz?activeTab=readme>
- [12] M. Page, D. MacLean, and C. Schudoma, "blastjs: a blast+ wrapper for node.js," *BMC Research Notes*, vol. 9, no. 1, Feb. 2016.
- [13] (2024, January) phylotree. [Online]. Available: <https://www.npmjs.com/package/phylotree>
- [14] (2019) What is user testing? definition - omniconvert. [Online]. Available: <https://www.omniconvert.com/what-is/user-testing>
- [15] A. Bhat. (2018, May) What is system usability scale? — questionpro. [Online]. Available: <https://www.questionpro.com/blog/system-usability-scale/>
- [16] (2024) Unit testing with jest: Beginners guide with examples. [Online]. Available: <https://www.lambdatest.com/learning-hub/unit-testing-with-jest>

Ramnick Francis P. Ramos is a BS Computer Science student from the University of the Philippines Los Baños. He was a former intern at the Computational Interdisciplinary Laboratory of UPLB (UPLB-CINTERLABS) where his interest in bioinformatics started.



APPENDIX I
USER TESTING CONSENT FORM AND RESPONSES

A. Collected User Testing Data

Field	Response
Respondent 1	
Consent	Yes
Date of Testing	11/20/2025
NICER Affiliation	Yes
Age	48
Sex	Female
Institution	CAS
Degree Program	—
Familiarity with RNA Sequences	4
Familiarity with Bioinformatics Tools	4
SUS Scores (Q1–Q10)	5,1,5,1,5,1,5,1,5,1
Usability Scores (GUI1–GUI4)	5,5,5,5
Helpful Features	Visualization and BLAST features
Difficulties	None
Respondent 2	
Consent	Yes
Date of Testing	11/20/2025
NICER Affiliation	—
Age	21
Sex	Male
Institution	IBS/CAS
Degree Program	BS Biology
Familiarity with RNA Sequences	3
Familiarity with Bioinformatics Tools	3
SUS Scores	3,2,5,2,5,1,5,2,4,2
Usability Scores	5,5,5,5
Helpful Features	Phylogenetic tree
Difficulties	Coordinates/map issues
Respondent 3	
Consent	Yes
Date	11/20/2025
Age	21
Sex	Female
Institution	CAS–IBS
Degree Program	BS Biology
Familiarity with RNA Sequences	2
Familiarity with Bioinformatics Tools	1
SUS Scores	5,1,5,4,5,1,5,1,5,4
Usability Scores	5,5,5,5
Helpful Features	—
Difficulties	—
Respondent 4	
Consent	Yes
Date	11/20/2025
Age	20

Field	Response
Sex	Female
Institution	CAS
Degree Program	BS Biology
Familiarity with RNA Sequences	3
Familiarity with Bioinformatics Tools	2
SUS Scores	3,1,5,2,4,3,4,3,4,5
Usability Scores	5,4,4,4
Helpful Features	Simple and informative interface
Difficulties	"Graduate soon kuya!"

Respondent 5

Consent	Yes
Date	11/20/2025
Age	19
Sex	Male
Institution	CAS
Degree Program	BS Biology
Familiarity with RNA Sequences	2
Familiarity with Bioinformatics Tools	1
SUS Scores	2,3,4,3,4,2,2,3,3,5
Usability Scores	4,4,3,4
Helpful Features	Straightforward features
Difficulties	—

Respondent 6

Consent	Yes
Date	11/20/2025
NICER Affiliation	Yes
Age	25
Sex	Male
Institution	UPLB Museum of Natural History
Degree Program	BS Biology (Microbiology)
Familiarity with RNA Sequences	4
Familiarity with Bioinformatics Tools	4
SUS Scores	4,2,4,2,4,2,4,2
Usability Scores	5,4,4,4
Helpful Features	16S rRNA search tool
Difficulties	FASTA sequences must be cleaned

Respondent 7

Consent	Yes
Date	11/20/2025
Age	20
Sex	Female
Institution	IBS/CAS
Degree Program	BS MST
Familiarity with RNA Sequences	3
Familiarity with Bioinformatics Tools	2
SUS Scores	5,4,3,3,4,2,3,3,4,5
Usability Scores	4,4,4,3
Helpful Features	More sample variety

Field	Response
Difficulties	—
Respondent 8	
Consent	Yes
Date	11/20/2025
Age	20
Sex	Male
Institution	CAS-IBS
Degree Program	BS MST
Familiarity with RNA Sequences	3
Familiarity with Bioinformatics Tools	1
SUS Scores	2,2,4,3,3,3,2,4,3
Usability Scores	5,5,5,4
Helpful Features	Search tab; CRUD
Difficulties	Hard for beginners; dashboard confusing
Comments	Needs better visual structure
Respondent 9	
Consent	Yes
Date	11/20/2025
Age	19
Sex	Female
Institution	IBS
Degree Program	BS Biology
Familiarity with RNA Sequences	4
Familiarity with Bioinformatics Tools	4
SUS Scores	4,2,5,3,5,2,4,2,4,2
Usability Scores	5,5,4,4
Helpful Features	Nucleotide sequencing features
Difficulties	Better naming/titles; more phylo visuals
Respondent 10	
Consent	Yes
Date	11/20/2025
Age	24
Sex	Male
Institution	ICS
Degree Program	BSCS
Familiarity with RNA Sequences	3
Familiarity with Bioinformatics Tools	3
SUS Scores	4,1,5,2,5,1,5,1,4,3
Usability Scores	5,4,5,5
Helpful Features	Running BLAST
Difficulties	—
Comments	Add genome visualization
Respondent 11	
Consent	Yes
Date	11/20/2025
Age	20
Sex	Male
Institution	IBS
Degree Program	BS Biology
Familiarity with RNA Sequences	4

Field	Response
Familiarity with Bioinformatics Tools	4
SUS Scores	5,2,5,2,5,1,4,2,5,2
Usability Scores	5,5,5,4
Helpful Features	Easy and intuitive
Difficulties	None
Comments	Add more data; improve phylo options