

BatSEQ: A WEB-BASED SYSTEM FOR CAVES'S RNA SEQUENCES WITH INTEGRATED LOCALIZED BASIC LOCAL ALIGNMENT SEARCH TOOL

Authors: Ramnick Francis Ramos and Arian Jacildo

Affiliations: University of the Philippines Los Banos - Institute of Computer Science, Computational Interdisciplinary Laboratory, Museum of Natural History



INTRODUCTION

In 2021, the Museum of Natural History of the University of the Philippines—Los Banos (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. 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 this observational biology paper that provided insights, particularly on the bacterial characteristics and identity, through the 16S ribosomal ribonucleic acid (16S rRNA) of isolates[2]. These sequences, 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. This developmental study then 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.

Objectives

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 CAVES for the general public. This website's administrators will be able to perform the fundamental CRUD database operations regarding the biological data information of cave ecosystems and its corresponding metadata. Specifically, the study aims to address the following objectives:

- 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);
- 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
- To assess the effective performance of said web-based database management system vis-a-vis its unit-specific functions.

METHODOLOGY

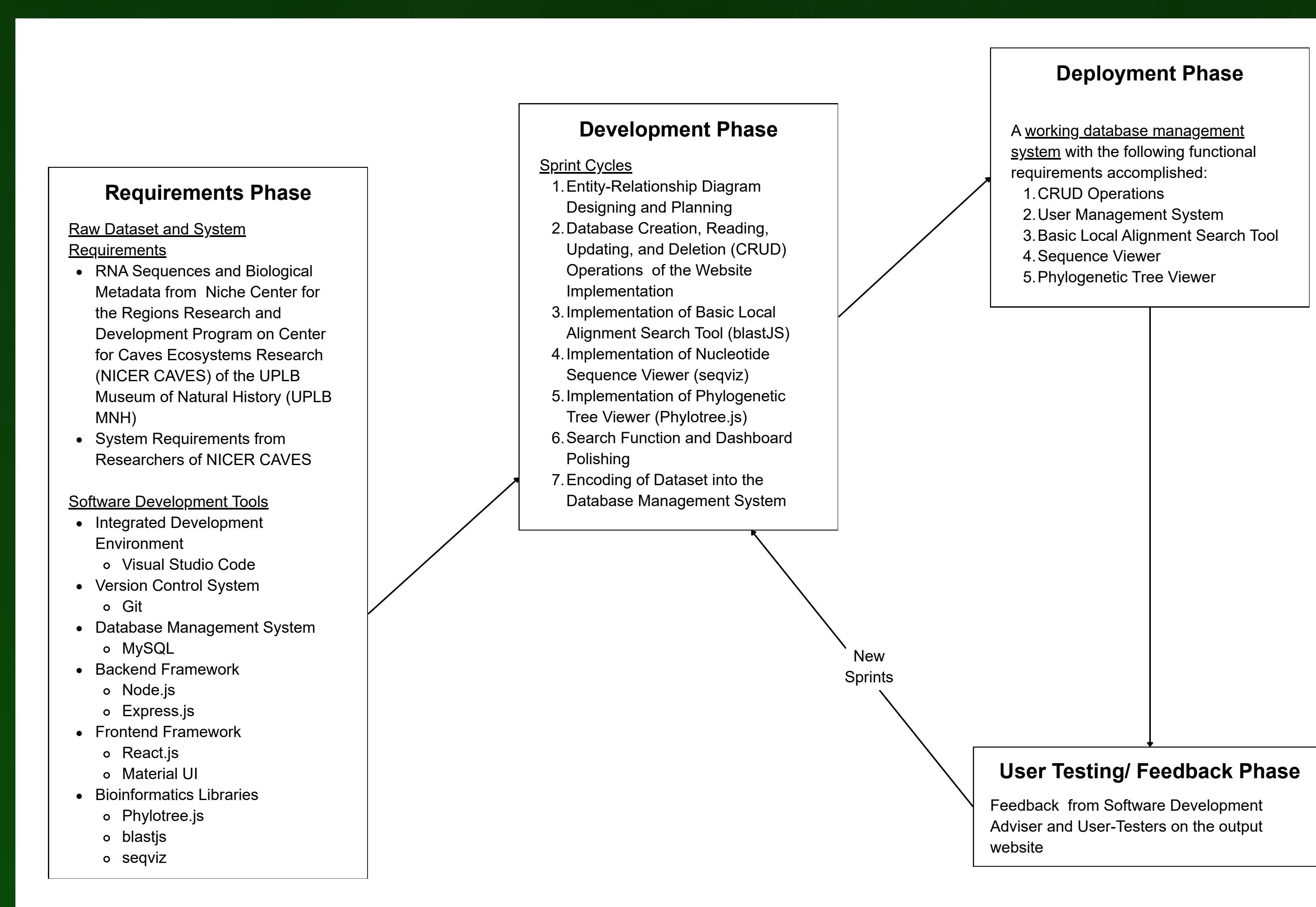


Figure 1. Operational Framework Diagram for Sprint-based Agile Methodology for Software Development

The process on how the web-based database management system BatSEQ was developed is visualized in the operational framework in Figure 1. The operations of web development for this project aimed to implement a sprint-based development process that incorporates an added feedback loop that allows the concurrent improvement of features during the development of the web development project. It was



For full list of references and journal-type paper, scan the attached QR code.

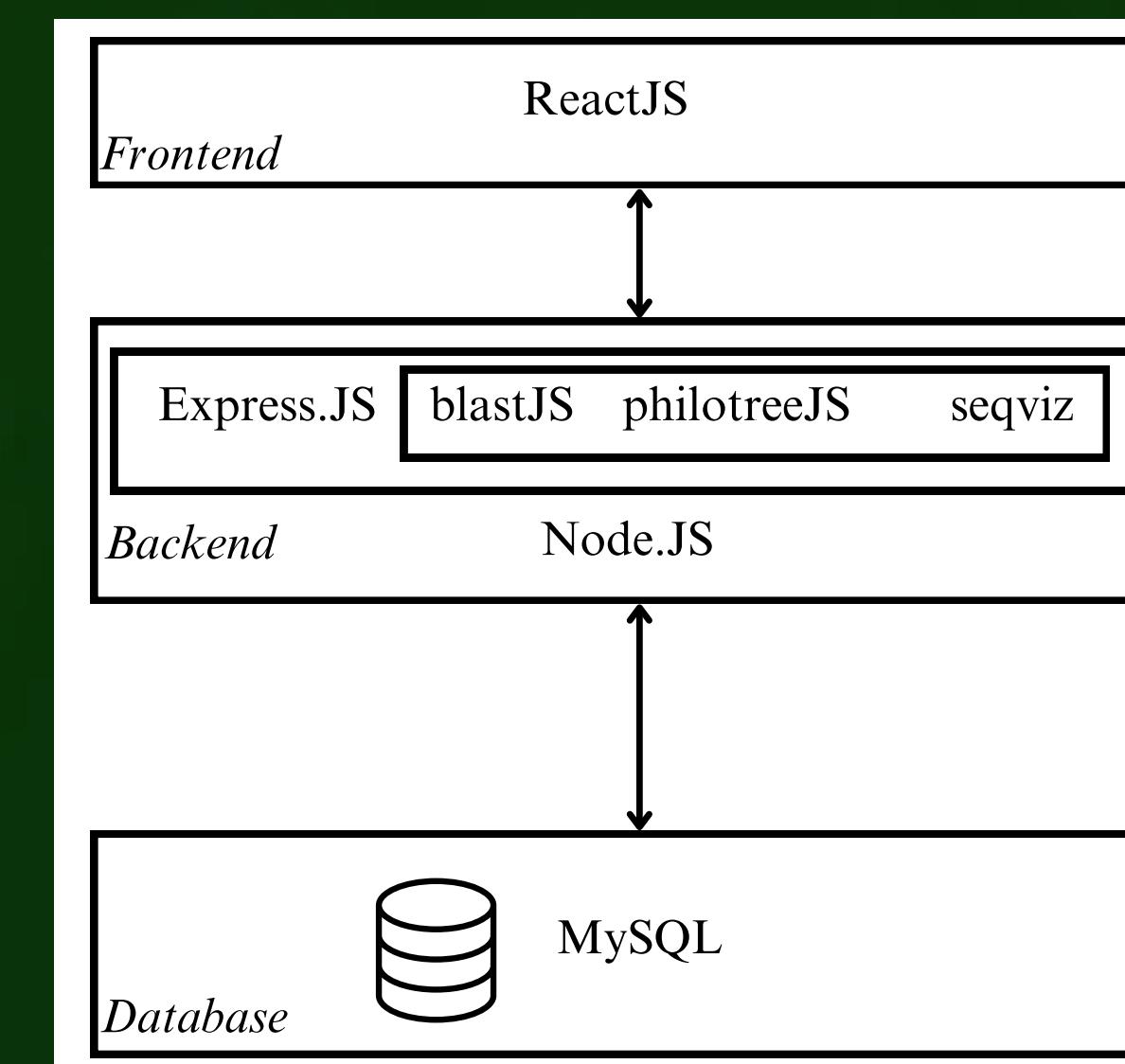


Figure 2. System Architecture Diagram of the MERN Stack Environment with the bioinformatics/visualization libraries

The system architecture of BATSeq is patterned after the standard interaction of a three-sectioned model for MERN (using MySQL instead of MongoDB) applications to ensure seamless data processing (see Figure 2). For the frontend portion of the system, ReactJS together with components from Material UI was used to handle the graphical user interface representation of the database management system. The backend ran primarily on Node.js and was structured using ExpressJS. Node packages were responsible for sequence alignment tasks as well as running visualization and nucleotide representation processes, aided by npm packages such as blastjs, seqviz, and phylotree.js.

Lastly, a MySQL database stored the RNA sequences, metadata for each isolate, user details, and image link files. The database schema supported CRUD operations for isolates, including updating RNA sequences via a FASTA file uploader.

The system was tested using the Standardized Usability Scale (SUS) and a 5-point Likert Scale User Functionality Specific Questionnaire to a pool of 11 respondents who were sample using purposive methods to ensure that all has background on microbiology.

RESULTS & DISCUSSION

Overview of the Respondents

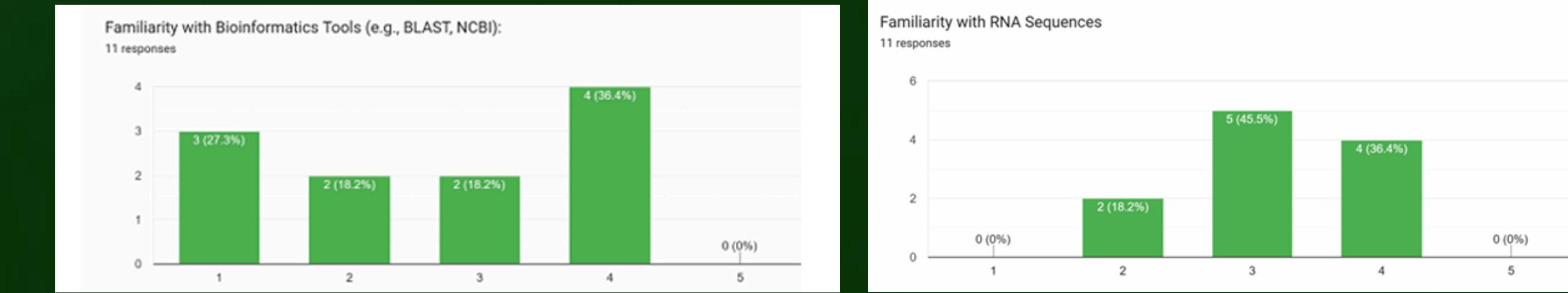


Figure 3-4. Familiarity of Respondents to Bioinformatics Tools (left) and Familiarity with RNA Sequences (right)

For the demographics of the respondents during the testing phase of the project, majority of participants the participants came from the College of Arts and Sciences, 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 ($\frac{1}{5}$ scale) to having advanced familiarity ($\frac{4}{5}$). (See Figure 3-4). Two of the participants were involved in the NICER Caves Program.

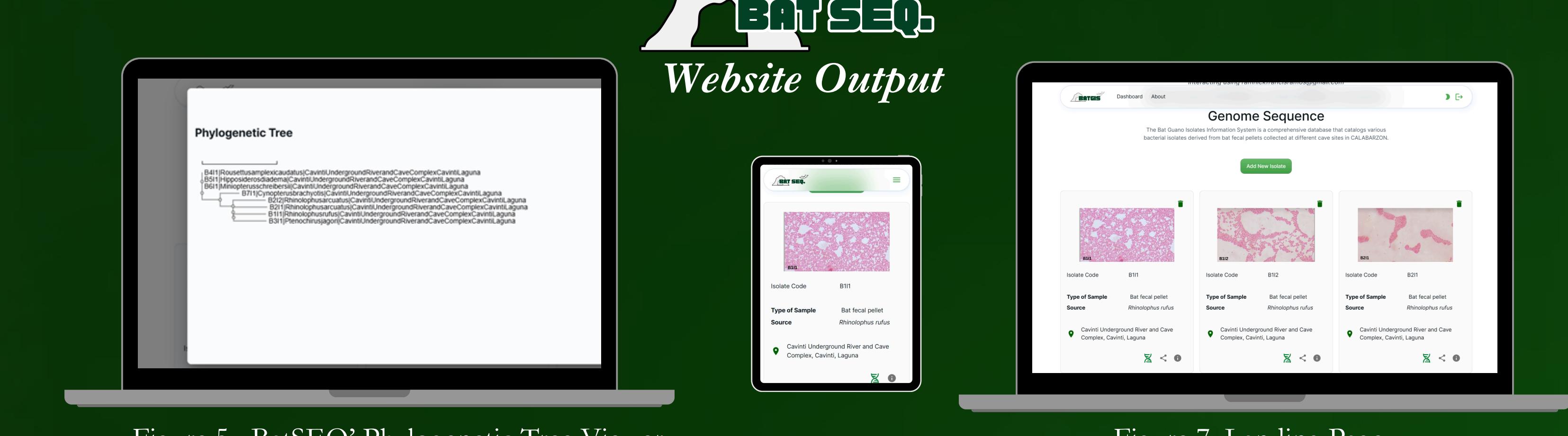


Figure 5. BatSEQ's Phylogenetic Tree Viewer

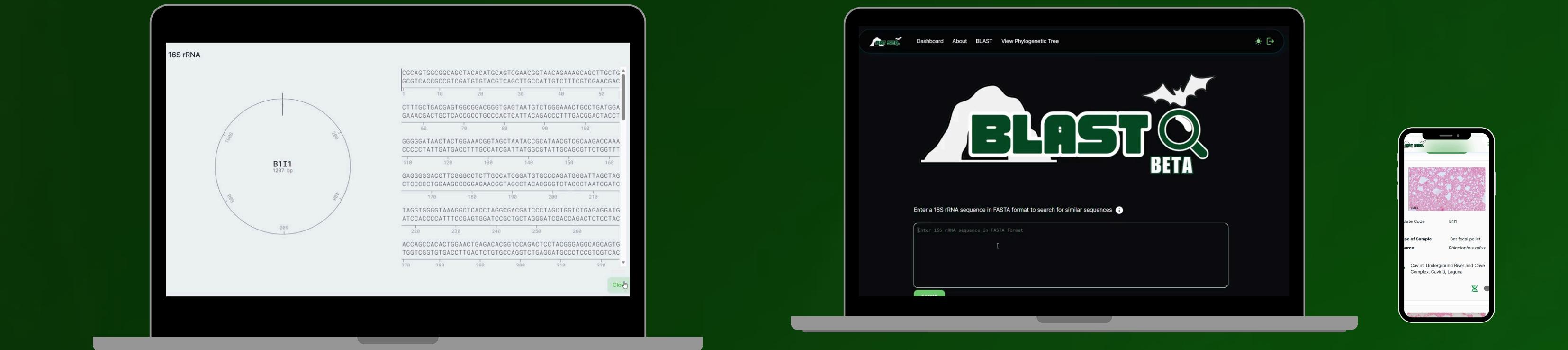


Figure 6. BatSEQ's Sequence Visualizer

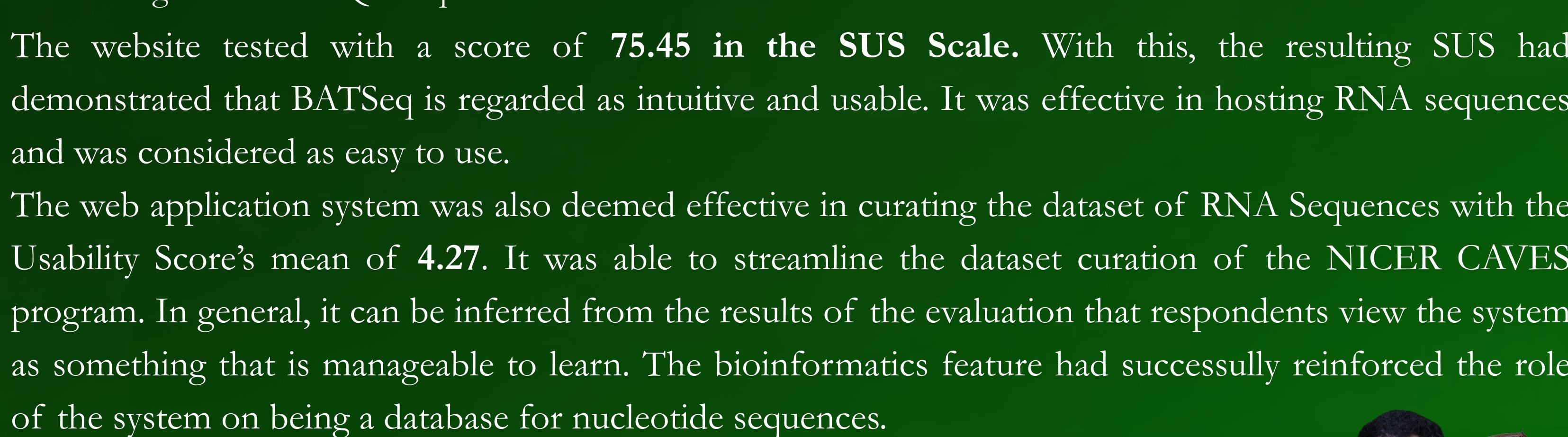


Figure 7. Landing Page

Figure 8. BatSEQ's Basic Local Alignment Search Tool

The website tested with a score of **75.45 in the SUS Scale**. With this, the resulting SUS had demonstrated that BATSeq is regarded as intuitive and usable. It was effective in hosting RNA sequences and was considered as easy to use.

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 dataset curation 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 successfully reinforced the role of the system on being a database for nucleotide sequences.

About the Author

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.

