

The Usability Testing of SSAAT, a Bioinformatic Web Application for DNA Analysis at a Nucleotide Level

Victor Mero

School of Computational and Communication Science and Engineering
Nelson Mandela African Institution of Science and Technology (NM-AIST)
Arusha, Tanzania
merov@nm-aist.ac.tz

Dina Machuve

School of Computational and Communication Science and Engineering
Nelson Mandela African Institution of Science and Technology (NM-AIST)
Arusha, Tanzania
dina.machuve@nm-aist.ac.tz

Abstract—Sanger sequencing remains the cornerstone method for Deoxyribonucleic Acid (DNA) sequencing due to its high accuracy in targeting smaller genomic regions in a larger number of samples. The analysis of Sanger sequence DNA data requires powerful and intelligent software tools. Most of the preferred tools are proprietary licensed tools that offer a user-friendly interface and have many features, however, their affordability, especially to individual scientists or students, is limited. On the other hand, a few free and open-source licensed tools are available but have limited features. This study focuses on the usability testing of the developed Sanger Sequence Automatic Analysis Tool (SSAAT), a free and open-source web tool for Sanger sequence analysis. Usability tests were conducted with potential users and the results demonstrate that the participants were able to use the tool easily and accomplish the test tasks at the given time. Moreover, the participants were excited with the easy-to-use interface and agreed that most users could use the tool with no need for technical assistance. However, the participants also identified some issues that require more development effort.

Keywords—Sanger sequence; usability; bioinformatics; web tool

I. INTRODUCTION

A. Background

Sanger sequencing technique is one of the most famous methods used for determining nucleotide sequences in DNA [1], due to its high sequencing accuracy compared to the Next Generation Sequencing (NGS) technologies and its efficiency in sequencing short fragments of DNA, ranging from 200 base pairs (bp) to around 1,000 bp. Sanger sequencing is extensively used to the fields of functional and comparative genomics, evolutionary genetics, and complex disease research. Particularly, the method was employed in sequencing the first human genome in 2000 [2]. The Sanger sequencing process is composed of a pipeline from the DNA extraction to the generation of a chromatogram which is stored as a file called AB1. This process can be seen in [3].

The Sanger sequencing quality relies on the "base calling quality", i.e. the relative certainty with which the nucleobases are determined [2]. Assessing the base-calling accuracy is usually performed using the visual inspection of the sequence trace chromatogram. Most often, proprietary software like CLC Genomics Workbench (Qiagen), SeqMan (DNASTAR), etc. are preferred due to their user-friendly interfaces and the features they provide. Nevertheless, some free open-source software tools for Sanger sequence analysis exist. Phred was among the earliest base-calling software tools reported to have less error rate than the ABI machine software [4-6]. However, Phred was developed as an open-source resource but it is not freely available [7]. Tracy on the other hand is a free open-source tool for Sanger sequence analysis that performs base-calling and other tasks including sequence alignment, assembly, and deconvolution of Sanger chromatogram trace files, all in a command-line interface [8]. Moreover, SangerseqR [9], Automated Sanger Analysis Pipeline (ASAP) [10], and SeqTrace [7] are also reported open-source tools for performing Sanger sequence data analysis but have limited graphical user interface and cross-platform capabilities. Some web tools are also available including Indelligent, CHILD, and Mixed Sequence Reader they are limited to a single feature usage [11-13]. The Sanger Sequence Automatic Analysis Tool (SSAAT), unlike the aforementioned tools, was developed as a web-based tool to eliminate the cross-platform issues while providing an easy-to-use interface and more DNA analysis features.

B. Usability

Usability refers to how easily a user of a specific product or design can use it to accomplish the intended goals effectively, efficiently, and acceptably [14]. In the field of human-computer interaction, usability is defined as a way to remove all possible frustrations that users may experience when using a product or design. On the other hand, usability evaluation refers to a method used in the central design to assess a product or design by testing it with a group of representative users [15], and a platform for users to give direct feedback and

recommendations [16]. Usability is a result of the basic quality components which are learnability, efficiency, memorability, error tolerance, and satisfaction [17]. There are several methods for testing products and among them, usability testing and heuristic evaluation have been the most appropriate ones [18]. Heuristic evaluation is mostly done by professionals who use generally accepted guidelines to evaluate the usability of the product through demos and report issues. In contrast, usability testing recruits users to evaluate a particular product's usability through their feedback after interacting with it [18]. The current study aims to explore the usability of SSAAT, which was developed as a user-friendly web tool for analyzing Sanger sequence data at the nucleotide level. The study focused on the task completion rates, the meantime to complete tasks, and system usability.

II. MATERIALS AND METHODS

A. SSAAT Tool Description

SSAAT was created to provide an easy-to-use interface for molecular biologists to perform Sanger sequence DNA analysis. SSAAT reads the DNA input files, performs base-calling, and provides chromatogram visualization, DNA sequence alignment, and polymorphism detection, and delivers a structured report of the analysis. The flow diagram of SSAAT (Figure 1) illustrates the working flow of the tool.

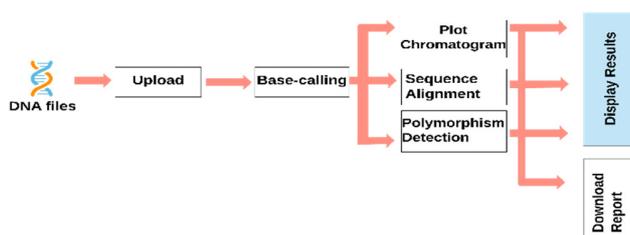


Fig. 1. Flow diagram of SSAAT.

SSAAT utilizes a web-based interface that allows the molecular biologist to perform DNA analysis in a user-friendly graphical interface and also enables working in cross-platform environments. One important element in accomplishing the goals of SSAAT is to ensure that its design satisfies the needs of the users and is flexible to handle the continued evolution of this emerging area.

B. Usability Testing Methodology

In this study, both qualitative and quantitative methods were used to capture user interactivity with the web tool. Qualitative data were collected through a Likert scale questionnaire while quantitative data, such as total users who were able to complete all the tasks, total complete tasks, complete task time, etc., were also collected. The testing session was designed in a way that each participant performed all the tasks and summative assessment was done to examine and evaluate participants' insights. The study aimed at capturing the indicating factors of usability such as learnability, efficiency, usefulness, and satisfaction through the conducted test session.

1) Participants and Duration

A total of 15 individuals participated in the usability test sessions. Based on previous studies, 5-20 participants are a valid sample for usability testing [19]. The first 3 participants were scheduled on the first day and were regarded as a pilot for the next sessions. The testing sessions were conducted for 5 days (3 participants per day). The duration for each session was 60 minutes and after every session, a break period of 60 minutes was given. The session duration was based on previous studies on usability which suggest that 60-90 minutes time is valid for test sessions [19]. During the testing sessions, the moderator provided a brief overview of the test session and requested the participants to fill in a pre-test questionnaire in order to collect the general data. The participants then read the task instructions and began to perform the tasks on the tool using a web browser. As soon as the participants completed all the tasks, the moderator requested the participants to rate the web tool (SSAAT) using the Likert scale questionnaire. This was done as a post-test session in order to find out more information about the overall usability of the web tool.

2) Tasks

Table I presents the results of the tasks that were obtained during the test sessions. Each participant was required to attempt the tasks and the moderator observed the time of completion and participant's behavior while attempting the tasks.

TABLE I. SUMMARY OF USABILITY TEST RESULTS BY TASK

Code	Tasks	Baseline time (min)/estimated time (min)	Mean time (min)	Completion rate (%)
Task 1	Identify the use of the web tool.	3/5	0.56	100
Task 2	File upload, view the sequence quality, and download the extracted sequence as a FASTA file.	3/5	4	87
Task 3	Navigate to Chromatogram Viewer, trim the 5' end 50 base and trim 3' end 100 base and download the chromatogram as a PDF file.	7/10	7.3	73
Task 4	Upload a reference sequence and calculate the global/local sequence alignment with the previously uploaded file as a primary sequence.	7/10	9	60
Task 5	Generate a report with sequence detail, chromatogram quality score plot, and sequence alignment results.	3/5	3.1	73

III. RESULTS AND DISCUSSION

A. Participant Characteristics

Among the 15 participants, there were 7 masters and three PhD students in life sciences, 3 molecular biologists, and 2 bioinformaticians. Eight were women and 7 were men, aged between 21 to 47 years old. All the participants claimed to use the internet daily. None of the participants received formal training or had the chance to review a user guide before participating in the usability test.

B. Task Completion Rate Results

All participants successfully completed Task 1 marking a 100% completion rate. Task 2 was completed by 87% of the participants. Among the long duration tasks, Task 3 scored 73% completion rate, while Task 4 scored 60% completion rate, and Task 5 was successfully completed by the 73% of the participants. The task definitions, completion rate, and mean time are illustrated in Table I.

C. Mean Time to Complete Task Results

The moderator recorded the task execution time for each participant. The allocated time for each task ranged from 5 to 10 minutes where simple tasks were allocated with less time and lengthy tasks with more. Task 1 had the shortest completion time with a mean time of 0.56 minutes. This was followed by Task 5 and Task 2, with times of 3.1 minutes and 4 minutes respectively. Task 3 and Task 4 were the longest to complete with mean times of 7.3 minutes and 9 minutes respectively. The overall completion time ranged from 0.56 to 9 minutes, with a commonly recorded time of less than 5 minutes for the majority of tasks.

D. System Usability Survey Results

During the post-testing session, the participants were asked to rate the web tool to capture the general usability aspects of the SSAAT. The detailed results are shown in Table II. The measurements which were captured from participants' post-test questionnaires included:

- Its ease of use
- If the users would prefer to use the web tool
- Its learnability
- If assistance from technical personnel was required
- System functionality integrations
- If the participants would recommend the tool to a colleague

The majority of the participants (86.67%) agreed that the web tool was easy to use. Additionally, most participants (93%) reported that they would prefer to work with the web tool often. Regardless of the higher scores of participants agreeing that the tool was easy to use, 40% of them agreed that technical assistance was needed to operate the tool effectively. More than half of the participants agreed that the integrated features were functioning well. Lastly, the majority of participants reported that they would recommend the tool to a colleague.

E. Discussion

The usability of SSAAT a bioinformatic tool for DNA analysis at the nucleotide level was assessed in this study. Our findings suggest that SSAAT is easy and learnable, and even new users may be able to use the tool without prior exposure or technical assistance and accomplish the required tasks at a given time. The participants identified a number of possible improvements to the tool such as the addition of batch processing capabilities, trace file editing, and connection to remote DNA databases. However, the mentioned suggestions would require more development effort and time, therefore we plan to work on them in the future versions. Other modifications such as the suggested lighter interface background colors and the creation of a user guide with some visual illustrations were easier and more straightforward to implement.

TABLE II. SYSTEM USABILITY SURVEY RESULTS

Statement	Response	Frequency
"The web tool is user friendly"	Strongly agree	5 (33.33%)
	Agree	8 (53.33%)
	Neutral	2 (13.33%)
"I would like to use web tool often"	Strongly agree	4 (26.67%)
	Agree	10 (66.67%)
	Neutral	1 (6.67%)
"I think most of the users will be able to use the web tool fast"	Strongly agree	4 (26.67%)
	Agree	8 (53.33%)
	Neutral	1 (6.67%)
	Disagree	2 (13.33%)
"I will not need technical assistance to be able to use the web tool"	Agree	7 (46.67%)
	Neutral	2 (13.33%)
	Disagree	4 (26.67%)
	Strongly disagree	2 (13.33%)
"I think the web tool units/parts are well integrated"	Strongly agree	4 (26.67%)
	Agree	6 (40%)
	Neutral	4 (26.67%)
	Disagree	1 (6.67%)
"I will recommend this web tool to my colleagues"	Strongly agree	4 (26.67%)
	Agree	9 (60%)
	Neutral	2 (13.33%)

IV. CONCLUSION

The primary focus of this study was to examine the usability of the developed SSAAT tool. This is an attempt towards eliminating the barriers to the availability of free and user-friendly software for Sanger sequence DNA analysis. The usability assessment results suggest that most of the users will be able to use the web tool without assistance. This is a good indicator that the tool is easy to use and hence most of the users are likely to often use it in their work and would probably recommend it to their colleagues.

The participants in this usability study encountered several minor usability issues while using the prototype of this web tool. To ensure the effective use of SSAAT, the issues identified during the usability testing sessions will be addressed in the upcoming version of the tool and the analysis results will affect the future development. Usability evaluations are invaluable to the success of technology in an emerging area, especially in a complex domain such as genetics.

ACKNOWLEDGEMENT

The authors would like to acknowledge the support from the Bioinformatics Unit at the Beca-ILRI hub in Nairobi Kenya and CoCSE Laboratory at NM-AIST in Arusha Tanzania.

REFERENCES

- [1] J. Shendure *et al.*, "DNA sequencing at 40: past, present and future," *Nature*, vol. 550, no. 7676, pp. 345–353, Oct. 2017, <https://doi.org/10.1038/nature24286>.
- [2] B. Ewing, L. Hillier, M. C. Wendl, and P. Green, "Base-Calling of Automated Sequencer Traces Using Phred. I. Accuracy Assessment," *Genome Research*, vol. 8, no. 3, pp. 175–185, Mar. 1998, <https://doi.org/10.1101/gr.8.3.175>.
- [3] <https://commons.wikimedia.org/wiki/File:Sanger-sequencing.svg>
- [4] "Applied Biosystems Genetic Analysis Data File Format." Applied Biosystems, Jul. 2006.
- [5] B. Ewing and P. Green, "Base-calling of automated sequencer traces using phred. II. Error probabilities," *Genome Research*, vol. 8, no. 3, pp. 186–194, Mar. 1998.
- [6] M. Machado *et al.*, "Phred-Phrap package to analyses tools: a pipeline to facilitate population genetics re-sequencing studies," *Investigative Genetics*, vol. 2, no. 1, Feb. 2011, Art. no. 3, <https://doi.org/10.1186/2041-2223-2-3>.
- [7] B. J. Stucky, "SeqTrace: A Graphical Tool for Rapidly Processing DNA Sequencing Chromatograms," *Journal of Biomolecular Techniques: JBT*, vol. 23, no. 3, pp. 90–93, Sep. 2012, <https://doi.org/10.7171/jbt.12-2303-004>.
- [8] T. Rausch, M. H.-Y. Fritz, A. Untergasser, and V. Benes, "Tracy: basecalling, alignment, assembly and deconvolution of sanger chromatogram trace files," *BMC Genomics*, vol. 21, no. 1, Mar. 2020, Art. no. 230, <https://doi.org/10.1186/s12864-020-6635-8>.
- [9] J. T. Hill and B. Demarest, *sangerseqR: Tools for Sanger Sequencing Data in R*. Bioconductor version: Release (3.12), 2021.
- [10] A. Singh and P. Bhatia, "Automated Sanger Analysis Pipeline (ASAP): A Tool for Rapidly Analyzing Sanger Sequencing Data with Minimum User Interference," *Journal of Biomolecular Techniques: JBT*, vol. 27, no. 4, pp. 129–131, Dec. 2016, <https://doi.org/10.7171/jbt.16-2704-005>.
- [11] C.-T. Chang *et al.*, "Mixed Sequence Reader: A Program for Analyzing DNA Sequences with Heterozygous Base Calling," *The Scientific World Journal*, vol. 2012, Jun. 2012, Art. no. e365104, <https://doi.org/10.1100/2012/365104>.
- [12] D. A. Dmitriev and R. A. Rakitov, "Decoding of Superimposed Traces Produced by Direct Sequencing of Heterozygous Indels," *PLOS Computational Biology*, vol. 4, no. 7, 2008, Art. no. e1000113, <https://doi.org/10.1371/journal.pcbi.1000113>.
- [13] I. Zhidkov, R. Cohen, N. Geifman, D. Mishmar, and E. Rubin, "CHILD: a new tool for detecting low-abundance insertions and deletions in standard sequence traces," *Nucleic Acids Research*, vol. 39, no. 7, Apr. 2011, Art. no. e47, <https://doi.org/10.1093/nar/gkq1354>.
- [14] M. Shitkova, J. Holler, T. Heide, N. Clever, and J. Becker, "Towards Usability Guidelines for Mobile Websites and Applications," in *Wirtschaftsinformatik Proceedings*, Mar. 2015.
- [15] P. T. Koziokas, N. D. Tselikas, and G. S. Tselikis, "Usability Testing of Mobile Applications: Web vs. Hybrid Apps," in *Proceedings of the 21st Pan-Hellenic Conference on Informatics*, New York, NY, USA, Sep. 2017, Art. no. 55, <https://doi.org/10.1145/3139367.3139410>.
- [16] J. R. Bergstrom and A. Schall, Eds., *Eye Tracking in User Experience Design*, 1st ed. Amsterdam, Netherlands; Boston, MA, USA: Morgan Kaufmann, 2014.
- [17] E. Folmer and J. Bosch, "Architecting for usability: a survey," *Journal of Systems and Software*, vol. 70, no. 1, pp. 61–78, Feb. 2004, [https://doi.org/10.1016/S0164-1212\(02\)00159-0](https://doi.org/10.1016/S0164-1212(02)00159-0).
- [18] A. Fernandez, E. Insfran, and S. Abrahão, "Usability evaluation methods for the web: A systematic mapping study," *Information and Software Technology*, vol. 53, no. 8, pp. 789–817, Aug. 2011, <https://doi.org/10.1016/j.infsof.2011.02.007>.
- [19] J. Nielsen, "Usability for the MassesJUS," *Journal of Usability Studies*, vol. 1, no. 1, pp. 2–3, Nov. 2005.