Usability Evaluation of Origin of Replication Finding Tools

Isra Al-Turaiki(✉), Maryam Aloumi, Nour Aloumi, Noorah Almanyi,
Khulood Alghamdi, and Sarah Almuqhim

College of Computer and Information Sciences, King Saud University,
Riyadh, Kingdom of Saudi Arabia
ialturaiki@ksu.edu.sa, maryaloumi@gmail.com, nouraloumi@gmail.com,
noorah556@gmail.com, khulood.alghamdi@gmail.com, s.almoqhim@hotmail.com

Abstract. Nowadays, there is a significant increase in the number of Bioinformatics tools and databases. Researchers from various interdisciplinary fields need to use these tools. Usability is an important quality of software in general, and bioinformatics tools in particular. Improving the usability of bioinformatics tools allows users to use the tool to its fullest potential. In this paper, we evaluate the usability of two online bioinformatics tools *Ori-Finder 1* and *Ori-Finder 2* in terms of efficiency, effectiveness, and satisfaction. The evaluation focuses on investigating how easily and successfully can users use Ori-Finder1 and Ori-Finder 2 to find the origin of replication in Bacterial and Archaeal genomes. To the best of our knowledge, the usability of these two tools has not been studied before. Twelve participants were recruited from four user groups. The average tasks completion times were compared. Many usability issues were identified by users of bioinformatics tools. Based on our results, we list recommendations for better design of bioinformatics tools.

Keywords: Origin of replication · Bioinformatics
Usability evaluation · oriC

1 Introduction

DNA replication is the process of copying a DNA molecule in order to produce two identical replica. This is a key step for the reproduction of living cells. It is initiated at a specific region in the DNA called the *origin of replication* (oriC). The whole process is regulated by a mechanism that recognizes both, the locations of oriC and the ideal time to start replication. In bacterial and archaeal genomes, there is usually a single oriC [13]. With the development of high-throughput sequencing technologies, there has been a rapid increase in the release of bioinformatics databases and tools. Ori-Finder 1 [9] and Ori-Finder 2 [14] are two web-based bioinformatics tools designed for predicting the location of oriC in bacterial and archaeal genomes, respectively. The two tools have been widely adopted by researchers for the identification and analysis of oriC.

Developers of bioinformatics software usually focus on validating the biological hypothesis. However, one feature that receives little attention is the usability of the bioinformatics tool itself. According to [5], the ISO 92491 defines usability as "*the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use*". This feature is very important quality of interactive bioinformatics tools. Failing to achieve a certain level of usability may result in many bioinformatics tools being unused. For this reason, evaluating the effectiveness and usability of bioinformatics tools is increasingly getting attention.

In this study, we investigate the usability of Ori-Finder 1 and Ori-Finder 2. To the best of our knowledge, the usability of these two tools has not been previously studied. Our goal is to highlight the main usability problems and to provide recommendations for better design. The rest of the paper is organized as follows: in Sect. 2 we discuss similar work for the evaluation of bioinformatics tools. Section 3 describes the two bioinformatics tools evaluated in this study. In Sect. 4, we discuss our methodology and the details of data collection. Results and design recommendations are discussed in Sects. 5 and 6. Finally, the paper is concludes in Sect. 7.

2   Related Work

The number of bioinformatics tools and users increase is growing. There is a need to ensure that the highest standards of usability are being met. Usability is an important aspect for the survival of software and bioinformatics tools are no exception. In this section, we discuss the literature of usability studies performed on bioinformatics tools and databases.

Mirel and Wright [15] highlighted the need to consider *usability* as a main goal when designing software for the scientific community. The authors focused on bioimaging software, proposing several criteria to be met, including: user and developer friendliness, interoperability, modularity, and results validation.

Bolchini et al. [5] conducted two usability studies in order to identify some critical usability problems in bioinformatics web-based databases. In the first study, a usability inspection using MILE+ protocol [6] was carried out to analyze the navigation and information architecture design of CATH database. CATH (http://www.cathdb.info/wiki) is a browsing-oriented protein classification database. Usability issues were identified in the navigation of different subsystems each with many releases. The authors showed that the user may be led to an old release of a subsystem. In addition, there were limited access paths available for content navigation. In the second study, user testing was conducted on three search-oriented databases: BioCarta (www.biocarta.com), Swiss Prot (www.expasy.ch/sprot), and NCBI (www.ncbi.nlm.nih.gov). Users had issues in search query formulation and in interpreting search results.

Mullany et al. [16] evaluated the effectiveness and usability of six existing bioinformatics databases. The authors used thirteen criteria, for each, a set of yes/no question were posed. The answers were coded with 1 for yes, −1 for no, and 0 for unknown. This enabled the calculation of a total score summarizing the

effectiveness and usability of a database. Many limitations were identified across multiple databases, including: poor documentation, lack of pathway output, lack of database updates, and inconsistencies in nomenclature.

Al-Ageel et al. [2] evaluated the usability of four web-based bioinformatics tools for structure and sequence motif finding. MEME [4], FIMO [10], RNAMST [7], and RNAPromo [17] were inspected using a list of heuristics proposed by [15]. Several usability issues were identified in the inspected tools, such as: too much detailed results, poor user interface designs, and lack of tools for powerful interactions. However, strength points were also identified in their study. MEME and FIMO provide adequate documentation and examples that help users search for information about the tool and find steps required to perform a given task. The study also showed that FIMO had the most usability issues while MEME had least usability issues.

3   oriC Finiding Tools

We conduct a usability study for two well-known web-based tools for finding oriC. Here, we give a brief description of each tool.

3.1   Ori-Finder 1

Ori-Finder 1 [9] is a web server for predicting oriCs in bacterial genomes. Locating oriCs using Ori-Finder is based on an algorithm incorporating base composition analysis using Z-curve method, distribution of DnaA boxes, and the frequency of genes near oriCs. The server accepts sequences in FASTA format as an input. In addition, other parameters need to be set, such as: species-specific DnaA boxes, protein table, and display parameters. The output page shows the predicted region of oriC along with detailed information, including: oriC length, genome length, and DnaA box distribution.

3.2   Ori-Finder 2

The Ori-Finder 2 [14] is also a web-based tool for the prediction and analysis of oriCs, but in archaeal genomes. For annotated genomes, the tool accepts a sequence file in GenBank format or in FASTA format with corresponding protein table file. Ori-Finder 2 can analyze unannotated genomes by utilizing ZCURVE1.02 [11], Glimmer3 [8], and BLAST [3]. The workflow of Ori-Finder 2 is composed of FIMO and REPuter [12] for searching motifs and repeats, respectively.

4   Methodology

4.1   User Groups

For the usability test of the two bioinformatics tools we recruited participants from King Saud University, Riyadh, Saudi Arabia. Four of them were postgraduate students in the *microorganisms biology* department with no previous experience in using bioinformatics tools. Three of our participants were HCI experts

who previously taught HCI courses in the *Information Technology* department. One of them had knowledge in both HCI and usability. Our study also included four beginners in bioinformatics. They were postgraduate students in *Information Technology* who took one course in bioinformatics. Finally, one bioinformatics expert was invited to participate. In total, there were twelve female participants between the age of 18–40 years. The usability test session took about two weeks.

4.2   Test Scenario and Goals

For this usability evaluation, a scenario similar to a real application was designed. The scenario describes a situation where there is a need to locate origin of replication in Bacterial and Archaeal genomes using Ori-finder 1 and Ori-finder 2, respectively. Here, we describe the scenario, goals and tasks.

Scenario: *You are a member of a group of researchers looking for a treatment for a disease caused by Bacteria (or Archaea). The process of genome replication is one of the most important tasks carried out in the cell. One way that can help in treatment is targeting the origin of replication of the Bacteria (or Archaea) in order to inhibit its replication. Therefore, you will use Ori-Finder1 or (Ori-Finder 2) web-based systems designed to predict the origins of replication in Bacteria (or Archaea).*

Goal 1: Predict the oriC in a Bacterial genome and interpret the results.

– Task 1: *Now, you want to try to use Ori-Finder 1 web-based system at* http://www.tubic.tju.edu.cn *using the sequence example provided by the tool for Escherichia coli. The mismatches between the DnaA boxes are allowed to be two mismatches as you have decided with your colleagues.*
  This task was further divided into the following three subtasks: Task 1.1: Use the example provided to upload a complete genome sequence in FASTA format. Task 1.2: Set the mismatch site to 2. Task 1.3: Submit sequence form.
– Task 2: *Now, you are about to inform your colleagues about the results you found using Ori-Finder 1. You want to report interesting information about the number of DnaA boxes, the location of the oriC region, the DnaA boxes identified in the sequence, as well as show them some relative Z-curves.*
  This task was further divided into the following subtasks: Task 2.1: Find the number of DnaA box. Task 2.2: Find the location of oriC region. Task 2.3: Find The sequence of oriC. Task 2.4: Review the results in the form of a curve.

Goal 2: Predict the oriC in an Archaeal genome and interpret the results.

– Task1: *Now, its time to try to use the Ori-Finder 2 web-based system and run the tool on an annotated archaeal genome. You are going to do the experiment for Pyrococcus abyssi GE5 species which is a sequenced and annotated*

*archaeal genome from NCBI genome database. Based to your colleague's deci-
sion, your are going to use Thermococcaceae for the motif taxonomy.*
Task 1.1 Select a sequenced and annotated archaeal genome from NCBI
genome database. Task 1.2 Select the Motifs taxonomy to Thermococcaceae.
Task 1.3 Submitting sequence form.
– Task 2: *Now, its time to inform your colleagues about the results you found.
You will report interesting information about the location of the oriCs with
ORB sequence and the DnaA boxes identified in the sequence.*
Task 2.1 Find the location of oriCs with ORB sequence. Task 2.2 Find The
sequence of oriC (i.e. the DnaA boxes identified in the sequence).

4.3 Pre-test and Post-test Questionnaires

Participants were asked to take a pre-session questionnaire. The purpose of the
pre-session questionnaire was to gather information about: demographics, expe-
rience and frequency of using Bioinformatics tools, and their expectations about
utilizing Bioinformatics tools compared to manual analysis. The questionnaire
included both open-ended, closed-ended, and scale questions. After the partici-
pants completed the usability test for each tool, we asked them to take a post-
session questionnaire. The questions were related to participants' impression of
the tool after performing the tasks. The questions were related to issues that
affect the usability of the system.

Our goal was to find out if the tools met the users expectation. We were
interested in knowing whether they have experienced any problems with the
design, layout, navigation, or output of each tool. Questions had a rating scale
to assess the participant's overall reaction to the tools usability and satisfaction
level. We also asked them whether they would use the tool in the future and
whether they would recommend it to a friend. Finally, we asked them to provide
their suggestions to improve these tools.

To get the participants overall impression, we finalized the usability test with
a short interview with the following questions:

1. How was your experience?
2. What did you like and did not like while using the systems?
3. What is your overall impression?

Most of the participants appreciated the availability of such tools to facilitate
the work in bioinformatics research. However, they pointed out the need to
enhance user interfaces for more clarity and better understandability. This will
help users better utilize such tool to their full potential.

4.4 Usability and Testing Session

The observation usability evaluation method was used during for the test ses-
sions. This involved watching the participants while they interact with the web-
based tools, taking notes, and asking questions. The testing environment setting

was an electronic observation room setup. In this setting, there are two rooms. The first room is for the test moderator and the participant. The test moderator sits close to the participant having an excellent view of what is going on with the participant, while making her feel comfortable. The observers are in the second room. They are physically separated from the testing activity.

The test takes about 20 to 30 min for each participant, depending on their experience.

The collected data from the test sessions were classified into: performance and preference data. The performance data included: task completion time, completion status (completed successfully, completed with difficulty, failure). We calculated the average value per user group. In terms of preference data, we have the participants' opinions, expectation, and experience from pre-session and post-session questionnaires. All data were collected, summarized, and recorded in a spreadsheet for easy processing.

5   Results

The average time of task completion was compared to the expected completion time in order to measure efficiency. Participant's opinions from pre-session and post-session questionnaires were used as a measure of satisfaction. Usability testing data, including questionnaires, were automatically collected using *Morea software* [1]. In this section, we summarize the main results obtained from our testing sessions.

5.1   Ori-Finder 1

Goal 1, Task1: Figure 2 shows the average task competition time for all subtasks across the four participant groups. As shown in the figure, all participant groups took longer time in performing subtask 1. Three user groups struggled in finding the example of the genome sequence. This implies that there would be an issue with the example provided within the tool. When a participant clicks on the example, a new page appears with the example sequence. The user is expected to copy and paste the sequence in the textbox.

The least time for all subtasks was taken by the bioinformatics expert. This was expected as she is familiar with similar bioinformatics tools. The group of postgraduate students in biology took longer time in performing all subtasks. Overall, participants spent more time performing the first subtask. In addition, we found that two user groups struggled in setting the mismatch site to 2, subtask 1.2. This option was not clear in the form. Issues regarding subtasks 1.1 and 1.2 are shown in Fig. 1. All user groups were able to submit the sequence form (subtask 1.3) easily. For subtask 1.1, 58% of participants were exceed the task expected time. However, 58% and 75% were able to complete subtasks 1.2 and 1.3, respectively, within the expected time.

Goal 1, Task2: Figure 3 shows the average task competition time for all subtasks across the four participant groups. As shown in the figure, surprisingly the

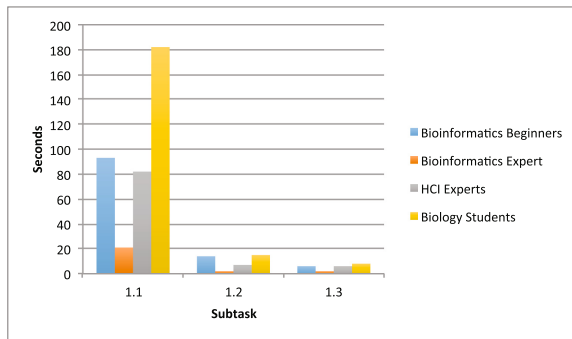**Fig. 1.** Issues found for subtasks 1.1 and 1.2.

Fig. 2. Average task completion time for Ori-finder 1 task 1 in seconds.
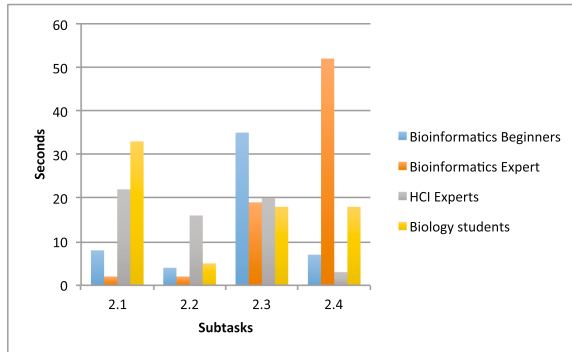

Fig. 3. Average task completion time for Ori-finder 1 task 2 in seconds.

bioinformatics expert took the longest time in identifying the oriC from the
a Z-curve (subtask 2.4). Beginners in bioinformatics took the longest time to
identify the sequence of oriC (subask 2.3). All groups found the location of
oriC region easily. Subtask 2.1, 2.2, 2.3, and 2.4 were performed within the
expected time by 75%, 83%, 58%, and 83%, of all participants, respectively.

5.2   Ori-Finder 2

Goal 2, Task1: Figure 4 shows the average task competition time for all subtasks across the four participant groups. We observed that two user groups had difficulty in selecting a sequenced and annotated Archaeal genome from the NCBI genome database (subtask 2.1). HCI experts and biology students took longer time since no action occurs when clicking on this choice to tell the user that the sequence has been selected. Users expected the sequence to be pasted in the textbox or a conformance message appears. Issue regarding subtasks 2.1 are highlighted in Fig. 5.

We also observed that all user groups had difficulty in submitting the sequence form (subtask 2.3). As shown in Fig. 6, the submit button is surrounded by other buttons and it does not appear at the end of the page as users expected. Overall, 75%, 66%, and 58% of all participants were able to complete task 1.1, task 1.2, and 1.3 within the expected time, respectively.

Goal 2, Task2: Figure 7 shows the average task competition time for all subtasks across the four participant groups. We observed that three user groups faced a problem with finding the sequence of oriC. As shown in Fig. 8, the DnaA boxes place was not clear and nothing is written to point to the place. Overall, 92% of all participants were able to complete task 1.1 within the expected time, respectively. However, 25% of all participants failed in completing task 2.3.
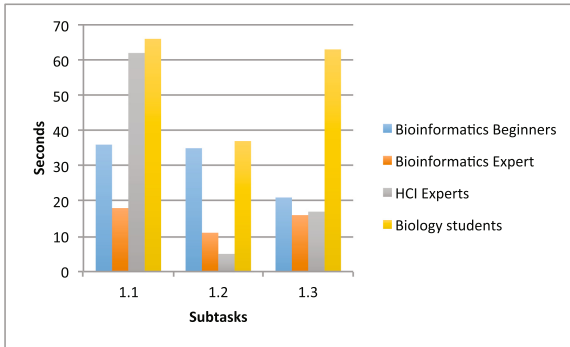

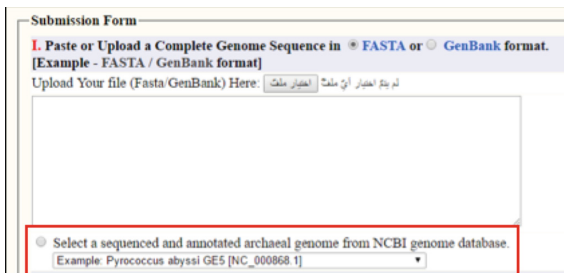**Fig. 4.** Average task completion time for Ori-finder 2 task 1 in seconds.

Fig. 5. Issues found for subtask 2.1.
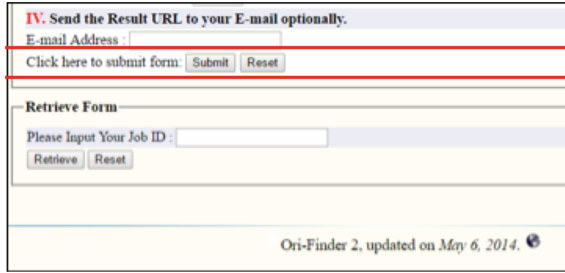
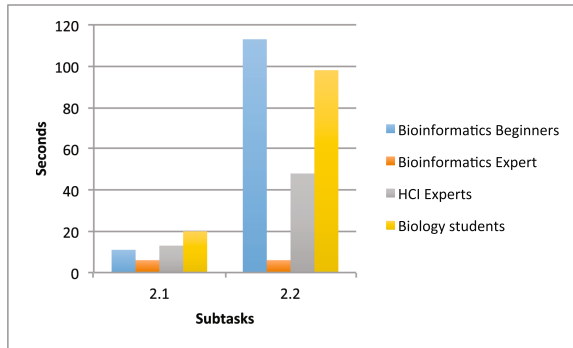**Fig. 6.** Issues found for subtask 2.3.

Fig. 7. Average task completion time for Ori-finder 2 task 2 in seconds.

Fig. 8. Issues found for subtask 2.2.

6   Recommendations

Considering all the usability issues identified in this usability evaluation of the two bioinformatics tools, we recommend the following improvements:

1. Enhancing colour scheme contrast between the text colour and the background.
2. Automatic load of sequence examples into textbox.
3. It is recommend to have a progress bar or a display of percentage of page loading in order to let the use know how long she needs to wait.
4. Provide messages and assistance to guide the user.

5. Provide explanations of technical words or use familiar icons in order to ensure that a non-expert can be conformable using the tool.
6 Improving the way of displaying the result of DNA boxes (the case of Ori-Finder 2).

7   Conclusion

Ori-Finder 1 and Ori-Finder are two popular bioinformatics tools for finding the origin of replication. In this study, the usability of these tools has been evaluated. Twelve participants were recruited from four user groups. The average tasks completion times were compared. Many usability issues were identified by users of bioinformatics tools. Based on our results, we discussed some recommendations for better design of bioinformatics tools. We hope that usability issues found in present versions of Ori-Finder 1 and Ori-Finder 2 can be addressed in future releases. We believe that research in bioinformatics usability is still in its infancy. There is a lot of room for improvement.

References

1. Usability Testing | Morae | TechSmith. https://www.techsmith.com/morae.html
2. Al-Ageel, N., Al-Wabil, A., Badr, G., AlOmar, N.: Human factors in the design and evaluation of bioinformatics tools. Procedia Manuf. **3**, 2003–2010 (2015)
3. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J.: Basic local alignment search tool. J. Mol. Biol. **215**(3), 403–410 (1990)
4. Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., Noble, W.S.: MEME Suite: tools for motif discovery and searching. Nucleic Acids Res. **37**(suppl–2), W202–W208 (2009)
5. Bolchini, D., Finkelstein, A., Perrone, V., Nagl, S.: Better bioinformatics through usability analysis. Bioinformatics **25**(3), 406–412 (2009)
6. Bolchini, D., Garzotto, F.: Quality of web usability evaluation methods: an empirical study on MiLE+. In: Weske, M., Hacid, M.-S., Godart, C. (eds.) WISE 2007. LNCS, vol. 4832, pp. 481–492. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-77010-7_47
7. Chang, T.H., Huang, H.D., Chuang, T.N., Shien, D.M., Horng, J.T.: RNAMST: efficient and flexible approach for identifying RNA structural homologs. Nucleic Acids Res. **34**(Suppl–2), W423–W428 (2006)
8. Delcher, A.L., Bratke, K.A., Powers, E.C., Salzberg, S.L.: Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics **23**(6), 673–679 (2007)
9. Gao, F., Zhang, C.T.: Ori-Finder: a web-based system for finding oriC s in unannotated bacterial genomes. BMC Bioinformatics **9**, 79 (2008)
10. Grant, C.E., Bailey, T.L., Noble, W.S.: FIMO: scanning for occurrences of a given motif. Bioinformatics **27**(7), 1017–1018 (2011)
11. Guo, F.B., Ou, H.Y., Zhang, C.T.: ZCURVE: a new system for recognizing protein-coding genes in bacterial and archaeal genomes. Nucleic Acids Res. **31**(6), 1780–1789 (2003)

12. Kurtz, S., Choudhuri, J.V., Ohlebusch, E., Schleiermacher, C., Stoye, J., Giegerich, R.: REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res. **29**(22), 4633–4642 (2001)
13. Leonard, A.C., Mechali, M.: DNA replication origins. Cold Spring Harb. Perspect. Biol. **5**(10), a010116 (2013)
14. Luo, H., Zhang, C.T., Gao, F.: Ori-Finder 2, an integrated tool to predict replication origins in the archaeal genomes. Front. Microbiol. **5**, 482 (2014)
15. Mirel, B., Wright, Z.: Heuristic evaluations of bioinformatics tools: a development case. In: Jacko, J.A. (ed.) HCI 2009. LNCS, vol. 5610, pp. 329–338. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-02574-7_37
16. Mullany, L.E., Wolff, R.K., Slattery, M.L.: Effectiveness and usability of bioinformatics tools to analyze pathways associated with miRNA expression. Cancer Inform. **14**, 121–130 (2015)
17. Rabani, M., Kertesz, M., Segal, E.: Computational prediction of RNA structural motifs involved in posttranscriptional regulatory processes. Proc. Natl. Acad. Sci. **105**(39), 14885–14890 (2008)