REVIEW on

A Comprehensive Resource for Exploring Antiphage Defense: DefenseFinder Webservice, Wiki and Databases.

doi: https://doi.org/10.1101/2024.01.25.577194

Title and abstract

Does the title clearly reflect the content of the article?

Mostly, Yes. I suggest the following change since the systems do not only target phages, but also other mobile genetic elements.

“Comprehensive Resource for Exploring Prokaryotic Defense Systems: DefenseFinder Webservice, Wiki and Databases.”

Does the abstract present the main findings of the study?

Yes. Slight improvements are suggested in the text below.

Introduction

Are the research questions/hypotheses/predictions clearly presented?

Yes

Does the introduction build on relevant research in the field?

Yes

Materials and methods

Are the methods and analyses sufficiently detailed to allow replication by other researchers?

Yes.

Are the methods and statistical analyses appropriate and well described?

Yes (mostly). See comments below.

Results

In the case of negative results, is there a statistical power analysis (or an adequate Bayesian analysis or equivalence testing)?

I don’t know

Are the results described and interpreted correctly?

Yes. But since this work is introducing a tool, databases and an encyclopedia, only some first (broad) insights are shown.

Discussion

Have the authors appropriately emphasized the strengths and limitations of their study/theory/methods/argument?

Yes.

Are the conclusions adequately supported by the results (without overstating the implications of the findings)?

Yes.

In the here presented work, done by Tesson et al., a website has been created to improve the bioinformatic detection of antiphage systems. This includes an updated version of DefenseFinder with a web service, plus three databases (a wiki on defense systems, a structure database with experimentally determined and AlphaFold2 predicted structures, and a precomputed DefenseFinder results database).

Overall, this is a great work and contribution for the community!
The arsenal and complexity of defense systems are huge and are expanding very fast, making them hard to track. This work facilitates transparency and a fast acquisition of knowledge on the systems by providing a clear overview. The creation of wiki pages was a lot of work and is highly appreciated. The experimental validation part is very nicely presented.

I have only a few comments and minor concerns:

- How are orphans (HMM-only hits) treated or should be treating. These orphans are genes involved in systems but do not make complete ones? This was not addressed in this or the previous article(s). Please add a paragraph what would be a recommend approach to deal with them.
- What about interactions or even synergy between defense systems? And, also, how is the concept of layers of defense systems considered and could be represented?
  These are just thoughts and suggestions that could be added in future versions. Please add some clarifications on these points since these topics are addressed in the community (as reflected in current literature).

Minor concerns

I tested the web service and it works well in my point of view. Is it possible to download figures of the genomic organizations (in good quality) without the need to make screenshots?
Moreover, hits on several contigs detected in a multi-FASTA file are all displayed on 'one genome arrow'. Can this be displayed on separate contigs (which visualization should be limited to 10 or an adjustable number)?

Abstract

- L.19-21: "all known antiphages defense mechanism" is in my point of view (slightly) overstated. There are also other tools that predict the presence of antiviral systems such as PADLOC (PMC9252829), which won't fully overlap with DefenseFinder, indicating that one tool cannot find all systems. I suggest to remove 'all known'.
- L. 26-29 I suggest the following change for a better readability:
  "To overcome these challenges, we present a hub of resources on defense systems, including: 1) an updated version of DefenseFinder with a web-service search function, 2) a community-curated repository of knowledge on the systems, and 3) precomputed databases, which include annotations done on RefSeq genomes and structure predictions generated by AlphaFold."

Methods

- L39-42. Sentence is unclear. Verb is missing (starting from L41).
- Creation of a homepage is quite specific and I do not have the know-how to give any comments this part. The only question that I have is, if the pages will be maintained on the long-term? Is there a funding behind?
- I suggest combining protein selection and structure prediction (L.304-319) together. What does "best hit" mean? Is the selection based on the 30% identity and 70% coverage? This threshold is not very high and may be biased by different structures. In other words, what are the maximum differences between sequences, and does it make sense to take just one sequence as representative?
- Pfam annotation (L.326-333): What is meaning of superimposed in this context? Per protein or domain/sequence position?

Conclusion

- I recommend rephrasing the first section to avoid frequent repetition of 'antiphage.'
- Additionally, I suggest to add a small section describing future plans for updating and improving this platform. This could include plans for exploring interactions between defense systems, conducting docking simulations using the predicted structures, providing tutorials, and even organizing workshops or events to foster collaborations in computational work on defense mechanisms.