

## **Reviewer's report**

**Title:** Automated analysis of ARISA data using ADAPT system

**Version:** 1 **Date:** 20 September 2009

**Reviewer number:** 2

### **Reviewer's report:**

If this software is published in BMC Bioinformatics, this software will likely be used by many colleagues who will not look carefully at how ARISA works and doesn't work. As a result, many analyses of microbial diversity will be highly flawed. Eventually, the community will learn the inaccuracy provided by the program but not before lots of scarce resources are spent and many meaningless papers are published.

The problem is the hypervariability of the ITS region between the 16S and 23S rRNA genes. The hypervariability is not correlated with taxonomy. Thus, a database of known ITS sequences has no predictive value.

In addition, ARISA has very low resolution. In an environmental ARISA community profile, each band can be comprised of ITS sequences from multiple organisms from widely diverse taxa. How can any ITS database be constructed to solve this problem?

Finally, there is considerable error in ARISA band size estimation. For bands as large as 1200 bp, that error can be 2% or more. Again, no software can correct for these errors.

For these three reasons, the software and database described in this paper cannot provide meaningful results to the user. My concern is that unsophisticated users will use this program and the results provided will provide very misleading interpretations.

My answers to the specific questions from BMC Bioinformatics are below each question. .

1. Does the software address a novel task? Alternatively, if there is already software available that performs this task, does the software outperform it in terms of speed, reliability, efficiency, or breadth of application?

This software performs a novel task but it provides very misleading results to the user for the reasons described above.

2. Is it easy to use?

Too easy - this will encourage users to use a program that provides incorrect results.

3. Does it satisfactorily address the task or application the authors intend?

No, it does not allow the user to quickly identify organisms in an environment.

4. Is the software freely available for non-commercial use (note that this is a condition of publication)? And is the availability of the software and any restrictions on use clearly stated in the manuscript?

Yes, web-based and freely available.

5. Does the manuscript clearly describe the problem the software is designed to address?

No, it does not describe the pitfalls that occur from the use of an ARISA database. Such a database ONLY useful if it is derived from a comprehensive set of ITS fragments present in the organisms from the environment being studied. The database described here is derived from fully sequenced genomes and may bear little relation to the community being studied in a given environment.

6. Does the manuscript clearly describe how the software is implemented?

Yes.

7. Does the manuscript clearly describe how the software performs and its advantages / limitations over existing applications?

Yes, I am not aware of any competing applications for ARISA. But there is a reason for this - what is described here will not provide meaningful results.

8. Does the manuscript state the software's operating requirements?

Web based, platform independent.

9. Are the discussion and conclusions of the manuscript well balanced and adequately supported by the data?

No. The database simply will not provide useful results.

10. Do the title and abstract of the manuscript accurately convey what has been found?

Yes.

11. Is the writing acceptable?

Yes.

**Level of interest:** An article of limited interest

**Quality of written English:** Acceptable

**Statistical review:** Yes, and I have assessed the statistics in my report.