20 April 2019

Dr. Ivan Baxter
Editor in Chief
Plant Direct

Dear Dr. Baxter and Members of the Editorial Board:
Attached please find our revised manuscript titled: *Accelerating structure-function mapping using the ViVa webtool to mine natural variation*. We appreciate the thoughtful comments and suggestions provided by you and reviewers. These comments have greatly improved the manuscript and the webtool. We have incorporated the reviewers’ suggestions wherever possible, improving accessibility and adding clarity and value to this manuscript. An MS Word version of the manuscript with tracked changes highlighting our revisions is attached.

A detailed response to the reviewers’ comments is included below.

Thank you again for your time and consideration.

Sincerely,

R. Clay Wright
Assistant Professor
Biological Systems Engineering
Translational Plant Science Program
Virginia Tech
wrightrc@vt.edu
Reviewer #1:

This manuscript describes a new datamining webtool, ViVa, and shows proof-of-concept application by beta testers toward characterizing natural variation within the well-studied auxin signaling pathway. These beta testers came from relatively wide-ranging educational backgrounds, thus further demonstrating the potential practical utility of this tool. Overall, I think the value and effectiveness of such a tool is well argued.

I do want to take the time to comment on the overall structure of this manuscript. The supplemental information (just referenced in the main manuscript a couple of times as a whole, and many of the figures are never referenced) presents analysis of other aspects of the auxin signaling pathway and further analysis of what is discussed in the main manuscript. This section is written almost like a separate manuscript. Overall this data adds to the demonstration of the utility of ViVa, and as a reader I would have preferred that it be better incorporated into the main body of the text.

We appreciate the reviewer’s support of our project. We agree that the supplement is primarily serving to provide additional demonstrations of the utility of ViVa. We structured the paper the way we did out of concern that the lengthy analysis included in the supplement would bog down readers for whom auxin is not their primary interest. To avoid this, we framed the paper as a concise vignette, and then included these additional analyses as a supplement. We agree with the reviewer that more integration was needed between these two components of the manuscript, and have added a paragraph outlining the nuclear auxin signaling pathway (including references to the supplemental figures) in the main text. We have also incorporated our reasoning as to this separation of analyses in the introduction.

Some other relatively minor comments/suggestions for improvement are provided below:
1) Line numbers start over on each page, which made referencing particular lines for review suboptimal.

We apologize for this oversight. In the revised version line numbers are continuous.

2) Figure 2 (and others like it): I am wondering if there is a way to show a consensus sequence or something on the “x-axis” that is not just codon position? Additionally, the boxes residues should be explained in the figure legend. Along these lines, the text mentions charged residues that show little variation in the EAR Domain. The authors should highlight which they are talking about in the figure. Also, the boundaries of the domains indicated EAR, etc.) should be clearly indicated.

We have added a visualization of the degree of consensus in the form of the evolutionary trace of each position in the alignment. We have also added a precise description of the domain boxes to the figure legend.

3) Lines 13-16, page 8: Sentence that begins “By comparing....” is confusing. Consider restating.
We have revised this sentence to make more clear and direct statements about the analysis of this visualization.

4) Please define $\pi_N$, and $\pi_S$ or at least the terminology $\pi_N/\pi_S$ (the nonsynonymous relative to synonymous nucleotide site diversity) when first used. It is stated in line 11, page 8, but not explicitly defined.

$\pi_N$, $\pi_S$ and $\pi_N/\pi_S$ are defined on page 5, in the paragraph starting at line 27. This paragraph has been edited to make the meaning of $\pi_N/\pi_S$ more clear. The sentence referenced in the comment has also been edited to provide additional clarity and accessibility.

5) Figure numbers referred in the text jump from 3 to 8 then 5 then 11. Consider re-numbering or reorganizing this information. This structural issue is discussed more broadly in the first paragraph above.

We have made our best attempt to keep the numbering of the figures in order. However, in keeping with our goal of a concise accessible manuscript presenting a ViVa usage example vignette, we retained some of the same numbering of the supplementary figures. We have also renumbered the supplementary figures as S1, S2… However, we would prefer to keep the Supplemental Analysis as a section of the final paper if possible, just as it is presented in our manuscript.

6) Figure 7 is never referenced explicitly in the text.

We have added a reference to this supplemental figure in the text as described above.

7) Figures 3, 5, 9, and 11: Were all of the same tools used to create the phylogenies? These should be explicitly stated, especially in the first figure of its kind (e. g. was Aliscore also used in figure 3? I assume so from the citation?)

Correct, we have added these references to each legend.

Reviewer #2:

Hamm et al., have generated a webtool, ViVa, that can analyze natural variants of Arabidopsis genes using the Arabidopsis thaliana 1001 genomes project database. Natural variants have aided breeders and evolutionary biologists to understand the importance of variation in the genome towards selection of desired phenotype and functional analysis of genes. Yet, the usage of such large-scale genomic data is empirical without a prior bioinformatic skillset. The authors have generated a webtool with a user-friendly interface for identification of natural variants within gene families along with related information from the repository. In a proof-of-concept, the authors identified variants in genes involved in the auxin signaling pathway and further built an alignment followed by phylogeny showing functional diversity.
Overall, this is a nice tool that will be of use to the community. I have a few comments for improvement of the manuscript and/or the ViVa tool.

Major concerns:
1. The authors have explained in the main results and methods about the different functionality of the web tool like diversity plot, SNP mapping, SNP browser. But the shown example using auxin signaling genes shows the diversity within the gene family but fail to show the robustness of the tool in utilizing the 1001 genomes and also SNP mapping. It would be useful for the authors to show the variant filtering option as this would be crucial for users of ViVa.

We thank the reviewer for the supportive comments about the potential use of the ViVa webtool. A new figure has been added that shows the general workflow of using the webtool as well as specific details on using the SNP Browser and SNP Mapping tabs. Additionally, a supplementary figure has been added showing the use of the SNP mapping tab.

2. In the introduction and abstract, there is information about the use of ViVa in mining data at the gene network level; however, this is now demonstrated.

The current tool is capable of visualizing individual genes and gene families, but as we have done for the auxin signaling network, each gene family comprising a network can be analyzed to make inferences about network variation. In the future we plan to add a network graph visualization. The "Overview of ViVa" sub-section of the results has been edited to say that functionality in ViVa for visualizing diversity at the gene and gene family level. We mention in the conclusion that network level visualizations are planned for future development.

3. This tool is meant to be easy for researchers of any career stage to use (the authors did not mention the level of expertise of researchers employed in using the tool and level of instruction provided by the tool developers). To this end, a more detailed explanation should be in the methods section (see two specific examples below).

We have added a Testing section to the Methods describing in more detail our process of testing the usability of ViVa across wide of expertise. We appreciate the reviewers comment and feel that this section is a great improvement to the manuscript.

a) A flow chart depicting the workflow with input file format, input parameters, output file format and how to interpret results needs to be provided.

We greatly appreciate this helpful suggestion. A new figure with a flow diagram of the general workflow for the webtool, including format of the input gene lists has been added.

b) The input sequence for detailed data analysis requires information "pulled" (line 16 of methods). The authors should provide a detailed stepwise instruction for obtaining this input file.
This procedure was performed in building the tool and is not necessary for the end user. We have rephrased this to indicate that this file was simply downloaded and provided a link. We have also added that the AraPort11 GFF file is included in the package.

Minor comments:
1. Provide citation in line 14-15 in introduction.

Relevant citations supporting the several key points of this sentence have been added.

2. Line 18-20 needs to be explained or removed since the current webtool is designed to employ known functional evidence to determine functional diversity but such information is still lacking in other crops. What is the possible future directions?

We have edited this sentence to be more direct in connecting with the variant effect assays referenced in the previous sentence. We have also added a brief explanation of how ViVa can be used to identify alleles of genes known to affect certain phenotypes. This ability to identify functional variation and genetic variation that may be of use to breeders in developing a desired phenotype is the future direction of ViVa, and we have added language to make that more clear.

3. The authors have provided both webtool based and R package. Is there any difference in terms of data analysis when utilizing the two modes in terms of analysis time, output formats, input parameters. A comparison of the two analysis pipeline should be discussed specifically in this context to obtain a general idea of advantages and disadvantages.

The subsection "ViVa R package: Programmatic access to ViVa’s functionality" briefly states that the R package provides all functionality of the webtool in addition to access to the underlying data structures. We have added a sentence to address that use of the package is meant for advanced R users, who wish to extend the capabilities of the web interface, while the web interface is intended for general use.

4. Line 28-29 in results needs to be rephrased.

We have simplified the language here to more discretely explain our logic to a broad audience.