Meinolf Ottensmann

Von: em.pone.0.539a24.3c3e86d6@editorialmanager.com im Auftrag von PLOS ONE <em@editorialmanager.com>
Gesendet: 02 June 2017 11:18
An: Meinolf Ottensmann
Betreff: PLOS ONE Decision: Revision required [PONE-D-17-07220] - [EMID:7397513e64c01e59]

PONE-D-17-07220
GCalignR: An R package for aligning Gas-Chromatography data
PLOS ONE

Dear Mr. Ottensmann,

I would like to apologize for the time it took to review the manuscript and thank you for submitting your manuscript to PLOS ONE. After careful consideration, we feel that it has merit but does not fully meet PLOS ONE's publication criteria as it currently stands. Therefore, we invite you to submit a revised version of the manuscript that addresses the points raised during the review process.

The reviewers provided a very careful and detailed critique, and brought up important points that need to be addressed. In particular, major concerns are:
How does this method and algorithm differentiates from previously reported alignment algorithms and methods? This is critical, as Plos One has strict criteria for methods manuscripts.
The implementation and success of the package will depend on the evidence and data showing overall improvement over other methods as well as a clear understanding of the use and effects of the different parameters on the alignment.
The methods performance needs to be addressed and compared against benchmark methods.

We would appreciate receiving your revised manuscript by Jul 17 2017 11:59PM. When you are ready to submit your revision, log on to http://pone.edmgr.com/ and select the 'Submissions Needing Revision' folder to locate your manuscript file.

If you would like to make changes to your financial disclosure, please include your updated statement in your cover letter.

To enhance the reproducibility of your results, we recommend that if applicable you deposit your laboratory protocols in protocols.io, where a protocol can be assigned its own identifier (DOI) such that it can be cited independently in the future. For instructions see: http://journals.plos.org/plosone/s/submission-guidelines#loc-laboratory-protocols

Please include the following items when submitting your revised manuscript:

- A rebuttal letter that responds to each point raised by the academic editor and reviewer(s). This letter should be uploaded as separate file and labeled 'Response to Reviewers'.
- A marked-up copy of your manuscript that highlights changes made to the original version. This file should be uploaded as separate file and labeled 'Revised Manuscript with Track Changes'.
- An unmarked version of your revised paper without tracked changes. This file should be uploaded as separate file and labeled 'Manuscript'.

We look forward to receiving your revised manuscript.

Kind regards,

Francisco J. Schopfer, Ph.D. MBA
Academic Editor
PLOS ONE

Journal requirements:

When submitting your revision, we need you to address these additional requirements.

Please ensure that your manuscript meets PLOS ONE's style requirements, including those for file naming. The PLOS ONE style templates can be found at
http://www.journals.plos.org/plosone/s/file?id=wjVg/PLOSOne_formatting_sample_main_body.pdf and
http://www.journals.plos.org/plosone/s/file?id=ba62/PLOSOne_formatting_sample_title_authors_affiliations.pdf

[Note: HTML markup is below. Please do not edit.]

Reviewers' comments:

Reviewer's Responses to Questions

Comments to the Author

1. Is the manuscript technically sound, and do the data support the conclusions?

The manuscript must describe a technically sound piece of scientific research with data that supports the conclusions. Experiments must have been conducted rigorously, with appropriate controls, replication, and sample sizes. The conclusions must be drawn appropriately based on the data presented.

Reviewer #1: No

Reviewer #2: Partly

---

2. Has the statistical analysis been performed appropriately and rigorously?

Reviewer #1: N/A

Reviewer #2: No

---

3. Have the authors made all data underlying the findings in their manuscript fully available?

The PLOS Data policy requires authors to make all data underlying the findings described in their manuscript fully available without restriction, with rare exception (please refer to the Data Availability Statement in the manuscript PDF file). The data should be provided as part of the manuscript or its supporting information, or deposited to a public repository. For example, in addition to summary statistics, the data points behind means, medians and variance measures should be available. If there are restrictions on publicly sharing data—e.g. participant privacy or use of data from a third party—those must be specified.

Reviewer #1: Yes

Reviewer #2: Yes

---

4. Is the manuscript presented in an intelligible fashion and written in standard English?

PLOS ONE does not copyedit accepted manuscripts, so the language in submitted articles must be clear, correct, and unambiguous. Any typographical or grammatical errors should be corrected at revision, so please note any specific errors here.

Reviewer #1: Yes

Reviewer #2: Yes

---

5. Review Comments to the Author

Please use the space provided to explain your answers to the questions above. You may also include additional comments for the author, including concerns about dual publication, research ethics, or publication ethics. (Please upload your review as an attachment if it exceeds 20,000 characters)

Reviewer #1: Ottensmann et al. introduce GCalignR, an R package to align GC data. They justify the necessity of their package since numerous tools exist for GC-MS data alignment, but there is no alternative for GC coupled to other (not-defined) detectors. However, there are indeed packages and tools for GC-MS peak alignment [1-4], and concretely [1,4] an R package called 'ptw'. Adapting those solutions, from GC-MS to GC should be straightforward, since all these align each m/z channel independently, and thus, a TIC (or any other single signal) generated by GC can be aligned as such (as if it was a single m/z channel). Second, methods and results are difficult to understand. From the text, it is not clear how GCalignR aligns peaks (see other comments). Also, the metrics used for the results section (Fig 6, the number of substances) are not indicative of the method performance. My suggestions about what this study needs to reach the novelty and quality required for publication are as follows:

1. Authors use a peak list, but they do not specify how they obtain it. Are all the GC vendors' software allow for automated peak picking? If so authors should state that, so users are aware that the peak list can easily be obtained from their vendors' software. However, a peak picking algorithm could be easily implemented, for instance, with the 'MassSpecWavelet' R package (Bioconductor). This would enable GCalignR to perform the complete GC data processing pipeline without third party vendor software.

2. A robust metric to evaluate the method performance is to analyze the same experiment with GC-MS and GC. This is how authors have tested GCalginR, however, I do not believe that the metrics employed are the most illustrative. The same peak list could be aligned with other GC alignment tools, such as [1]. Then, the performance could be assessed by introducing a random noise, and then evaluate the quantitative performance (R-squared) between GCalignR and [1], given a reference (ground truth) quantitative peak area of a selective m/z for each compound. In other words: provide the R-squared for each molecule, between GCalignR and the reference (selective m/z) peak area across samples; and the same between 'ptw' and the selective m/z.

I believe that the implementation of these suggestions have a low technical complexity, and that they will greatly improve the applicability of GCalignR in GC studies.

Minor:

- Line 42: anonymous should read unknown
- Although analytical methods are described in another paper, a brief courtesy methods paragraph should be included. Authors could name which type of detectors coupled to GC is GCalginR compatible with.
- Section (ii) Peak alignment, this section is confusing and more effort should be devotted to clarify the GCalignR algorithm. For instance, how is the first matrix (line 109) calculated?
- Line 248, 257-258. Units are mandatory.

[1] Ron Wehrens Tom G. Bloemberg Paul H.C. Eilers. Fast parametric time warping of peak lists. Bioinformatics

(2015) 31 (18): 3063-3065.

[2] Perera, V., De Torres Zabala M., Florance, H., Smirno , N., Grant, M., Yang, Z. R. Aligning extracted LC-MS peak lists via density maximization. Metabolomics. 8 (2012) 175–185.

[3] Wei X, Shi X, Merrick M, Willis P, Alonso D, Zhang X. A method of aligning peak lists generated by gas chromatography high-resolution mass spectrometry. Analyst. 138 (2013) 5453-60.

[4] Bloemberg, T. G. et al. (2010) "Improved Parametric Time Warping for Proteomics", Chemometrics and Intelligent Laboratory Systems, 104 (1), 65-74

Reviewer #2: The authors present a technique implemented in R for retention time alignment of GC data. They demonstrate the data on real data to show noise insensitivity.

Major:
Page 2, paragraph starting with line 48:
Algorithms and software are not synonymous. You have opted to describe both in one manuscript, which is fine, but should realize that the description and evaluation of each are very different. What you've done is created a piece of software for alignment (and anyone could substitute your algorithm for their own), and created an algorithm for alignment.

A split treatment would facilitate the reproduction of the algorithm, including for external validation, without being tied to the implementation, and would allow the evaluation of the implementation independent of the alignment algorithm it is used with.

I would like to see you tweak the text to separately discuss them. If you are introducing a novel algorithm, its description should be in the methods, not the introduction. It should also be described in sufficient detail that it can be reproduced from the paper. The current description is vague enough to not only be unimplementable, but also too vague to see how your algorithm differs from the over 100 alignment algorithms already published.

It is very difficult to write English language in a way that carries the mathematical precision of code. I would recommend replacing English descriptions of code with pseudocode or a flowchart. These descriptions should definitely not be in Figure legends (Fig 2), or the introduction.

Your method has a dozen user-set parameters. In your evaluation against a ground truth benchmark, you explored 100 combinations of parameters. What strategy should users use to find optimal settings for 12 parameters on real (non-ground truth) data, where supervised scoring is not available? This is a very significant issue.

On a related but separate note, I would like to see a figure showing the results of those 100 combinations to give the user a sense of what random parameter settings (which may be the best they can do with real data) will yield. In a sense, this would be the most valuable result in the paper, as it shows what a user can expect in terms of real world performance.

This paper is well written, but the content is written from the perspective that there are no other alignment methods published. In reality, there are dozens. This paper cannot be published without a comparison to existing methods. While it is too much to ask for a comparison to the 100+ papers that have been published, the following is a bare-minimum to show that the algorithm is superior to what is already available:
-A quantitative comparison of results for 2-3 top (either quantitatively best or most popular) alignment methods.
-An algorithmic comparison to these other methods (how is what you are doing different from what they did).

An extensive review on alignment was published in 2015 by Smith et. al: "LC-MS alignment in theory and practice: a comprehensive algorithmic review." The authors will find a comparison of 50 alignment algorithmic approaches, and an analysis of continuing limitations of all published algorithms. Notable among the latter: Model assumptions that fail to capture real behavior, long run times due to algorithm complexity and user-defined parameter optimization, pairwise comparisons and reference samples, current methods have not been empirically compared. The authors will note that their method as described in this manuscript is subject to all of those shortcomings.

If your algorithm is as non-novel as your current description suggests, the paper will have to be re-written with the contribution being an R program for facilitating an alignment algorithm, rather than an R program with a novel alignment algorithm. The focus and content would need to be revised to highlight how users can implement the algorithm of their choice, how they can use the program's visualization and other inspection tools to tune algorithm parameters, etc.

Minor:
1. What data formats are accepted by the application? Are they standard? Does any other pipeline use these data formats? Is there a mapping between standard data mappings and what your program accepts/outputs?
2. Page 2, line 64. There are literally dozens of open source alignment programs available. This is not an advantage over what is currently available.

---

6. If you would like your identity to be revealed to the authors, please include your name here (optional).

Your name and review will not be published with the manuscript.

Reviewer #1: (No Response)

Reviewer #2: (No Response)

[NOTE: If reviewer comments were submitted as an attachment file, they will be attached to this email and accessible via the submission site. Please log into your account, locate the manuscript record, and check for the action link "View Attachments". If this link does not appear, there are no attachment files to be viewed.]

Need assistance with your figure files?

While revising your submission, we encourage you to use PACE (the Preflight* Analysis and Conversion Engine, http://pace.apexcovantage.com/), a digital diagnostic and conversion tool for figure files. PACE helps users ensure that their figures meet PLOS requirements and that the quality of published figures will be as high as possible. To use PACE, you must first register as a user. Then, login and navigate to the UPLOAD tab, where you will find detailed instructions on how to use the tool. If you encounter any issues or have any questions when using PACE, please email us at figures@plos.org.