**Response to Reviewers:**

First, we would like to first thank the reviewers for the *considerable* amount of time they clearly put into the review of our manuscript. We are truly appreciative of the effort, and agree with a significant majority of their comments.

Second, we want to immediately address one issue raised by two reviewer: their inability to reproduce our claimed runtime. This is easy to explain: in the rush to get our manuscript out the door for the NAR deadline, we made last minute change to the code that introduced a bug with a significant impact on runtime. We realized this when we were able to reproduce the reviewer’s results with the submitted version, and our own results when we backtracked to an earlier version. The problem has been fixed, all results have been re-generated to reflect the current version, and the reviewers should be able to reproduce them. We very much appreciate the review’s leeway on this (and their willingness to further consider the manuscript).

All other points will be addressed on a case-by-case basis as follows:

**Reviewer #1**:

|  |
| --- |
| **Comment:** I suggest to authors the use of MetaSim tool. |
| **Respose:** We were not aware of this tool, and appreciate the pointer. We have used MetaSim for an extensive set of tests, adding the relevant to discussion to both the manuscript and supplementary discussion, as appropriate. |
| **Manuscript changes:** |
| **Status**: Unfinished |

|  |
| --- |
| **Comment:** During the simulation, authors should take all execution times in order to be compared with those from other tools. |
| **Response**: |
| **Manuscript changes**: |
| **Status**: Done |

|  |
| --- |
| **Comment:** I tested PEACE on a human dataset downloaded from the Easycluster web page, a program cited in the manuscript. Such database contains 111 well-defined clusters. Using PEACE with default parameters I was able to recover only 67 clusters. It suggests that type 2 error may be high. For this reason I think that authors should take into account also type 2 error other than the type 1 and sensitivity. A summary table may be useful in the main manuscript. |
| **Response**: |
|  |
| **Status**: Unfinished |

|  |
| --- |
| **Comment**: During the testing I found that PEACE took more than 20 min. to process about 20,000 human ESTs. On the same machine, wcd took less than 5 min. However, improved performances were registered running PEACE on a cluster. |
| **Response**: This would be a direct effect of the bug introduced before submission, as discussed at the beginning of this document. We fully acknowledge that any clustering tool requiring 20 minutes to cluster 20K ESTs would be worse than useless, and again thank the reviewer for not dismissing us out-of-hand. The bug is fixed, and we can now… |
| **Manuscript changes:** |
| **Status**: Unfinished |

|  |
| --- |
| **Comment**: Regarding real data I have several doubts about their reliability. In case of mouse dataset I directly found incorrect relationships between genes and ESTs. This could limit the evaluation step. |
| **Response**: We are not quite sure how to interpret this statement. Most major bioinformatics tools have a heuristic aspect: BLAST, GenScan, Hummer and almost all other major tools have some sacrifice of result quality in favor of speed: the size of the data sets make this unavoidable. The clustering problem is no different: it is impossible to perform a large clustering job in a reasonable amount of time, requiring that WCD, Cap3 and very other tool accept some errors in order to achieve a usable execution time. We make no secret of the fact that there will be some errors in the clustering, but provide evidence that our solution quality is at least on part with other tools.  Without knowing more specifics of the reviewers complaints, we are unable to further address this issue. Nor are we clear what the reviewer means in his concern that this “could limit the evaluation step”. |
| **Manuscript changes**: None |
| **Status**: DONE |

|  |
| --- |
| **Comment**: In order to reliably assess the performance of PEACE on real data, I suggest the use of ad hoc benchmarks in which the gene-to-est relationship is well known. Genome browsers such as UCSC could be very useful. |
|  |
| **Manuscript changes:** |
| **Status**: Unfinished |

|  |
| --- |
| **Comment:** Using PEACE I found few errors in the GUI interface:   * in the “choose workspace” window replace “This is directory is called…” with “This directory is called…” * in the alert of the “clustering Setup” window replace “ESTs more making more clusters” with “ESTs making more clusters” (if I correctly understood the meaning of the sentence) |
| **Response**: We appreciate the observation, and have fixed these problems. |
| **Manuscript changes:** No change to manuscript, but the GUI has been changed as requested. |
| **Status**: Needs Varification |

|  |
| --- |
| **Comment**: About the output, it should be useful to get results in a standard format like GFF in order to be processed by other tools. |
| **Response**: We assume the reviewer is referring to *General Feature Format* (define at <http://www.sanger.ac.uk/resources/software/gff/>). This format is for tracking sequence annotations. We have no sequences to annotate; our output is a cluster of ESTs which correspond to an unknown transcript. Without having the reference transcript on which we could describe the features (presumably the GFFs), we are not sure how to make use of this format. |
| **Manuscript changes**: None |
| **Status**: DONE |

|  |
| --- |
| **Comment**: PEACE has been developed for clustering and assembly but running the program I was not able to assemble ESTs for each detected cluster. |
| **Comment**: As is stated in the manuscript (see ???), PEACE is a clustering tool – it does not intended to perform assembly and has no capacity to do so. We merely claim that it can pick out those ESTs associated with a given cluster in preparation for assembly. The expectation is that the user will the employ the assembly engine of their choice to each cluster, taking advantage of the smaller problem set sizes and improved clustering quality to get an overall result in a smaller amount of time. |
| **Manuscript changes**: None |
| **Status**: DONE |

|  |
| --- |
| **Comment**: Here quality scores could be introduced and PEACE should be able to read SFF files generated by 454 reads. |
| **Response**: This was a good idea, and we have incorporated it into the tool. |
| **Manuscript changes**: |
| **Status**: Unfinished |

**Reviewer II**

|  |
| --- |
| **Comment #2**: I did have an installation problem on the ﬁrst server I tried, not sure why – pretty standard Linux Ubuntu Karmic Koala trying both the packed openmpi available and my own manually installed. I attach a log for feedback |
| **Response**: We appreciate the forwarding of the log. The information allowed us to isolate and fix the problem. |
| **Manuscript changes**: No change to manuscript, but the instillation process has been modified to eliminate this bug. |
| **Status**: Unfinished |

|  |
| --- |
| **Comment #6**: Your new data sets should be publicly available for experimentation. |
| **Response**: In the original submission, or data sets were a combination of those available by your groups and certain sets currently under study in the lab of Dr. Liang. However, analysis of Dr. Liangs sets required the use of the **gmap** tool, which was objected to by Reviewer #1. So we have instead removed them, and instead are using benchmark sets obtained from both your site and the EasyCluster. We are happy to put these sets on our sites, but it seems unnecessary – and we would prefer not to without specific permission from the sources of the data. For the moment we will provide pointers to both sites.  The original chlamy dataset was available on our site (though we have since removed it). If needed we can make our own Arabidopsis data available, but it is no longer relevant to the paper. |
| **Manuscript changes**: |
| **Status**: Unfinished |

|  |
| --- |
| **Comment #8**: I think the abstract’s claim on performance are cherry  picking. Obviously you have to compress in the abstract but you still need to be fair in assessment.  While I do think the number of experiments are sufficient here, there is a lot of subtlety in both performance and quality assessment. You seem to pick in your comparison with wcd (and CAP3) the best results you got for your tool. With respect to the quality results you pick one of the simulated parameters which gives you very good results. On real data you report, the quality (sensitivity) difference is not big (0.958 vs 0.94, 0.936 and 0.932 – with  wcd having better Jaccard indices). I don’t think that this justiﬁes as a headline result a 52% improvement in quality. |
| **Response**: Upon reflection, we find this to be a very justified comment, and we have modified the manuscripts accordingly. To be clear: it is our position that PEACE is on-par with WCD, but not necessarily superior. Each tools has does better than the other under certain metrics (though I suspect this is largely accounted for by difference in parameter settings), and we feel the work is worth publishing in this NAR issue large because of the GUI that (to our knowledge) WCD cannot currently match. We hope this is now clearer. |
| **Manuscript changes**: |
| **Status**: Unfinished |

|  |
| --- |
| **Comment** **#9**: With respect to performance, in the last paragraph of the abstract you claim a 40% improvement in performance on a 361MB (you say Mb, I assume MB) Arabidopsis data set. I presume this is our Aful l or the A686 data set? I cannot see this claim substantiated in either the paper or the supplementary material (I just don’t see the experiment reported – perhaps I am being dense here but I did read the paper clearly and have been through it again). |
| **Response**: First: no, the Arbidopsis was not a WCD set (see response to comment 6), but based on the comments of Reviewer #1 we have stopped using it anyway. |
|  |
|  |

|  |
| --- |
| **Comment #10**: On p7 you talk about Figure C.4. I presume you mean S.4. |
|  |
|  |
| **Status**: Unfinished |

|  |
| --- |
| **Comment #11**: In the captions of Figures S4 and S5, please make clear what data sets are being used. |
|  |
|  |
| **Status**: Unfinished |

|  |
| --- |
| **Comment** **#12**: In Table S2 (and where you reference it), you call the data here Mouse data – in fact it is a subset of the Arabidopsis data. Could you just double check all names of data sets? |
| **Response**: (blush) Thank you. We at some point confused the WCD Mouse and Arabidopsis sets. All sets have been doube-checked. |
| **Manuscript comments**: Fixed as appropriate. |
| **Status**: Unfinished |

|  |
| --- |
| **Comment** **#13**. The documentation is clear and well written. There is a glitch in that the page numbers in the table of contents are not correct. (e.g., the non-graphical mode is claimed to be at page 24 – it’s at page 37). Also, in the CLI discussion, you call the program PEACE, but the make seems to create peace |
| **Comment**: Thank you for catching this.. |
| **Manuscript changes**: No change to the manuscript, but the documentation has been fixed. |
| **Status**: Needs Varification |

|  |
| --- |
| **Comment #14**: In all cases when you report performance results, please be clear on which computer you ran the data. Also, saying 3.2GHz Intel Xeon is not a well deﬁned concept – there are many machines with this label varying dramatically in technology, L2 cache size and bus speed, which can have a profound effect on time taken. Please quote model number. |
| **Response**: Each processor was a 3 GHz Intel Xeon EM63T CPU with a 2MB cache and an 800 MHz front side bus (model number Xeon LV 3.0, released 2005). We have added this information in both the paper and supplementary materials with the discussion of runtime. |
| **Manuscript changes**: The information has been added to both the manuscript (modifying the previously provided description) and the supplementary materials (in the section on runtime). |
| **Status**: Unfinished |

|  |
| --- |
| **Comment #15**: You don’t talk about memory consumption. My experiments seem to show that the consumption is modest compared to some other tools which is very important since some other tools fail on this. If your program does have modest memory requirements, it’s a very good thing. Either way you should say so explicitly. |
| **Response**: |
|  |
| **Status**: Unfinished |

|  |
| --- |
| **Comment** **#16**. However: I could not run our A686904 data set – either from the command line or the GUI. |
| **Response**: |
|  |
| **Status**: Unfinished |

|  |
| --- |
| **Comment**: 17. Please state version of wcd you used for your testing – in what I cite below, I used wcd 0.5.1 (May 2009). |
| **Response**: The previous version was using 0.4.5; for the resubmission we have updated to 0.5.1. |
| **Manuscript changes**: We have noted this information in both the manuscript and supplementary materials. |
| **Status**: DONE |

|  |
| --- |
| **Comment #18**: With respect to the MPI implementation, when the system is installed from the GUI I cannot see how the list of nodes where the jobs should run is speciﬁed. Are you assuming that it is already listed in the MPI set up as a default ﬁle? This may be a reasonable assumption but perhaps it could be documented. |
| **Response**: |
|  |
| **Status**: Unfinished |

|  |
| --- |
| **Comment** **#19**: Most seriously, I have not been able to reproduce your performance claims either from the command line or from the GUI. |
| **Response**: This is a direct result of the bug we introduced right before review. We have fixed the bug, posted the newest version, and rerun *all* results using that version. We expect that you will now find runtimes consistent with our claims. |
|  |
| **Status**: Unfinished |

|  |
| --- |
| **Comment**:  20. The points that you make about short read sequences is important. With more pro jects using  this new data the ability to cope with such data is important. I can accept because of the  way in which wcd’s heuristics are implemented that PEACE will do better with short read  sequences. However, we have used wcd very successfully with 454 data. Recent results of one  of my students showed that provided the average sequence length was over 100, reasonable  results will be obtained. wcd’s parameters can be changed so window lengths of less than 100  can run (though as the length drops to 50, the hard coded non-parameterisable heuristics  mean the quality suffers – but it actually can run even with Solexa data – just doesn’t do  well at all). I am surprised that you couldn’t get wcd to run on the data.  So: I think for your experiment you should characterise more carefully your data set (average  length, proportion under say 80 in length). This could be a big selling point of the tool and  so a little more discussion on this would be worth it. For example, with some short read  sequencing technology now producing 60bp length sequences, would PEACE be suitable (I  doubt wcd would be). |
| **Response**: |
|  |
| **Status**: Unfinished |

**Reviewer #3**

|  |
| --- |
| **Comment**: 1. The tool took a long time to install after hacking the installation commands and the GUI simply would not run. We tried this on several UNIX systems, running different flavors of UNIX - all with Java as required. |
| **Response**: We need more information to address this. We have not encountered any such problems, nor have any of our testers. And we cannot fathom why the GUI would not run on any machine supporting Java. Without further information, there is nothing we can do. |
| **Manuscript changes**: None |
| **Status**: DONE |

|  |
| --- |
| **Comment**:  2. The manuscript looks good: h/e there is no comparison with any similar clustering method to permit a reasonable evaluation:  1.SOAP (http://soap.genomics.org.cn/), which assembled excellent short reads from novel organisms without genome data;  2. TGICL: the well established clustering tool.  The manuscript would be much improved if such comparisons can be provided. |
| **Response**: |
|  |
|  |