**Response to Reviewers:**

First, we would like to thank the editor and the reviewers for the considerable amount of time they have clearly invested into the review of our manuscript. We are truly appreciative of the effort and we have strived to address all of the comments.

We would like to immediately address one concern raised by two of the reviewers regarding increased run times of our software. This is easy to explain: in an endeavor to address technical difficulty reported by one of the reviewers (via the editor), we decided to upload our current working version. Unfortunately, we missed noting that this version was not yet performance-tuned – hence the reviewers were working with a default parameter configuration that was considerably less than optimal. The problem has been fixed, all results have been re-generated to reflect the current version, and the reviewers should be able to reproduce our results. We apologize for this oversight. We very much appreciate the opportunity to resolve this issue and the willingness of the reviewers' 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. |
| **Response:** 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.  We continued to use ESTSim for simulating Sanger Reads for consistency with the WCD analysis in the Hazelhurst 2008 paper. |
| **Manuscript changes:** Added analysis of tools as applied to MetaSim output for 454 and Illumina technology simulations (results section, last paragraph). |
| **Status**: DONE |

|  |
| --- |
| **Comment:** During the simulation, authors should take all execution times in order to be compared with those from other tools. |
| **Response**: The runtimes for PEACE and Cap3 are reported in Figures 4 and Tables 1 of the manuscript for purposes of comparisons. We do not report Cap3 runtimes because we believe any comparison is inappropriate: while we use Cap3 as a clustering tool in this study, it is in fact performing assembly as well. Since the two stages are integrated in Cap3, runtimes cannot be separated – hence the runtime of Cap3 will naturally be significantly greater than that of the other tools. Therefore we believe any run time comparison with Cap3 would be both inherently unfair to Cap3 and essentially meaningless. |
| **Manuscript changes**: Add a comparative plot of runtime (Table 4) and runtime results for benchmarks sets (Table 1). |
| **Status:** Unfinished |

|  |
| --- |
| **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**: We have responded by both fixing the specific problem and adding the requested analysis. The reduced quality for this benchmark was due to the value used for our threshold parameters; redone *all* analysis with a new threshold value, and present those results. As shown in Table 1 of the manuscript, we now do quite well on this benchmark – with almost identical results to WCD.  More generally, we agreed with the reviewer about the need for looking at type 2 error. We have added plots of both the Jaccard Index (which reflects type 2 error) and of cluster-level type 2 error to the manuscript, as well as presented these results in the analysis of the real data. |
| **Manuscript changes**: Figure 3 and Table 1 now reflect Type 2 errors. |
| **Status**: DONE |

|  |
| --- |
| **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 performance-tuning introduced before submission, as discussed at the beginning of this document. The bug is fixed, and we find in simulations that our runtime is almost identical to WCD for sequential runs, and somewhat better in multiprocessor runs (see Figure 4 in the main manuscript).  When applied to real data, results are somewhat more mixed, as shown in Table 1 of the main manuscript.  All results should now be re-producible. |
| **Manuscript changes:** Fixed bug in software; added Figured 4 and Table 1 to the manuscript. |
| **Status**: Finished |

|  |
| --- |
| **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**: Similar to most major bioinformatics tools (such as: BLAST, GenScan, Hummer), including WCD and Cap3, we too have a heuristic aspect: we tradeoff some quality in favor of speed because the size of the data sets make such a tradeoff unavoidable. Without heuristics it would be practically impossible to perform a large clustering job in a reasonable amount of time. Consequently, similar to WCD, Cap3 and other tools, we 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 par with the best cluster tools in the literature. |
| **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. |
| **Response**: Given the reviewer’s apprehension concerning gmap results, we have dropped the data sets and instead adopted a number of common benchmarks from the WCD and EasyCluster studies. Specifically, the human Benchmark dataset and the A076941 dataset [provide the requested mapping. |
| **Manuscript changes:** Removal of the analysis of the chlamy and arabidopysis data; addition of the analysis of two benchmarks with known solutions (see Table 1 of manuscript) and several other benchmarks for runtime analysis (see Table S1 in supplementary materials). |
| **Status**: DONE |

|  |
| --- |
| **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**: DONE |

|  |
| --- |
| **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**: I believe this comment reflects confusion addressed in the next comment. 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**: currently PEACE is a clustering tool – it does not perform assembly and currently has no capacity to do so. The “assembly” in the acronym refers to the computational environment we have developed, which is configured to support an assembly algorithm that will be introduced in a future study. We presume that this is causing the reviewer’s confusion, but would prefer not to change the title and have tried to clearly state in the first sentence of the abstract that this is (at the moment) just a clustering tool. |
| **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**: No changes to the manuscript, but it is now on option in the GUI. |
| **Status**: DONE |

**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 software has been modified to eliminate this bug. |
| **Status**: DONE |

|  |
| --- |
| **Comment #6**: Your new data sets should be publicly available for experimentation. |
| **Response**: In response to comments from Reviewer #1, we have replaced our datasets with standard benchmark sets used in both the WCD and EasyCluster analysis. These are publically available sets, and we have added links to them on the <http://peace-tools.org/> website. |
| **Manuscript changes**: Have added appropriate links to the PEACE website. |
| Status: DONE |

|  |
| --- |
| **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 reflect, we agree. This has become even more true with our change in the use of threshold, which has rendered are our results almost identical to yours. (In short: with the old threshold, we have some significant improvements over WCD in simulations in both Se and JI, but not a significantly worse JI on benchmark data. We could not justify this, and had to revert to a lower threshold that makes our results close to identical.)  We have changed the abstract and results section to emphasize the attributes of our tool that justify a place in this journal. Specifically, we feel that we match WCD in quality on Sanger Sequences, that we do have (justifiably claimed) significant improvement in quality on short-read data, and a very user-friendly GUI that enables easy instillation and use of the parallel computations. |
| **Manuscript changes**: Modifications to the abstract and results sections. |
| **Status**: DONE |

|  |
| --- |
| **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**: 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. We have replaced with benchmarks from both EasyCluster and your own cite, as discussed in the results section (see Table 1) and the supplementary materials. |
| **Manuscript changes**: Modifications to the results section and the addition of Table 1. |
| **Status**: DONE |

|  |
| --- |
| **Comment #10**: On p7 you talk about Figure C.4. I presume you mean S.4. |
| **Response**: Correct, though in the revisions this has been removed. |
| **Manuscript changes**: Reference removed in the course of other changes. |
| **Status**: DONE |

|  |
| --- |
| **Comment #11**: In the captions of Figures S4 and S5, please make clear what data sets are being used. |
| **Response**: Done |
| **Manuscript changes**: Information added to relevant figures. |
| **Status**: DONE |

|  |
| --- |
| **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**: We at some point confused the WCD Mouse and Arabidopsis sets. All sets have been double-checked. |
| **Manuscript comments**: Fixed as appropriate. |
| **Status**: Done |

|  |
| --- |
| **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**: DONE |

|  |
| --- |
| **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 EM64T CPU with a 2MB cache and an 800 MHz front side bus (model number Xeon LV 3.0, released 2005); this has been noted. |
| **Manuscript changes**: The information has been added to both the manuscript (in the caption for Figure 5) and the supplementary materials (in the section on runtime). |
| **Status**: DONE |

|  |
| --- |
| **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**: The A686904 data set has a very short EST that caused a sanity check assertion to trip in our code. We have now fixed this problem and our tools successfully processes the A686904 data set. |
| **Manuscript changes**: No changes to the manuscript, but the code has been fixed. |
| **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**: We have added references to configuring MPICH (via mpd) and open MPI (via torque) to utilize an appropriate set of hosts (via hosts file) our user manual. We have also added a brief message directing the user to the manual for details in the GUI. |
| **Manuscript changes**: None to manuscript, but changes have been made in the tool, GUI and manual. |
| **Status**: DONE |

|  |
| --- |
| **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 performance-tuning issue we discussed earlier on. We have fixed the bug, posted the newest version, and rerun *all* results using that version. We expect that the runtimes will be consistent with our current claims. |
| **Manuscript changes**: All results have been re-run and figures redrawn to be consistent with the latest version of the tool |
| **Status**: DONE |

|  |
| --- |
| **Comment #20**: The points that you make about short read sequences is important. With more projects 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**: Using the MetaSim simulation tool we did some experiments on simulated 454 and Illumina data – verifying that we can handle the shorter sequences. (Note the added discussion of our “adaptive d2” strategy in the supplementary materials.) In regards to our comparison to WCD, we ran into a problem. We could run your tool on the output data, but as far as we could tell, the window size was hard coded; it was not clear how to change this. Its not clear from above whether your student is changing the parameters for 454, but based on the our testing we are not seeing these results when using the default WCD parameters (v. 0.5.1). We have gone ahead and posted a sample simulation output to the peace page for your use if you want to verify this.  So we have, to the best of our ability, tested our tool against WCD – but have acknowledged in the paper that this may be an unfair test. |
| **Manuscript changes**: Addition of analysis of results on 454 and Illumina data, and the requested characterization, in the results section of the manuscripts. |
| **Status**: DONE |

**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 are surprised and concerned to note that a pure Java program would not run on a machine supporting Java version 6. We (and several collaborators) have tried our software on a broad range of computing platforms and have not encountered such a problem. We cannot fathom why the GUI would not run on a machine supporting Java version 6. We would need additional information to address this problem. |
| **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**: A significant portion of our paper discussed comparisons between PEACE, the WCD clustering tool, and the Cap3 assembly tool. Previously some of these comparisons were regulated to the supplementary materials, and we have shifted them into the main manuscript. That document now contains a results comparison between the three tools on simulated data (Figure 3), a runtime comparison on simulated data (Figure 4), comparisons on benchmark sets (Table 1), and several figures in the supplementary materials.  While SOAP and TGICL are good tools, we are constrained by the length limit and do not think adding them in would improve the quality of our data. SOAP is primarily an assembly tool, and TGICL essentially does its clustering through MEGABlast (using Cap3 to complete the assembly). Given the space limitations, we feel that it makes more sense to compare against WCD (a tool specifically for clustering) and Cap3 (a tool commonly used for clustering)—and that such comparisons provide a reasonable evaluation of PEACE. |
| **Manuscript Changes**: Shifted come results from the supplementary materials to the manuscript (e.g. parts of Figure 3, all of Figure 4). |
| **Status**: Finished |