DETAILED POINT BY POINT RESPONSES

Reviewer 1

-- Ref 1.1 –Comparison to single species clustering --

Reviewer Comment	But it would seem a simpler comparison of the individual species runs with your method or simple method like pam TO your multi-species results should be the first thing to try. The correct control is a comparison to simpler single species methods.
Author Response	This is a great suggestion. In the revised manuscript, we added a new sect "Comparison with single-species clustering" with a new main Figure 6. Specifically, we applied the classical clustering algorithms, including K-m hierarchical clustering and PAM, to cluster the expression profiles of wor and fly separately, resulting at two sets of co-expression modules. We calculated the enrichment of orthologous gene pairs for every worm-fly module pairs, and compared with the enrichment of orthologs for cross-species modules in OrthoClust (Figure 6). We found that even thoug there are certain combined worm-fly modules with significant enrichment orthologous gene pairs, the enrichment is lower than the cross-species constructed by OrthoClust. This is not surprising because the clustering depends merely on the co-expression connectivity in individual species, it less effective in finding the corresponding sets of genes in two species responsible for the same function. The observation points exactly to the essence of OrthoClust - to incorporate evolutionary information across species.

-- Ref 1.2 – Alternative benchmark--

Reviewer Comment	The benchmark chosen is not the best for these organisms. There is a huge quantity of regulatory data, instead of benchmarking with GO shared/enriched function benchmark by seeing how many known co-regulation pairs you recover by each method or parameter choice.
Author Response	Thanks for the great suggestion. In the revised manuscript, we added a ne section "Benchmarking modules based on co-regulation patterns" with a r Figure S4. More specifically, we used worm and fly regulatory networks constructed by using modENCODE ChIP-Seq data as an alternative benchmark. There are 26 TFs in fly and 79 TFs in worm with ChIP-Seq data across various conditions. The networks are thus the superposition of

	the TF-target interactions in worm and fly respectively. Though the sets o TFs are rather small, the experiments were performed in conditions simila the expression data. This presents an advantage over regulatory data collected in literature. As the reviewer suggested, we studied the likelihoo having a common TF for two genes in a module as compared to two gene different modules. We found that pairs of genes within a module, in avera have higher number of common TFs than pairs of genes in different modu In summary, there is a 2.6 fold increase in worm (1.6 in fly), meaning tha genes in a module are indeed more likely to be co-regulated by a commor TF. Nevertheless, using regulatory data as a benchmark cannot allow us to directly compare the performance of our cross-species modules. A gene ir worm and a gene in fly will not share common transcription factor. Thoug one could attempt to look at the orthology between worm TFs and fly TFs is not very feasible as the number of TFs in two networks are rather small We therefore keep the original GO comparison in the revised manuscript provide the co-regulation pairs as an alternative.

-- Ref 1.3 – Language --

Reviewer Comment	The language is not tailored to the audience. The term "spin system" does not resonate with anyone not trained as a physicist or physical chemist. Thus the analogy to spin systems is a major distraction and makes the paper less clear to many (including Bio, CS, comp bio and bio-stats people: including 99% of readership). There are other methods that optimize similar cost functions using MCMC or simulated annealing that describe what they do much better. Cut the spin system stuff and spend more time describing what the method does in a step-wise/procedural active voice. Work on improving the clarity of the methods description.
Author Response	We see the reviewer's point. We have revised the language in a way targe to Bio, CS audiences. While the reviewer suggested that there are other methods that optimize similar cost functions using MCMC or simulated annealing, we would like to point out that the essence of our method is exactly an optimization procedure based on MCMC / simulated annealing The spin system description is merely analogy. In the revised manuscript, explain the method using standard terms in optimization, and only mentio one sentence the spin analogy.

-- Ref 1.4 – Scalability--

Reviewer Comment	No mention of scalability.
Author Response	Though for each iteration, the MCMC updating algorithm scales linearly the size of the system, depending on the energy landscape, there is no sim theory to address the convergence time of simulated annealing. Practically could be quite slow. We agree that this is a downside of simulated anneali Nevertheless, the conceptual framework of OrthoClust is an optimization process and simulated annealing is one of the many optimization approac While simulated annealing may have scalability issue, it is easily to imple and in principle capable of getting the optimal solution. These pros and co were addressed in the discussion section.

-- Ref 1.5 – Other data-types--

Reviewer Comment	What about other data-types? Why compare to GO or P-P interactions, why not use them to get the right answer? My opinion is that clustering just expression data is useful.
Author Response	As the reviewer probably have noticed, OrthoClust can be easily applied t other networks like protein-protein interaction networks. We have drastically revised the manuscript to emphasize the generality of OrthoClust to cluster high-dimensional genomics data or other co-associa networks across species (not restricted to expression network). Nevertheless, the current study focuses on generalizing the traditional clustering method, which is usually unsupervised, to a cross-species fashi Though a supervised learning scheme based on GO or P-P interactions sounds like a fruitful avenue, it is beyond our scope to incorporate in the study. However, we discussed the possible generalization in the discussio section.

Reviewer 2

-- Ref 2.1 – Comparison with Network alignment--

Reviewer Comment	There are other methods that have been used to identify conserved modules by integrating data from multiple different networks, such as IsoRankN and RankProp. As the authors point out, these methods perform network alignment

	rather than potentially cross-species clustering. However, a network alignment can be easily interpreted in terms of conserved clusters, and there are a few papers at least that use this to analyze a co-expression network together with gene homology between species (e.g. Ficklin & Feltus, Plant Physiol 156:1244, 2011). The authors should compare their results in detail to results from network alignment. How do their results differ and why?
Author Response	We thank the reviewer for raising this concern. We added a new results section "Comparison with network alignment" in the revised manuscript. More specifically, we applied IsoRank to align the fly and worm co-expression networks using the same procedure as described by the original paper (Singh et al. PNAS 2008). We repeated the steps for differe alpha (an intrinsic tuning parameter defined by isorank), arriving at sets o worm-fly gene-pairs that significantly aligned (P-value cutoff based on IsoRank default setting). A worm-fly pair means the two genes are simila sequence and exhibit similar topological features in the corresponding networks. Motivated by our original analysis that made use of a set of metagenes (Stuart et al. 2003) as gold standard, we compared the fraction metagenes recovered by OrthoClust versus IsoRank. We found in our original analysis the fraction recovered is 81% for OrthoClust, but 88% fo Isorank (see Figure S6). In this sense, OrthoClust's performance is very close to IsoRank in identifying the corresponding functional genes betwee two species. Nevertheless, network alignment does not immediately report how genes form clusters. As outlined by the reviewer, we used the IsoRank pairs to construct conserved modules and compared the modules with the conserv modules constructed by IsoRank. First of all, we connected the IsoRank pairs to form seed modules. By connecting such edges in the network, we generated aligned subgraphs that could potentially be interpreted as modu conserved across two species. Using the suggested value 0.5 for alpha, we found that the seed modules agree pretty well with the OrthoClust module Among the gene pairs whose components are predicted to be in the same seed module, 43% of them also fall on the same OrthoClust module. We want to emphasis that the aligned subgraphs are not exactly the conserved modules in OrthoClust, because conserved modules in OrthoCl automatically consists not only the orthologs but related genes in two spec In general, it is not trivial to construct conserved modules based on the aligned subgraphs. We did so by following procedures similar to Ficklin & Feltus 2011, but found that the resultant modules do not agree very well w OrthoClust. As the agreement is reasonable for seed modules (aligned subgraphs), but worse by including extra genes, we think the last step is th heart of the differences. While the procedures used by Ficklin & Feltus 20 can be extended or modified in various ways, we believe that a novelty of

	OrthoClust is to approach the same problem in a more straightforward manner (no need to perform alignment and then additional procedures). Furthermore, unlike Ficklin & Feltus 2011 that focused on merely conserv modules, OrthoClust goes one step further by generating specific module and conserved modules in an integrated algorithm. We believe this is also novel aspect of OrthoClust compared to algorithms suggested by the reviewer. In summary, we think the comparison with network alignment i really important. We have shown additional analysis and elaborated the discussion further in the revised manuscript.

-- Ref 2.2 – Exploration of the parameter space (1)--

Reviewer Comment	The number of edges d from each gene in the co-expression network. Also, what happens if they use a correlation coefficient cutoff instead of a fixed number for each gene? The explanation of their choice of d on p. 17 isn't clear.
Author Response	Thanks for the comments. First of all, the explanation on the choice of d i the method section was revised to make the point clear. We understand that there are various ways in literature to construct a unweighted co-expression network from expression profiles. By calculati the gene-by gene Pearson correlation matrix, there are essentially two cla of algorithms: to impose a global threshold on the values of the correlatio coefficients for all genes, or to locally allow each gene to connect to the t most correlated genes (rank-based). The pros and cons between a rank-based approach and a correlation-based approach were studied in Re 19. We tried both ways, and found that the network based on a fixed cuto has only a couple of very big modules (see Supplementary further analysi It suggests that the network is too densely connected, and therefore modu cannot be easily resolved. On the contrary, the rank-based network can be highly resoluted into modules. This is essentially why the rank-based approach was adopted in our study, and we have explained the reason in revision. Nevertheless, in the revised manuscript, we put our focus on the generality of OrthoClust on a variety of networks. The clustering of gene expression profiles is an application. We therefore do not plan to include details of how to construct co-expression networks in the manuscript.

-- Ref 2.3 – Exploration of the parameter space (2)--

Reviewer Comment	Are the results affected if the ortholog graph is unweighted (a simpler assumption) rather than weighting by

	the number of orthologs?
Author Response	Thanks for the question. We added a new section "Weights associated wit the orthology relationships" in the revised manuscript" to explain the importance of the weight. In short, it is very common to have many-to-ma orthology relationships in eukaryotic genomes. Such mappings form bipa clique-like structures in the ortholog graph. We found that using an unweighted orthology graph suggested by the reviewer, for many ortholog cliques, genes are clustered in a single module (see Figure S5). It means t cost function is in favor of the huge cliques and bias against the conserved clusters that are linked by one-to-one orthologs. In contrast, the original weighted orthology graph can be used to further divide the cliques into sub-modules.

-- Ref 2.4 – Robustness with respect to the number of runs--

Reviewer Comment	What happens if simulated annealing is run 16 times instead of 32? What about 64? How robust are the results? The algorithm is not guaranteed to find the minimum so it could potentially give very different answers for purely stochastic reasons. If simulated annealing is run 32 times, for example, and this is done separately several times, how similar are the co-occurrence matrices and resulting clusters each time?
Author Response	We agree with the reviewer that this is indeed an important issue. We therefore included a new section "Robustness analysis" in the revised manuscript. In short, we generated clusters for running the algorithm 8 tir 16, 32, 64 and 128 times. We calculated the consistency between the co-occurrence matrices among each case (as suggested by the reviewer). A expected, the consistency increases if more runs were included (Figure S7 For the case the number of runs is 32, the overlap is 65%, and the overlap between clusters made by 32 times and 128 times is 76%. We think the ca 32 is a reasonable compromise between computational cost and robustnes Indeed, all the other analysis we presented with the clusters constructed b running the algorithm 32 times show statistical significant results.

-- Ref 2.5 – The importance of negative correlation--

Reviewer Comment	How important is it to have a term in the energy function that penalizes genes in the same cluster that are anti-correlated? Most clustering methods simply use the positive correlations. This could be an important

	novel contribution. The authors could show that the results are improved when this term is included.
Author Response	In this particular application, in the networks constructed by our expressio data, the number of negative edges is about 10% of the total number of edges. We explored the differences between with and without the negativ edges by clustering only the fly network, arriving at two sets of modules (and II). We found that the 2 sets of modules are similar. About 80% of pa that are identified as partners in a module for set I are grouped to a modul group II. Moreover, we did not find systematic differences between the I a II. To demonstrate the importance of negative correlation, we further ran algorithm on an artificially generated network (see Supplementary further analysis). In the artificial network, new clusters separated by negative edg are separated. We therefore believe that the negative correlation is in principle important, even though it is not entirely obvious for the data we used. Nevertheless, we want to emphasize that in the revised manuscript, we ha put the emphasis on a more general cross-species clustering framework rather than a particular focus on expression clustering. We therefore toned down language so that the penalty for negative correlation is not a must b possibility.

-- Ref 2.6 – Assignment of ncRNAs--

Reviewer Comment	The assignment of noncoding RNAs to clusters does not yield any interesting biological insights, nor does it help to illustrate the methodology. I would suggest that it be removed.
Author Response	Again, as we have put our emphasis on the generality of OrthoClust in the revised manuscript, the ncRNA section was greatly shortened. The origina Table 1 was removed and Figure 7 was moved to Supplementary information. However, as the concept of cross-species modules is a novel of OrthoClust, and mapping uncharacterized genomic elements from two species to such modules have the potential to infer their analogous functio we believe it is worthwhile to mention this application and highlight a few interesting examples.

-- Ref 2.7 – Discussion on multiple species--

Reviewer Comment	The discussion of how the approach could be extended to multiple species seems to be irrelevant, as the authors state the simulated annealing approach would not scale.

	This section could also be removed.
Author Response	First of all, it is not impossible to use simulated annealing to perform OrthoClust for more than 2 species. Indeed, we did that for three species (worm, fly and human). Nevertheless, we want to emphasize the goal of c manuscript is to present a conceptual framework based on a novel cost function, where simulated annealing is one of the possible way to optimiz cost function. For completeness of the framework, it is still worthwhile to present the generalization by including the multiple species discussion. Although simulated annealing does not scale easily, we, and may be other researchers, may come up with a faster way to optimize the cost function the near future.

-- Ref 2.8 – Similarity based on GO--

Reviewer Comment	On p. 10 the authors define a new similarity metric based on the Gene Ontology. Many such metrics have previously been published and some have been widely used. The authors should justify why those would not be applicable, and why a new metric would be more appropriate.
Author Response	We are sorry that our writing may have confused the reviewer here. We d not invent this metric. It is based on the well known inverse document frequency. It has been widely used in text mining, and it has been employ in quantifying similarity in GO Ontology (cite xxx). In the revised manus we have included a few more references. The major reason of using the similarity metric is for its simplicity. We understand that there are many metrics. As this particular analysis is constitute a rather minor point, we believe it makes sense to use a simple metric (an inner product of two inv document frequency vectors).

-- Ref 2.9 – Modularity--

Reviewer Comment	For Figure 5A the definition of modularity isn't clearly explained.
Author Response	The definition was introduced by Ref. 29. We have updated the text accordingly to make it clear.

-- Ref 2.10 –Software availability --

Reviewer Comment	This is a methods paper and it would be very beneficial for the software to be made available.
Author Response	In this stage, we are happy to provide our MATLAB scripts for readers upon requests. Making a distributable software is one of our future goals. Mainly because of the speed of simulated annealing, we find it pretty hard make a user-friendly software. We are considering other approximate solutions for optimizing the cost function. It may speed up the process.