In:

> Random genetic drift sets an upper limit on mRNA splicing accuracy in metazoans

Benitiere *et al.* have investigated the role that limited purifying selection may have had in the evolution of splicing complexity. Alternative splicing is often cited as an explanation for the evolution of organismal complexity in the absence of an increase in the number of coding genes. However organismal complexity is in itself associated with a decrease in the effective population size (N_e). Hence the alternative hypothesis, that complexity in alternative splicing results from splicing errors that appear due to the lack of purifying selection may also explain this relationship.

Benitiere *et al.* have explored the relationship between selection and splicing complexity by comparing rates of alternative splicing with proxies for N_e in a range of metazoan species, and considered the extent of purifying selection at splice sites in human and Drosophila. Their results argue convincingly that much of the complexity in AS is likely to have evolved due to lack of purifying selection and is thus unlikely to underpin organismal complexity. I think that the work has been done thoroughly and supports their arguments and I have no major issues with the manuscript. However, I note that there is a large discrepancy between their title and the concluding statement of their abstract:

> All these observations are consistent with the hypothesis that variation in AS rates across metazoans reflects the limits set by drift on the capacity of selection to prevent gene expression errors.

I think that the tone of the latter is more appropriate, and that the title overstates the certainty of the conclusions that can be drawn from the work. This is not because of any obvious weaknesses, but because it is inherently a difficult question to answer conclusively. In particular, Chen *et al.* (2014) claimed to have excluded an explanation based on N_e. Benitiere *et al.* do cite Chen, but they do not provide any reason as to the difference in the conclusions reached. There can be a large number of reasons, but the conclusions are incompatible and for Benitiere to be correct Chen must be wrong and this needs to be addressed directly.

I am also concerned that more recent work using long-read sequencing technology (Leung *et al.* Cell Reports, 2021, 10.1016/j.celrep.2021.110022) does not seem to show more AS in humans compared to mice (if anything the opposite was observed). This contrasts with several studies based on short read sequencing and again I feel that these discrepancies ought to be discussed.

I think that the weakest point of Benitiere *et al.* is related to the composition of the data that they have used. They seem to be aware of this, but consider that it could only lead to an under-estimate of the affect of drift on AS. I am not completely convinced by this, and am concerned that the data is likely to comprise sequences from a range of technologies that can influence their

observations. Unfortunately, there is a good chance that the different sequencing technologies will not be uniformly distributed between species owing to the fact that analyses of non-model organisms is likely to have been carried out at later dates and thus with more up to date technologies.

I think that the work would benefit from including analyses from more carefully collated data sets where care is taken to make sure that the underlying technologies are equivalent. Ideally this would be done from species that differ in N_e but which are otherwise similar (eg. marine and fresh-water teleosts). There is also transcriptome data and estimates of N_e in asellid isopods (Lefebure *et al.*, Genome Research 2017, http://www.genome.org/cgi/doi/10.1101/gr.212589.116), who argue that smaller N_e leads to larger genomes as a consequence of less effective selection. If Benitiere *et al.* are correct, there should also be an increase in the amount of low-frequency splicing events in species with lower N_e.

Evaluation of the different components of the article

Title

> *Check that the title clearly reflects the content of the article.*

The title clearly reflects the content of the article, but I think that it is rather too conclusive (especially compared to the conclusion of the abstract).

Abstract

> *Check that the abstract is concise and presents the main findings of the study.*

The abstract is relatively concise (268 words) and clearly summarises the work.

However, I do not think that the work should be considered as a *meta-analysis*. As I understand it, a meta-analysis is an analysis of the results of a set of analyses. Here the authors have made an original analysis of published data and their work does not rely at all on any results of prior analyses. Hence it is simply an analysis.

Introduction

> *Check that the introduction clearly explains the motivation for the study.*

The motivation is abundantly clear.

> *Check that the research question/hypothesis/prediction is clearly presented.*

The questions are also clearly presented.

> *Check that the introduction builds on relevant recent and past research performed in the field.*

The debate about the extenct of the role of alternative splicing is nicely introduced; however, it would be nice to include more recent work making use of long-read technologies that are more suitable for studying alternative splicing (eg, Leung *et al.* Cell Reports, 2021, 10.1016/j.celrep.2021.110022).

Materials and methods

> *More generally, check that sufficient details are provided for the methods and analysis to allow replication by other researchers.*

The methods section of the main manuscript does a reasonable job of explaining what was done, but is unable to provide sufficient detail to describe how the analyses were carried out. This additional detail is provided from an external source (zenodo.org) which provides a large number of data files and scripts. However I've not been able to find a description of the overall pipeline. For example, there are individual R scripts that generate the different figures which is nice; however, these scripts read data from files of processed data, and worse the locations of these files are sometimes outside of the data archive itself.

What is worse is that I am unable to find tables of the original data sources; they may well be there, but to my mind I shold not need to go looking for them as they (eg. identifiers for all of the SRA data, genome assemblies and annotations) are fundamental to the description of the materials used. Hopefully the authors need only provide a more detailed README.md file to address these issues.

> *Check that the statistical analyses are appropriate.*

As far as I can tell the statistics are reasonably chosen; however, I cannot confirm that they have been correctly carried out. But in any case I am not overly concerned about the details of the statistical tests as these do not matter as much as the nature of the data upon which they were applied. That is, I am much more concerned about what unknown factors may affect the analyses in a non-random manner. In this case there may be issues that relate to the sequencing technologies used as well as the choice of species and individual samples that could affect the validity of the conclusions. Unfortunately, although they provide a list of species analysed I have not found more detailed descriptions of the individual samples from which sequencing data was obtained. These details should be included in order to be able to address the validity of the analyses.

However, it is likely to be difficult to address these issues even with such additional meta-data as the problem is inherently complex. To my mind the validity of their conclusions is better assessed by testing predictions made in better characterised species than by tweaking statistical methods.

Results

If possible, evaluate the consistency of raw data and scripts.

This is difficult to do in the absence of additional description of the methods and materials used.

If necessary, and if you can, run the data transformations and statistical analyses and check that you get the same results.

This is not possible within the time frame of the review process.

In the case of negative results, check that there is a statistical power analysis (or an adequate Bayesian analysis).

Not applicable.

Inform the recommender and the managing board if you suspect scientific misconduct.

I do not suspect any scientific misconduct.

Tables and figures

Check that figures and tables are understandable without reference to the main body of the article.

The figures are generally understandable; however in many cases (see below) the authors use terms that are not explained in the captions making it difficult to understand the details of the analysis.

Fig. S1. The term 'average AS rate' is used. The term is defined in figure 2 and in the main text. However, the equations used in the text and figure 2 are not the same and this could be confusing, particularly since the figure refers to AS rate whereas the text refers to RAS and RANS rates.

Fig. S2. The terms RAS and RANS are used without definition; there is a description, but I find it difficult to understand even though I know what RAS and RANS refer to.

Fig. S3. Caption refers to N2 without definition. N2 is from the definition of RAS and its definition can be found in the text. But it would not be possible to know this from the figure alone.

Fig. S4. 'Low AS' and 'High AS' major introns; definitions of low and high not given in caption.

There are other similar examples.

Check that figures and tables have a proper caption.

See above.

Discussion

> *Check that the conclusions are adequately supported by the results and that the interpretation of the analysis is not overstated.*

In general I think the discussion is well supported by the analyses performed. However, I take issues with statements like:

"As predicted, this estimate of the prevalence of functional SVs tends to decrease with decreasing N_e"

As they did not measure N_e, but proxies of N_e, and they are careful to point this out in other places.

> *Check that the discussion takes account of relevant recent and past research performed in the field.*

The discussion is admirably concise whilst including relevant research; however it does not comment sufficiently on past research that claims to exclude the role of genetic drift in the evolution of splicing complexity (see comments above).

References

> *Check that all references are appropriate and that the necessary references are present.*

I have not checked all references, but most seem correct. However, I don't think that Torson *et al.* (2015) says anything about alternative splicing.

I think that the manuscript is adequately referenced.

> *Report any reference cited in the text that does not appear in the reference list.*

(This should not be done manually)