by Ignacio Bravo, 13 Feb 2023 11:20

Manuscript: https://doi.org/10.1101/2022.12.09.519597 version 2

Recommendation for revision of the first version of the manuscript.

The preprint by Bénitière and coworkers has been evaluated by three experts in the field. The three reviewers agree on the importance and on the interest of the manuscript, in terms of relevance of the question addressed as well as in terms of validating a null, neutralist perspective. I largely share the enthusiasm of the the reviewers, which is very evident in their comments. Notwithstanding, I also share with the reviewers some concerns, concerning the concordance (or the lack of concordance) with previous works in the literature addressing similar or parallel questions around fidelity during transcription. Further, the existence and potential impact of possible systematic biases linked to a diversity of sequencing technologies in the databases analysed, and to a differential sensitivity for detecting exon junctions needs to be more explicitly addressed in the text. The reviewers have extensively commented the text in several passages, and I globally agree that addressing the different concerns raised will undoubtedly ameliorate the manuscript for its clarity and soundness. I am convinced that an improved version may be a substantial contribution to the field.

We thank the editor and the referees for their constructive and encouraging comments. These comments were very helpful to improve the presentation of the analyses and the discussion of the results. We have revised our manuscript, and we think we were able to address all the points raised by the referees. Detailed point-by-point responses are given below. The original comments are written in black, and our response in blue.

We hope you'll find our revised manuscript suitable for recommendation by PCI Evol Biol.

Florian Bénitière, Anamaria Necsulea and Laurent Duret

Reviewer 1

Bénitière et al "Random genetic drift sets an upper limit on mRNA splicing accuracy in metazoans"

Lynch's drift barrier hypothesis postulates that, because the efficiency of purifying selection is inversely related to population size, mutation rates should be expected to show a similar correlation. Under this view (https://pubmed.ncbi.nlm.nih.gov/20594608/), selection against "mutators" (genetic variants responsible for higher mutation rates) is driven mainly by selection against the increase in mutation load they induce. As per the first sentence, this selection is expected to be more efficient in larger populations, explaining their typically lower mutation rates. This same perspective (due ultimately to Kimura and especially to Ohta) has been used to explain variation in levels of genetic variability of all sorts observed in nature, and represents a critical benchmark relative to which inferences of positive selection should be made.

Of course, we should have cited Ohta and Kimura. This is now corrected (line 43).

The present contribution applies this line of thinking to splice site variation. Eukaryotic mRNAs often contain introns that are removed (spliced out) before being translated. Moreover, mRNA sequence data makes clear that many mRNAs exist in more than one splice isoform within cells. Previous work has demonstrated that the number of such alternatively spliced mRNAs is positively correlated with organismal complexity (defined as cell type heterogeneity), suggesting the possibility that greater levels of diversity are required to support said complexity. On the other hand, it has also long been understood that organismal complexity is inversely correlated with population size (see also https://pubmed.ncbi.nlm.nih.gov/14631042/),

We agree, this paper by Lynch & Conery is also relevant. We added this reference (line 72)

suggesting the opposite interpretation of the data. Namely, because the efficiency of purifying selection is negatively correlated with population size, we might on first principles predict greater mRNA splicing diversity in smaller populations (as is also seen in their mutation rates). In this reading of the data, that those species happen also to exhibit greater cell-type heterogeneity would be a coincidence, rather than the selective driver of mRNA splicing diversity.

And indeed, the present contribution demonstrates a strong, negative correlation between three proxies for population size and splicing variability using large,

published datasets from both insects and mammals. In both groups of species, putatively smaller populations exhibit higher splicing variability, consistent with the Kimura/Ohta/Lynch drift barrier hypothesis.

This is a very important finding, because it suggests a more appropriate "null hypothesis" under which to test adaptive explanations for splice variability, and I generally regard this work quite favorably for that reason. It will be appreciated both by individuals interested in mRNA splice function and evolution, as well as those more broadly interested in general principles of evolutionary and population genetics.

At present however, I find the manuscript hard to follow in its technical details, which risks minimizing its impact. For the second camp of anticipated readers (which includes me), the complexity of the biology requires more hand-holding to allow easy comprehension of the paper's result. I enumerate points of confusion below, but before that, I also note two points of contact with the literature that seem to be missing at present.

First, the present work brings to mind the drift barrier's impact on transcription error rate (see https://pubmed.ncbi.nlm.nih.gov/26884158/). This phenomenon seems very closely related to the present work, yet those authors find quite a different pattern than that described here. I would therefore be interested in hearing the present authors' views.

Traverse and Ochman (2016) measured transcription error rates in *Escherichia coli* (high Ne) and in two endosymbiotic bacteria (low Ne). They reported that the transcription error rate was higher in *E. coli* (~8e-5 per site) than in the two endosymbionts (~5e-5), which is *a priori* in contradiction with the predictions of the drift barrier hypothesis. Even more surprisingly, these estimates were ~10 times higher than what had been reported in eukaryotes (~4e-6; Gout et al. 2013). However, a more recent study evaluated the transcription error rate in *E. coli* at only 5.8e-6 per site (Li and Lynch 2020), *i.e.* >10 times lower than what was reported by Traverse and Ochman (2016). According to Li and Lynch (2020), the very high error rates measured by Traverse and Ochman probably result from RNA damage during the preparation of sequencing libraries, and thus, do not reflect real in vivo transcription errors. Given these large uncertainties in the measures of transcription error rates, it seems for now difficult to interpret the differences reported across species. This is why we were hesitant to cite the paper by Traverse and Ochman.

One important point however, is that the paper by Traverse and Ochman triggered interesting discussions regarding the predictions of the drift barrier model. Indeed, as highlighted by Xiong et al. (2017), the relationship between the genome-wide error rate and Ne is not expected to be monotonic: in species with very high Ne, selection on each individual gene should favor genotypes that are robust to errors of the gene expression machinery, which in turn, reduces the constraints on the global level of gene expression errors (Rajon and Masel 2011; Xiong et al., 2017). Thus, paradoxically, species with very large Ne are expected to have gene expression machineries that are more error-prone than species with very small Ne (Rajon and Masel 2011). It is therefore possible that error rates at different steps of the gene expression process respond differently to the drift barrier.

We added a paragraph in the discussion (line 407) to discuss this point, and to mention the case of transcription error rates in bacteria and eukaryotes.

And second (more esoterically), I am reminded of much older work on intron phase (e.g., https://pubmed.ncbi.nlm.nih.gov/8618928/), and wonder whether there are any interesting correlations between this intron attribute and splice variability.

We are not sure to understand the point raised by the referee. Long et al. (1995) reported that, among 11,117 introns from eukaryotic genes, 48% were in phase 0, 30% in phase 1, and 22% in phase 2. This non-random distribution of intron phases is even more striking for regions encoding highly-conserved protein domains (1,496 introns) where 55% corresponded to phases 0, phase 1 and 2 corresponded respectively to 24% and 21%.

We investigated those ratios for each species in our dataset. We computed the proportion of phase 0, phase 1 and 2 among major introns from all protein-coding genes and from BUSCO genes. In agreement with Long et al. (1995), we observed an excess of phase 0 introns (Fig. R1A, R1B). In all species, the average AS rate tends to be higher in phase 0 introns, intermediate in phase 2 introns and lower in phase 1 intron (Fig. R1C, R1D). However, the difference in average AS rate across phases is quite limited (0.4% between phase 0 and phase 1 introns) compared to variation in AS rates across species (from 0.8% to 3.8%; Fig 2B). Moreover, the distribution of intron phases is very similar across all species (Fig. R1A, R1B). Hence, variation in intron phases cannot contribute much to variation in AS rate across species.

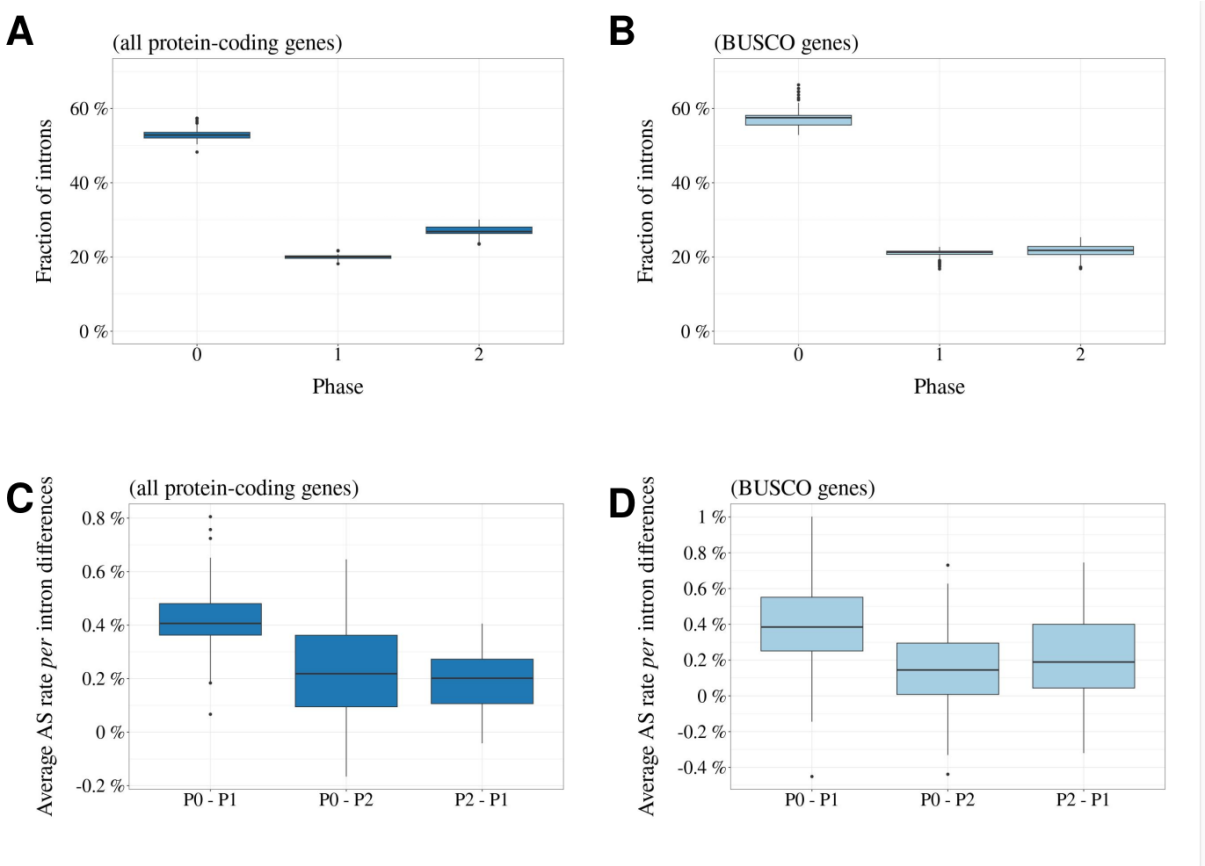

Figure R1: The distribution of intron phases does not account for variation in AS rate across species. (A,B) Boxplot of the fraction of introns in phases 0, 1, and 2 among the 53 species. Per species, average difference in AS rate between phase 0 and phase 1 introns (P0-P1), between phase 0 and phase 2 introns (P0-P2) and between phase 2 and phase 1 introns (P2-P1), (C,D). Analysis was performed on annotated major introns from all protein-coding genes (A,C) and from BUSCO genes (B,D).

1- Line 2: this is trivial, but because of the diversity of definitions (and opinions) of biological complexity (and its evolution), I recommend that the authors explicitly state theirs, e.g., "...noticed that the complexity of organisms (i.e., the number of distinct cell types) correlates positively with…" One would hate to lose readers on such a "partisan" point.

Done (line 3).

2- Lines 33-34: one or two more sentences on the variability of intron splicing efficiency in nature would be most welcome, both with respect to spliceosome and splice signal "quality." I myself have no previous knowledge of how these facts are understood. Adding some details here will help all readers to imagine the mechanisms of splicing accuracy that are putatively under purifying selection..

We added two sentences to add more details (lines 35).

3- Major vs minor isoforms. If I understand the biology correctly, isoform abundances fall into two entirely disjunct classes, as illustrated clearly by figure S2. If that's correct, I would encourage the authors to make that point before the material that begins at line 65. As it stands, I read the sentence that begins "This pattern is mainly driven by low-abundance isoforms…" as casting the situation as a two-dimensional problem, with mRNAs having a spectrum of isoforms, varying both in number and frequency.

Indeed, clarifying that point much earlier might also improve comprehension of the material beginning on line 42. How do major/minor isoforms correlate with the 1% of isoforms that produce detectable amounts of protein? How do we know constraints are weaker on the protein products of minor isoforms? And how might minor isoforms be involved in gene regulation? To be clear, I have little doubt that the facts as stated are correct, and that they reflect the authors' deep understanding of the biology. Moreover, that understanding is likely shared by most readers in the mRNA splicing community. But as noted above, this work also has exciting implications for evolutionary biology, and much of that readership would appreciate more information on the basic facts.

We modified the introduction (lines 47) to try to clarify the different points raised by the referee.

Relatedly, is it true that figure S2 is a histogram of AS frequencies across all species in the final dataset? So each gray dot is a frequency bin for some species, with lines connecting bins within species ? And then the D mel and human bins and lines are highlighted ? Esp because of what I regard as the centrality of this figure for the basic biology (does it perhaps thus deserve promotion to the main text?), I encourage the

authors to explain its construction more clearly. Finally, I don't see the yellow trace for Apis that the legend promises.

Following the referee's suggestion, we moved this figure to the main text (now Fig. 2B). We also extended Fig. 2 to describe more clearly how the different parameters were measured (Fig 2A and 2C). We also modified its legend to clarify its content.

4- The size and shape of the dataset. If I understand correctly, N = 978 is the number of single-copy orthologs across metazoans, ≈80% of which could be unambiguously identified. What does "unambiguously defined" mean? Is this a reflection of incomplete annotation, or something else? Please explain.

We modified this sentence (line 94) to refer to the method section, where this point is explained in details (line 438): "BUSCO genes were removed from the analysis if they were associated to more than one annotated gene or to an annotated gene that was associated to more than one BUSCO gene. "

And how many orthologous introns are there among those genes before and after you apply your >N=10 reads filter? (Trivial point, but recycling the symbol N risks confusion. When I first read line 100, I thought you were down to just N=10 genes. And as just noted, I would have liked to know the number of surviving introns, rather than only the percent surviving.)

We added the number of introns (line 105) and modified Table S1 to include the number of annotated and analyzable introns among BUSCO genes.

Finally, similar to my question about perhaps promoting Fig S1, I had a hard time following lines 86-96 without ready access to the figure. My thought would be to try rewriting this paragraph so as to simply cite the punchline of Fig S1, directing the reader to the supplement to learn more. Or, fully unpack all the details here. The present "hybrid" approach seems suboptimal.

Following the referee's suggestions, we shortened this part (line 99) and moved the details into the Supplementary material (Supplementary Fig. 1).

5- Figure 1A. I was surprised to see a single phylogeny for insects and mammals. We know they exhibit reciprocal monophyly, a fact which in any case has no implications for this study. More seriously, I worry about the positive correlation in 1B, which seems to be driven entirely by the non-insects. (Equivalently, I can

discern little to no trend among points corresponding to organisms whose body length is below 5 cm.) Similarly, removing non-insects from 1C seems to change its message.

I would encourage the authors to explore the story that emerges by the independent analysis of insects and mammals. That might be the more appropriate framing of the data.

As proposed by the referee, we investigated the relationship between the alternative splicing rate and proxies for the effective population size in both vertebrates and insects. We observed that this tendency was consistent in vertebrates and insects using linear regression (significant positive correlation with $p<0.05$ in all cases, except for body length in vertebrates) (Fig. R2).

Note that for these analyses we have smaller sample sizes, so we have a weaker power to detect signal. We therefore prefer to present the combined analyses, using PGLS to account for phylogenetic inertia.

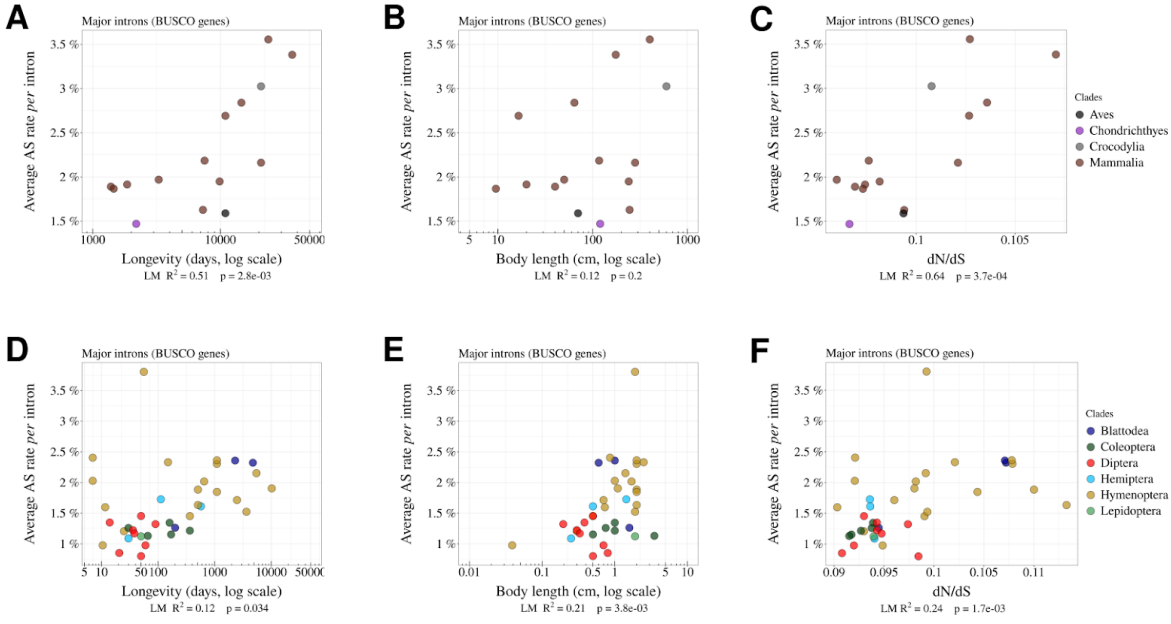

Figure R2: The rate of alternative splicing correlates with life history traits in both vertebrates and insects. Correlation between the average AS rate *per* intron and longevity of each species (days, log scale) (A,B), body length (cm, log scale) (B,E), or the *dN /dS* ratio on terminal branches of the phylogenetic tree (C,F). In vertebrates (A,B,C) and insects (C,D,E). Only the BUSCO genes were included in the analysis.

6- Line 115: that three poorly correlated measures provide noisy estimates of a fourth could be construed by some readers as wishful thinking. I'm not saying it's not so, but especially since the fourth quantity is the fulcrum of the whole study, I feel the authors owe the reader a much stronger explanation here.

The point we wanted to make is that probably none of these proxies provides a perfect estimate of N_e. We modified the sentence accordingly (line 120).

7- I found the section beginning line 116 exceedingly difficult to follow. N1, N2, RANS, RAS? Fig 2A is excellent, but my head is swimming nevertheless! It might help to use more informative variable names, but perhaps more importantly, I encourage the authors to add more English prose to illustrate how these quantities each work.

Following the referees' comments, we changed the names of variables N1, N2, and N3 respectively to Ns (number of spliced reads), Na (number of alternatively spliced reads), and Nu (number of unspliced reads). In addition, we modified Fig. 2 to define all the variables and present how the various parameters were computed (Fig. 2A, 2C).

Unfortunately, this is as far as I was able to get with the manuscript. My inability to internalize these key statistics left me unable to push further.

To summarize, I am very excited about the overarching thesis of this work, and its implications for two communities of readers: AS works and evolutionary geneticists. However as presented, it fails to reach this member of one of the second of those communities. I very much hope the authors will attempt a revision that includes enough intellectual and conceptual hand-holds to help my community appreciate their work. If successful, I have no reason to believe that this paper won't be an important contribution to my field.

Reviewer 2

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 (Ne). 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 Ne 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 over-states 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.

We agree that in the end, we just propose a model (as always), and of course, a short title cannot give all the nuances that can be developed in the text. But we think it is important that the title gives a clear statement of our main conclusion.

In particular, Chen et al. (2014) claimed to have excluded an explanation based on Ne. 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.

Chen et al (2014) measured the rate of alternative splicing across 47 eukaryotic species. They observed a strong positive correlation between the AS rate and the number of cell types (CTN) of an organism. To test whether this correlation might be a consequence of the drift barrier, they focused on 12 species, for which they had measured levels of polymorphism at silent sites (π). They found that the correlation between AS rate and CTN remained significant after controlling for π. They therefore concluded that the association between CTN and alternative splicing was not a by-product of reduced effective population sizes among more complex species.
This conclusion was however based on a very small sample of species. More importantly, it assumed that π could be taken as a proxy for Ne. At mutation-drift equilibrium, π is expected to be proportional to $N_e u$ (where u is the mutation rate per bp per generation). Thus, if u is constant across taxa, π can be used to estimate variation in Ne. However, the dataset analyzed by Chen et al (2014) included very diverse eukaryotic species, with mutation rates ranging from 1.7×10^{-10} mutation/bp/generation in budding yeast, to 1.1×10^{-8} mutation/bp/generation in humans (Lynch et al. 2016). Hence, at this evolutionary scale, variation in Ne cannot be directly inferred from π without accounting for variation in u. Moreover, the drift barrier hypothesis states that the AS rate of a species should reflect the genome-wide burden of slightly deleterious substitutions, which is expected to depend on the intensity of drift over long evolutionary times (i.e. long-term Ne). Conversely, π reflects Ne over a short period of time (of the order of Ne generations), and can be strongly affected by recent population bottlenecks (too recent to have significantly impacted the genome-wide mutation load). The drift barrier hypothesis therefore predicts that the splicing error rate should correlate more strongly with proxies of long-term Ne (such as dN/dS, life history traits, or organismal complexity) than with π. The fact that AS rates remained significantly correlated to cellular diversity after controlling for π (Chen et al. 2014) is therefore not a conclusive argument against the drift barrier hypothesis.

We added a paragraph in the discussion (lines 288) to mention this point.

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.

We agree with the referee that using long-read RNA-seq data would likely improve our estimates of AS rates. However, this type of data is not yet publicly available for enough species, in contrast with short-read RNA-seq data, which is abundant in public databases. We now discuss this point in our manuscript (line 379).

Regarding the differences in AS rates between human and mouse, we would like to point out that the manuscript by Leung et al. did not aim to quantitatively compare human and mouse brain transcriptomes. The data they generated is indeed not directly comparable between the two species: this dataset includes considerably more Iso-seq reads for mouse (5.66 million) than for human (3.30 million). The number of analyzed individuals is also higher for mouse (12) than for human (7). Thus, it is possible that the sequencing depth, which is still a limiting factor for long-read transcriptome sequencing, could affect the authors' estimates of AS rates.

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.

Among the 3496 RNAseq dataset that we analyzed, 3463 (99%) were sequenced with Illumina. The sequencing technologies are therefore very homogenous across taxa. We added a sentence (line 108) to mention this point. We controlled for sequencing depth, which should be the main technical factor affecting AS detection.

It should also be noted that the main results (Fig. 3A) were confirmed when using a subset of species for which the exact same protocol was used to prepare RNAseq data from seven vertebrate species (Fig. 3B).

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 tech-nologies are equivalent. Ideally this would be done from species that differ in Ne but which are otherwise similar (eg. marine and fresh-water teleosts). There is also transcriptome data and estimates of Ne in asellid isopods (Lefebure et al., Genome Research 2017, http://www.genome.org/cgi/doi/10.1101/gr.2125 89.116), who argue that smaller Ne 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 Ne.

We agree with the referee: it would be interesting to extend the analysis by comparing closely related species with contrasted effective population sizes, to limit potential sources of variation that we might have overlooked. We did analyze the asellid isopods dataset (we were co-author of this 2017 study): unfortunately, the RNAseq sequencing depth is not sufficient to quantify AS accurately, and furthermore, a reference genome assembly is lacking for most of these species.

It would be worth investigating whether appropriate data (reference genome + deep RNAseq data) are available for other clades (e.g. marine vs fresh-water teleosts or endemic insular vs mainland passerine birds). However, this would considerably delay the publication of our results (it took us two years to collect the data presented here). We believe that the results reported here are already sufficient to support solid and original conclusions.

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.

Corrected (lines 11 and 268).

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 extent of the role of alternative splicing is nicely intro-duced; 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).

We now mention this point in the discussion (line 379)

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.

We provided in supplementary figure (Supplementary Fig. 10) a description of the pipeline used to process the data.

We also added information regarding the computing resources that are required to process these datasets. (line 461)

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 should 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.

The identifiers of SRA data, genome assemblies and annotations are provided on the zenodo archive, in the file data/Data1_supp.tab. We added a sentence in the 'Data and code availability' section (line 599) to mention this point, and to give a brief description of the main content of this archive. As suggested by the referee we extensively completed the README.md file.

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 analyzed 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.

We now include in the zenodo repository a table (data/data10_supp.tab) providing information on the samples used. Most of the samples come from Illumina platform (3463) and also PACBIO (4), ION_TORRENT (2), ABI_SOLID (15), L454 (4) and BGISEQ (8).

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 characterized 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 sta-tistical 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 sci-entific 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.

We used the 'relative abundance of spliced variants' (RAS) and the 'relative abundance of non-spliced variants' (RANS) to identify minor and major introns, and then used the AS rate of major introns to evaluate the alternative splicing rate in a given species.

We agree that the first version was too complicated. We have changed the main Fig. 2 to describe the different variables and the equations used to compute RAS, RANS (Fig 2A) and the AS rate (Fig 2D).

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. S2 has been moved to the main manuscript (now Fig. 2), where RAS and RANS are now explicitly defined.

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.

N2 (now named Na) is defined in Fig. 2. We modified the caption to mention that point.

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

Corrected.

There are other similar examples.

We checked all captions and corrected them where necessary.

> 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 Ne"

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

We edited this sentence (line 353): 'As predicted, this estimate of the prevalence of functional SVs tends to decrease with decreasing Ne proxies (e.g. Fig. 3A, where *Ne* is approximated by longevity).'

> 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).

We added a paragraph in the discussion (lines 288) to discuss the analysis that Chen et al. (2014) had performed to reject the drift barrier hypothesis (see response above).

References

> Check that all references are appropriate and that the necessary ref-erences 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.

This was an error. We now refer to a review article by John et al (2021) (line 170).

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)

Reviewer 3

Report on

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

by Florian Bénitière, Anamaria Necsulea, Laurent Duret

The authors analyze transcriptome sequencing data from 53 metazoan species to evaluate the hypothesis that genetic drift explains the positive correlation between genome-wide al- ternative splicing rate and organismal complexity (drift-barrier hypothesis). This hypothesis is a (neutral or null) alternative to the (adaptive) explanation that alternative splicing contributes to the evolution of complex organisms. I am very supportive of this idea to evaluate the accordance of this observed correlation with a neutral evolutionary model.

The drift-barrier hypothesis bases on the assumption that many detected alternative splices are splicing errors. These errors are less efficiently purged in species with small effective population sizes, thus explaining the negative correlation of the alternative splicing rate with proxies of Ne. To evaluate this hypothesis, first proxies for the effective population size (Ne) of all the species investigated are defined: body size, longevity and dN/dS. Indeed, based on the available data, the authors find that the alternative splicing rate negatively correlates with proxies of Ne, which is consistent with their suggested hypothesis (Fig. 2).

To further validate their claim, the authors differentiate between functional and non-functional splicing variants. The expectation from their suggested hypothesis is that functional splicing variants, being under selective pressure, should be enriched among abundant splice variants, whereas non-functional variants should be enriched among rare splicing variants. Similarly, increasing the effective population size should decrease the alternative splice rate for non-functional splices and increase the alternative splice rate for functional splices, which is precisely what the authors find (Fig. 3). The authors even move on to further support the drift-barrier hypothesis by two additional tests: The selection strength on splice sites should increase for increasing population sizes (Fig. 4) and the abundance of rare splice variants should decrease with increasing levels of gene expression (Fig. 5). The findings are, again, largely consistent with the drift-barrier hypothesis.

Overall, this is a very convincing assessment of the drift-barrier hypothesis to explain different levels of alternative splicing across metazoans. The manuscript is well written. I particularly liked the introduction and the careful language, i.e., to not jump to conclusions too quickly. The manuscript is also well structured, which helps to follow the line of arguments. There are a few passages though that, in my opinion, need some clarification (more details in my list of comments below). Also, the mathematical model and Fig. 6 do not add value to the manuscript in my opinion. They should be removed or at least moved to the appendix (more details on this below). Nevertheless, I think that this is a very well done and scientifically sound and thorough analysis of a neutral hypothesis to explain the variation of alternative

splice rates across organisms, which merits publication. A list of comments, suggestions and questions follows.

Comments

1- Line 124: The definition of RANS is unclear to me. Why is N 3 divided by 2? I am sure there is a simple reason that escapes my attention. I suggest to add a short explanation (either here or in the Methods section – line 416).

Following the referee's recommendation, we have added more details in the method section (line 472).

2- Line 125: '... at least 10 reads." → I guess this refers to the sum of N 1 + N 2 + N 3, is that correct? I suggest to clarify this.

RAS and RANS are ratios : RAS=Na/(Ns+Na) ; RANS=Ns/(Ns+Nu/2) [see Fig. 2A]. To limit noise, we estimated RAS and RANS only when the denominator was at least 10 (i.e. [Ns+Na]≥10 for RAS, [Ns+Nu/2]≥10 for RANS. Following the referee's comments, we modified this sentence (line 134) to refer to the method section, where this point is explained in detail (line 471).

3- Lines 125-135: The phrasing could be more streamlined in my opinion to avoid ambi- guity. For example, I think that to describe a splice variant only one word should be used consistently (at the moment isoform and transcript are used interchangeably?). Also, again to avoid confusion, I suggest to use minor splice variant instead of splice variant for RAS≤ 0.5 and major splice variant instead of intron for RAS> 0.5 (is this actually correctly interpreted?) – alternatively minor and major intron would also be fine, but just writing intron for splice variants with RAS> 0.5 is unfortunate. As it is, two different terms (splice variant and intron) refer to related concepts (larger or smaller RAS values), which should also be reflected in the words used in my opinion. At least I had problems to remember the definitions and it is sometimes difficult to figure out whether intron refers to any intron or an abundant splice variant.

We agree, it is important to clarify the terminology. But we think that the term 'major splice variant' is not appropriate because the term 'variant' implies that there are several isoforms (at least two), while in many cases, just one single isoform is observed for a given intron. We therefore prefer the term 'major intron', as suggested by the referee, and now use it consistently across the whole manuscript.

We also rephrased this section to avoid confusion (line 140).

4- Fig. 2B,C: I suggest to use the same y-axis scale in the two plots.

Done (NB: Fig2B,C is now Fig3A,B in the revised manuscript) .

5- Lines 159ff.: I think it would strengthen this test of robustness substantially if data from an invertebrate would be added – of course only if feasible. Alternatively, I suggest to emphasize again at the end of the paragraph that all seven species are vertebrates.

The dataset provided by Cardoso-Moreira et al. (2019) is unique and ideal for studying differences between tissues in several species. Indeed, the same protocol was performed by the same laboratory to produce RNA-seq with a large sequencing depth.

For insects, this type of study is not common. Most RNA-seq data come from pooled whole bodies because of the difficulty in dissociating organs. However, we found in our dataset one insect (*Dendroctonus ponderosae*, Coleoptera), for which there are multiple RNA-seq datasets for the ovary, for the testis, and for the head. We observed very little variation in AS rate across these 3 organs (Supplementary Fig. 9), which is consistent with what we reported for vertebrates. We added a sentence (line 179) to mention this new analysis.

6- Line 191: I was confused by the definition of MIRA. Is there a mistake in the denominator? Should it be N 1 minor intron? (see also line 426)

We have modified Fig. 2 to better describe the variables and equations used to define MIRA (Fig 2D).

7- Fig. 3B-D: These panels are not referenced in the main text (or just later in the discussion). I suggest to either move them to the Appendix or, better, to comment on them in the main text close to the figure. I think they make a good case for the drift-barrier hypothesis, which should also be mentioned (earlier) in the main text.

These panels are presented in the discussion section (line 322). We agree that these observations strongly support the drift-barrier model (and this is what we argue in the discussion). We tried to comment on these results earlier in the manuscript, but we felt that this was too heavy, and redundant with the discussion.

8- Line 225: 'significant' → I personally try to avoid using the word 'significant' if it does not relate to a statistical test. Here, I think it belongs to a statistical test, but then also the p-value should be given. (This also refers to other places in the manuscript.)

We added the p-value (line 235).

9- Line 228: This is actually a strong argument against the adaptive hypothesis (large alternative splicing rate in complex organisms) and I suggest to spell this out explicitly.

Yes we agree. We added a sentence to highlight this point (line 238).

10- Line 259: I would write 'proxies of Ne values' for the sake of precision.

In that sentence, we state that our dataset includes species covering a wide range of Ne values. We think that this statement is correct, even though we cannot directly measure Ne.

11- Lines 267ff.: I suggest to add ranges of values throughout this paragraph.

Done (line 278).

12- Line 273: I suggest to cite the Bush et al. paper already in the introduction where the drift-barrier hypothesis is introduced because the idea is put forward in this paper (e.g. their Section 4). In general, I think that this paper would merit to receive some more credit for the drift-barrier hypothesis idea earlier in the paper. Essentially, the manuscript by the authors is exactly doing the suggested

comparative analysis across multiple species to assess the roles of genetic drift and selection on the alternative splicing rate.

We agree. We now cite Bush et al. in the introduction (line 74).

13- Line 278: I suggest to replace 'the others' with the respective precise term (I guess rare SVs with MIRA< 5%).

Done (line 312).