PCI answer to reviews on "Genomic imprinting mediates dosage compensation in a young plant XY system"

We are thankful for the anonymous reviewers' and the Editor's constructive feedback. We find their comments allowed us to significantly improve the manuscript and we hope that it will be considered for recommendation by PCI.

This manuscript has been evaluated by two referees, who agree that the findings represent an important advance in our understanding of the early stages of dosage compensation in sex chromosomes, an important topic in evolutionary biology. The second referee nevertheless raises concerns about the lack of a reverse cross, which impedes disentangling maternal/paternal imprinting from strain effects. The second referee also suggests that the discussion should be more balanced, highlighting the differences in pattern strength found between figures and experiments. Both referees further suggest some improvements regarding clarity, mainly on the hypotheses in the introduction and discussion, on the outgroup and the methods, and they have a list of questions which the revised manuscript should answer to. The main figures are missing and should be added in the PDF. In addition to these referees' suggestions, I also have the following recommendations for improving the manuscript. As a non-specialist of dosage compensation, I found the main text hard to understand and had to read the text several times to fully understand hypotheses and findings. If the authors target a broad audience, it would be useful to expose more clearly hypotheses and inferences, using less specific jargon. If room is needed, I would move the discussion on proximal mechanisms to supplementary text, and also the discussion about buffering explaining sex equality, as this pattern is not found in the present study. In addition, I wondered whether "using the ratio of Y over X expression levels in males as a proxy for Y degeneration" could not be circular. Indeed, the ratio of Y over X expression levels in males includes both the effect of Y degeneration and X compensation in males, and not only degeneration. Looking at dosage compensation as a function of a measure of Y degeneration that integrates dosage compensation may sound circular. I do not think the inferences are circular though, but I would be more careful in the definitions in the text and in the figures, and it may be better to take as a measure of Y degeneration differences or ratio of expression between Y and the outgroup autosome? About the main figures, why are they plotted and statistically analysed as discrete classes rather than as continuous variables? To sum up, there is potential for a highly interesting and novel contribution to the field of evolutionary biology. However, the paper needs careful revision along the lines above. If you are able to accommodate these points, I would encourage resubmission to PCI Evol Biol for recommendation.

The missing figures have been added and the text revised to improve clarity.

Figures have been plotted as discrete classes to allow for statistical testing and also to better visualise patterns as variance is high in the data maybe due to noise in the data and also varying degrees of dosage compensation for the different genes.

We used Y/X ratio because, although it is a bit circular, it allows to categorize genes with high dosage compensation levels together, which Y over outgroup ratio does not do. In future a better way to categorize genes would be to have their position along the X chromosome in order to categorise genes by age of X-Y divergence. However, this is not doable with current data (gene position was not resolved on the genetic map). Another solution to categorise genes by age would be to estimate X-Y divergence but this requires DNA sequences for X-hemizygous genes. Using

RNA-seq data for this purpose would be highly biased. This is why we stick to the Y/X ratio in this manuscript.

Reviewed by anonymous reviewer, 2017-10-17 19:30

Muyle et al aim to address an important question in sex chromosome research, the mechanism by which dosage compensation evolves. They study gene expression patterns across sex-linked genes in Silene latifolia, a plant with young sex chromosomes. They show that Y degeneration is associated with upregulation of the maternal X allele, presumably to compensate for unequal gene dose, and provide support for the steps proposed by Ohno in his theory of dosage compensation. Not only does this research provide further insight into the status of dosage compensation in plants, which have been relatively understudied relative to animals, but also represents an important advance in our understanding of the early stages of sex chromosomes evolution. I very much enjoyed reading this paper; it was very clear and represents an important advance. I have some suggestions that should help to improve the study.

We are thankful for this positive consideration of our work.

The discussion of the steps associated with the evolution of dosage compensation in the introduction are a little confusing. Currently, the introduction reads that equal expression between the sexes is a form of dosage compensation. This has been debated (Mank & Ellegren 2009 Heredity; Melamed et al 2009 Heredity) however, according to the definition proposed by Ohno, expression equality between the sexes is just a side effect of dosage compensation, and maintaining dose between the X and autosomes is the key step. Unequal gene dosage will select for the upregulation of the single X in the heterogametic sex in order to re-establish ancestral expression levels, and then compensatory evolution in the homogametic sex will restore balanced gene expression. Following this, it would be helpful if the introduction were revised to more clearly outline these steps. In addition, it might be useful here to mention briefly why unequal gene dose between the X and autosomes is thought to negatively affect fitness.

We modified the introduction following these suggestions.

It would be useful to have information in the main text about the outgroups and when they diverged from S. latifolia. When did the sex chromosomes in S. latifolia evolve? This would also be helpful in understanding why the authors averaged expression level across the two outgroups – are they equally diverged from S. latifolia? It is difficult to interpret how this may have influenced the results. Can the authors show that expression is S. viscosa and S. vulgaris is highly correlated for all genes and for Y/X genes with Y degeneration?

We added the information on divergence between *S. latifolia* and the two outgroups in the main text (~5My) and also provide a new supplementary figure (S1) showing a phylogeny with ages of sex chromosomes and outgroup divergence time.

The two outgroup expression levels are highly correlated: for flower buds R-square = 0.7 and p-value < 2.2e-16, for leaves R-square = 0.5 and p-value < 2.2e-16. This information has now been added in the Supplementary material.

Figure 1 and Figure 2 appear to be missing. However, the supplementary figures were very clear and detailed. There appears to be a bracket missing on the Y axis.

Sorry for not including the main figures, they have now been added. However, we were unable to find the missing bracket.

The difference in dosage compensation between X hemizygous and X/Y genes is very interesting but only discussed in the supplementary text. I would recommend moving this to the main text and discussing there the potential dosage insensitivity of X hemizygous genes. The authors don't report whether X genes in general are significantly depleted in ribosomal protein coding genes. This is an important test, is it just X hemizygous genes, is there a significant difference between X/Y genes? Related to this, it would be useful to mention in the figure legends that X hemizygous genes have a mean Y/X expression ratio in males = 0 whereas all other categories are X/Y. For example, on page 3 it says 'for X/Y contigs, the difference between maternal and paternal X in females increases with Y degeneration', but clearly this is not the case for genes with mean Y/X expression ratio of 0.

Unfortunately, we are running out of space and chose to keep the discussion on X-hemizygous genes which could be challenging for a broad readership to the supplementary material. We however added one sentence in the main text which summarizes our results on lower dosage sensitivity of X-hemizygous genes.

More detail has been added on the GO-term analysis in Supplementary text 1. Under-representation of ribosomal protein is only significant when comparing X-hemizygous genes to autosomal genes. This is not the case when comparing X/Y genes to autosomal genes.

Details were added in the supplementary Figures regarding the X-hemizygous category (for simplicity this category is absent from the main figures).

For genes with no Y decay (mean Y/X expression ratio in males = 1-1.5), the Y allele appears to be upregulated in males relative to the ancestral allele. Is this because of sex-specific selection?

Note that most genes that have evolved sex-specific functions have been studied in Zemp et al. (2016) and were removed from this study. However, it is indeed possible that these genes with higher Y expression are undergoing specialisation in the male function.

It would be useful to mention in the methods why it is important for SEX-DETector that X and Y sequences are assembled in the same contig. This is mentioned in the reference transcriptome section but not explained.

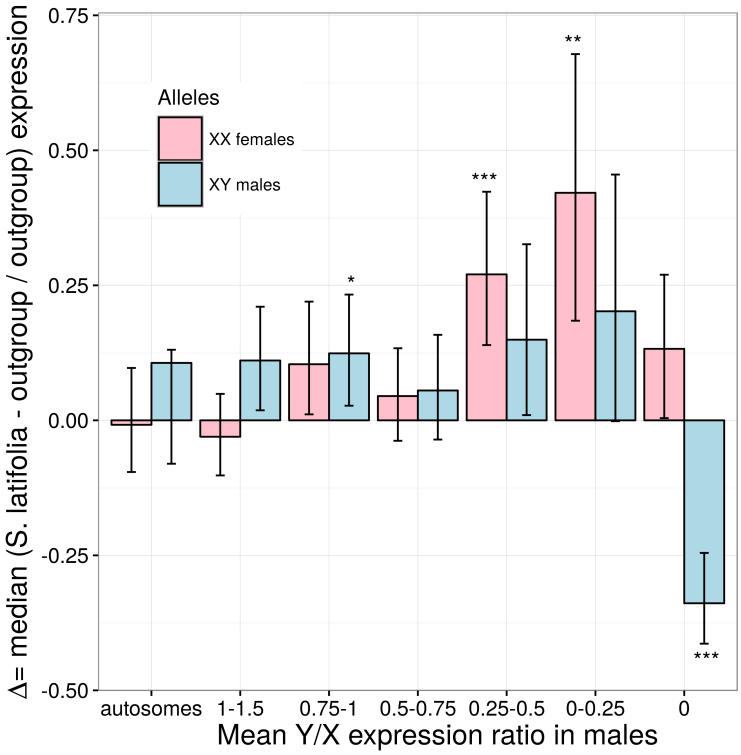
This is explained in Supplementary Text S1: If the X and the Y copy are too divergent to be assembled together, the X contig will be wrongly inferred as X-hemizygous by SEX-DETector because Y alleles will be absent from the contig.

It would be good to mention dosage compensation in Tribolium, which seems to have a parallel mechanism to the pattern the authors find. The female X chromosome is hyper transcribed relative to autosomal expression levels but the X and autosomes in males are compensated. Prince, E. G., Kirkland, D. & Demuth, J. P. Hyperexpression of the X chromosome in both sexes results in extensive female bias of X-linked genes in the Flour Beetle. Genome Biol. Evol. 2, 336–346 (2010).

We now refer to this work in the main text, however recent analyses suggest that these previous results were due to biases from inclusion of gonads in whole body expression level analyses (see recent review by Gu et al in GBE).

I assume that the genes shown in Figure 1, Supplementary Figures are genes with unbiased expression (as sex-biased genes were removed). It would be useful to have another supplementary graph showing the magnitude of expression differences between the sexes for these genes. I assume that the expression equality is achieved despite upregulation of the maternal X in females, because the upregulation of the maternal X in males is much greater?

Sex-biased genes were identified using a differential expression analysis, which is conservative and only identifies genes showing very large differences between the sexes (typically higher than a two-fold difference). It is important to realize that "unbiased" genes can still show some differences in expression between the sexes. Indeed we only showed contigs with "unbiased" expression in all figures. We did not include a figure showing global male and female expression in the manuscript because we thought the allelic expression figure would be enough. But as indicated in the main text and as can be deducted from Figure 1, sex equality is not achieved, see for example here a figure of global male and female expression in seedlings:



Reviewed by anonymous reviewer, 2017-10-23 09:56

Mule et al. look at allele-specific expression of S. latifolia, a plant whose young sex chromosomes have acquired partial dosage compensation. They find that, contrary to the expectation if this was simply due to general buffering mechanisms, the maternally derived X chromosome of females is also up-regulated relative to the outgroups. They interpret this as evidence for an imprinting-based mechanism of dosage compensation. S. latifolia is a great model in which to study the evolution of dosage compensation, and the examination of paternal and maternal allele expression is an important step forward.

We are thankful for the anonymous reviewer's positive feedback.

My main concern is that, as far as I understand, the authors did not perform the experiments on the reverse cross (Leuk144-3*mother x U103*7_father), and therefore cannot control for strain of origin effects. My impression is that this would be essencial for imprinting to be conclusively detected. As it is, the results are very suggestive but difficult to interpret. Since they pick genes that:

-do not have sex-biased expression

-do not differ in expression with the outgroup

-whose expression is lower for the Y copy than the X copy

Then in males these genes will by definition have increased X (maternal) expression. Since the

same maternal X (or at least an X from the same strain) is being transmitted to the daughters, increased maternal expression may be expected even if there is no imprinting.

As explained in a response to referee 1 comment, only genes with very large differences in expression between the sexes were removed. They represent a minority of the genes, that is 4.5% in leaves, 15.2% in seedlings and 25.3% in flower buds. The majority of genes were kept in our analysis.

We did not filter genes for having similar expression to the outgroup. Also, we considered different categories of Y/X ratio but did not filter for having lower expression of the Y copy, we have categories where X and Y copies are expressed at a similar level. We did not control for the reverse cross, however, the parents both come from Switzerland and we expect little effect from the population of origin as they are fairly close. The direction of the cross is known to have an incidence in interspecies crosses but here there is little divergence between the two parents. Of course it would be ideal if our results could be confirmed on a wide range of parental samples across Europe but we are unfortunately unable to provide these controls at the moment.

Another issue is that the patterns are not as clear as figure 2 and the main text seem to imply:

***The 0.75-1 category looks very different in figure 1 (no difference between maternal/paternal allele) and figure 2 (very significant difference).**

Figure 1 and 2 cannot directly be compared. Figure 1 shows medians of the delta value, while Figure 2 shows the differential effect of paternal and maternal origin of alleles, in interaction with degeneration level, while controlling for the variability due to genes and individuals and taking the outgroup as a reference (offset). It is thus expected that the patterns are different between both figures. Also, even if there is a significant effect in Figure 2 for category 0.75-1 but not in Figure 1, the intensity of the effect is low. Lastly, significance is higher in Figure 2 because the data consists of SNP read numbers for every individual while in Figure 1 there is only one value per contig averaged over all individuals and SNPs.

***For 0.5-0.75 there seems to be a difference in figure 1, but not due to overexpression of the maternal allele.**

The difference in Figure 1 is not significant for this category, both maternal and paternal alleles are expressed at levels similar to the outgroup.

***in flower buds the effect is stronger for the 0.25-0.5 category than 0-0.25.**

Indeed but the difference is not significant as error bars overlap.

***in leaf and flower bud an excess of maternal allele expression is detected for the autosomes in females, and in leaf an excess of paternal expression is detected in males for the 0.57-1 category.**

Indeed there is a pattern that looks like imprinting in the leaf and flower bud dataset for the autosomes. We think this is due to a bias we were unable to control for. In this case we rather choose to take the autosomes as a reference in order to consider sex-linked genes, and indeed imprinting is much higher in sex-linked genes compared to the autosomes for these two tissues.

About the significantly higher Y expression of some contigs for the 0.75-1 category, as suggested by the other reviewer, it could be the result of sexual selection.

***there is generally no significant excess of female maternal allele expression for the validated contigs (figures s4 to s6).**

It is now clarified in the main text that the analysis on the validated set of contigs lacks statistical power.

I understand that these could simply be due to noise/lack of power, but they should be

addressed more openly in the text. It was particularly confusing that even though flower bud data was used for the sequencing of the parents, and for the reference transcriptome, seedling is the tissue shown in the main figures. This should be justified.

Historically we sequenced the flower buds first. Only later during the project funding was obtained for two other tissues. This is why the parents were sequenced for flower buds only. In the end the flower bud tissue was the most sex-biased (supplementary Table S2) and therefore not the best to study dosage compensation. The seedlings are interesting because they correspond to an early developmental stage where dosage balance could matter more than in adult tissues. It was indeed shown in some animals that the degree of dosage compensation is higher for larval stages than adults.

Other:

***It seems surprising to me that there is no differential expression between S. latifolia and its outgroups - such differences are usually found even between strains of the same species.**

The species are very closely related (5My) probably explaining why globally there is no expression changes. However if we did a differential expression analysis between the two species then of course some genes would show significant differences but we are interested in global patterns.

***Random selection of 200 autosomes: I think an important additional control would be to use the smallest sample size (79) and run a bootstrapping procedure to get a probability of being different from 0 in female.**

We chose 200 in order to be conservative and have a higher sample size than the X/Y category with the lowest number of genes. The aim is to reliably test whether autosomes in S. latifolia have an expression level similar to the outgroup species, so taking a lower sample size would make this test less reliable.

***"Autosomal normalised expression levels in the two outgroups (S. vulgaris and S. viscosa) were averaged together." --> this does not seem like a standard approach. Was the analysis checked using each separately?**

The graphs are slightly different when taking the two outgroups separately, however we believe that averaging expression levels between the two outgroups allows to get closer to an estimate of the ancestral autosomal expression level, regardless of specific expression changes in species. That is why we prefer this procedure when possible (for the seedling data we only have sequencing for one outgroup).

***The numbering in the supplementary methods is inconsistent (e.g. 4) Validation of sex-linked contigs and 4) Expression level estimates).**

Thank you for noticing these typos, they are now corrected.

***Normalized fpkm: was this normalized any further than just calculating fpkm, which is often not sufficient? If not, did you check the distribution of fpkm between samples to exclude biases?**

We did not normalise further than with FPKM for Figure 1 results. However, for Figure 2 the effect of individuals on expression is accounted for in the model so any bias is corrected. For the identification of contigs with unbiased expression we have used TMM normalization.

***The text could do with a few more details, even if most of the methods are supplementary. For instance:**

-you don't mention the crossing scheme that you used at all.

The cross has been used in other published papers and we only cite these papers to avoid redundancy.

-In page 2, a sentence explaining how you defined sex biased genes would have been useful.

Indeed we are unfortunately running out of space in the main text and the details are shown in the supplements.

-The number and type of tissues used is never mentioned in the main text.

It is mentioned in the third paragraph.

-In page 3, it would be useful to have a qualification of the close relatedness of the species.

This information has been added in the main text (~5My) and in a new supplementary figure S1.