We thank the reviewers for providing thoughtful comments that have helped us to improve this manuscript. We have incorporated changes that address the reviewer’s concerns, and significantly improved the clarity of the manuscript. We hope that you find the revised manuscript suitable for publication. Below we have included each of the reviewer's comments and our responses in red. Any text that has been added to the manuscript is italicized in this response letter.

**Reviewer 1**

A main issue is that the authors praise their current study to go deeper than a former study that looked at the impact of holocentricity at the level of orders (line 75) and indeed the authors use several genera per order. However, the results amalgamate the inferences again to an order level (same in the abstract - line 26). It would be important to assess and discuss the variation within each order also because such variation may provide hints about the underlying processes. This has been shown for Lepidoptera in De Vos et al. (2020 Philosophical Transactions of the Royal Society B) where the authors compared rates of speciation in relation to chromosomal variation.

The pervious study the reviewer mentions only assessed simple difference in chromosome number. We have clarified this in the introduction.

Lines XXX.

*“This previous study was limited to an order level analysis and only tested whether the mean chromosome number among monocentric and holocentric clades was different.”*

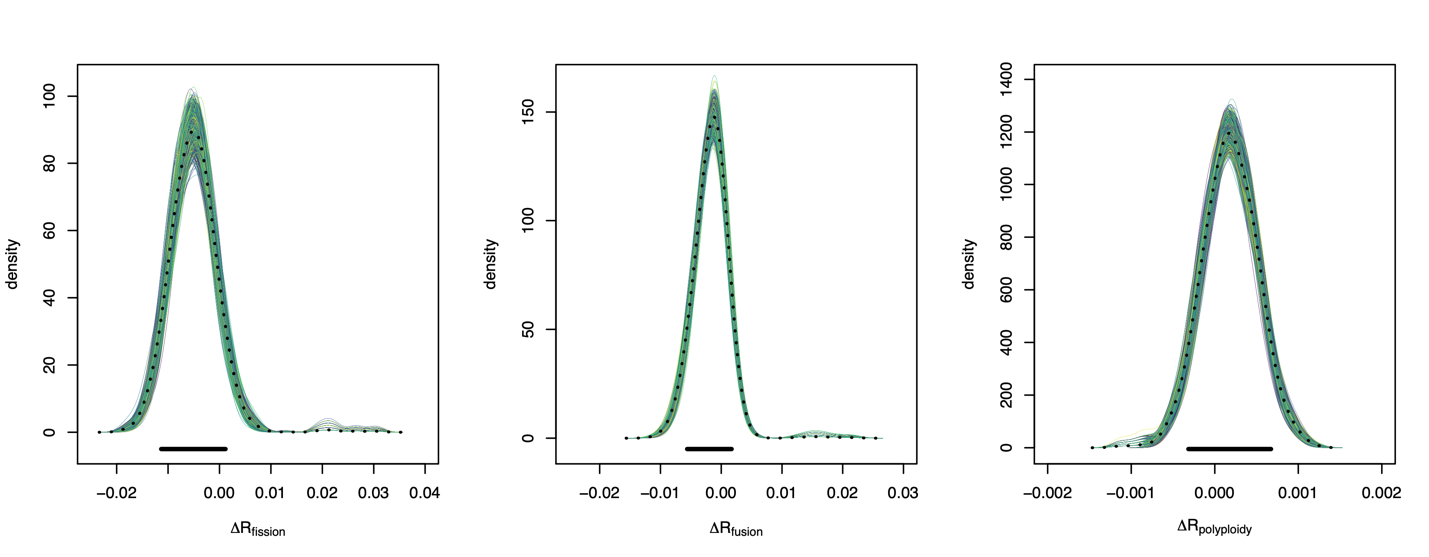
The study we present here uses a recently developed method of estimating rates of chromosome fissions and fusions [1]. Since all clades are monomorphic with regard to centromere type reviewing the data at an order level still seems appropriate (Figure 3) but this is done in addition to the insect wide analysis that incorporates all available data for all orders simultaneously (Figure 2).

With regard to variation within an order, we agree that this is an interesting idea. However, there is currently a lack of methods to evaluate rate variation within a phylogeny for the evolution of discrete traits. The only existing methods are like the one we have applied here where the investigator must a priori provide a hypothesis about what groups of species are allowed to vary in rates of evolution (e.g. in the current study we allow holocentric and monocentric lineages to have different rates of evolution). We are currently in the process of developing comparative methods that would allow for a more agnostic approach to investigating regions of increased or decreased rates of evolution in discrete traits but it is still under heavy development and testing and is not ready to be applied to any empirical analyses.

Another potential issue I see is that chromosome numbers per genus can be very variable. The authors tried to get around with this by sampling from the distribution (lines 238/239), but how robust is this and how does such variation relate to the variation in R?

The sampling approach that we use is relatively common in comparative analyses. However, we were unable to find any robust analysis that shows how it might impact inferences of this type. We chose to perform a bootstrap analysis to test for sensitivity to sampling of chromosome number and for the use of only 100 trees from the posterior distribution. This analysis is now included in the supplemental material:

*One potential concern with our analysis approach is that we are using a phylogeny with only a single tip for each genus included in our analysis, but in many cases, we have multiple species in a given genus and sometimes they vary in chromosome number. Our solution was to randomly sample from all species in a given genus and assign one of the observed chromosome numbers for each genus tip in our phylogeny. Using these sampled chromosome numbers, we then estimated rates for the current tree. Next we repeated this process for each of the 100 trees from the posterior distribution. Finally, we combined the post-burnin portion of our MCMC performed on each tree to generate a posterior distribution incorporating uncertainty in both phylogeny and tip state. This combination of a sampled tree from the posterior and a sample of possible chromosome number assignments to each genus will be referred to as a sample set below. To assess the impact uncertainty in trees and chromosome number data we conducted a bootstrap analysis with 1000 replicates. Briefly, for each bootstrap replicate we took our existing MCMC log files and chose 100 of them with replacement. This led to an average of 63 sample sets being used for parameter inference and in most (greater than 90%) bootstrap replicates one or more sample set was included four or more times. With this approach if some sample sets lead to very different answers we expect to see variation in our calculation of the delta R statistic that is reported in the paper.*



**C**

**B**

**A**

***Supplemental Figure 3. Comparison of bootstrap and empirical estimates****. In each plot we show the statistic for the three parameters of interest in our model. In each plot colored lines show the density distribution of 1000 bootstrap datasets. The black dashed lines show the density distribution from the empirical dataset. The solid black line at the bottom of each plot shows the limits of the most extreme credible intervals from all 1000 bootstraps. If a bootstrap dataset conflicted with our empirical analysis it would have a credible interval where the lower value was greater than zero or its higher value was less than zero. All 1000 credible intervals span zero.*

*Our results indicate that sample sets have no significant impact our inference approach. None of the 1000 bootstrap replicates led to inferences different from those reported in the main body of the manuscript. More broadly we would suggest that this type of bootstrap replicate should become a more standard part of comparative methods to assess whether estimates have been marginalized over a sufficient sampling of trees.*

Line 27: Here and throughout the manuscript it would be important to state in which orders polyploidy is common. It is for example highly debated if polyploidisation occurred in Lepidoptera – the authors only highlight that there were ancient whole genome duplications (line 186-188) but even that has been debated (ref 41 in the manuscript). See also Lukhtanov et al (2015, Proceedings of the Royal Society B).

We agree that it would be beneficial to add those orders where WGD is common when discussing polyploidization in insects, therefore we have added the following to lines XX in the manuscript: “There is support for WGDs in 18 hexapod species encompassing 12 orders (Ephemeroptera, Zoraptera, Plecoptera, Embioptera, Blattodea, Thysanoptera, Psocodea, Hymenoptera, Coleoptera, Trichoptera, Lepidoptera, Diptera). 5 of these 12 orders with WGDs are included in our study.” Additionally, the authors appreciate that there is a debate over polyploidization in Lepidoptera and have added the additional literature to the manuscript.

Line 48 onwards: It is important to note that these processes described here, e.g. Robertsonian translocations, are based on monocentric chromosomes. I would suggest to expand this section to also indicate processes involved for holocentric chromosomes (reviewed in ref 21 in the manuscript).

This is a good point, Robertsonian translocations are defined by the position of the translocation relative to the centromere and thus are not possible in a species with holocentric chromosomes. We have reworked the text in this region to make it inclusive of various processes.

*First, translocations followed by the possible loss of a small fragments of one chromosome can decrease chromosome number (e.g. Robertsonian translocation in monocentric species) [2].*

Lines XXXX

Line 63: However, many species are often not variable at all in terms of chromosome numbers, such as Lepidoptera where most known species show a karyotype close to the putative ancestral state (ref 36 in the manuscript and De Vos et al. 2020).

We agree that segregating variation in chromosome number is rare. However, at this point in the manuscript we are giving examples where individuals with heterozygous chromosome number have been evaluated with regard to fitness loss due to segregation errors. This is the logical basis of the widely accepted idea that we are testing in this paper (does holocentricity lead to higher rates of chromosome fission because they are able to be segregated without difficulty?) For this reason, we have not changed the text in this passage.

Line 70 onwards: This argument makes only sense if holocentricity would result in fission events, yet fusion is also possible.

We disagree with the reviewer’s statement here. Even if holocentricity leads to an equal increase in both fissions and fusions if both rates increase the expected maximum chromosome number will be higher in the clade with higher rates of fusions and fissions. To illustrate this point, we have performed a simulation study.

Here is the R code necessary to replicate this simulation

# load packages

library(chromePlus)

library(diversitree)

# set seed

set.seed(1)

# simulate a phylogeny under birth death model

tree <- tree.bd(pars = c(3,1), max.taxa=100)

# scale the tree to unit length

tree$edge.length <- tree$edge.length/max(branching.times(tree))

# simulate chromosome number evolution

# we will simulate 100 datasets with

# a fusion and fission rate of .1 (slowdat)

# and an additional 100 datasets with

# a fusion and fission rate of 1 (fastdat)

# for each simulation we will record the minimum and maximum

# chromosome number

fastmax <- fastmin <- slowmax <- slowmin <- c()

for(i in 1:1000){

print(i)

temp <- simChrom(tree, pars=c(1, 1, 0, 0, 10),

limits = c(1, 20), model = "2010")

fastmax[i] <- max(temp)

fastmin[i] <- min(temp)

temp <- simChrom(tree, pars=c(.1, .1, 0, 0, 10),

limits = c(1, 20), model = "2010")

slowmax[i] <- max(temp)

slowmin[i] <- min(temp)

}

df <- data.frame(c(fastmax, fastmin, slowmax, slowmin),

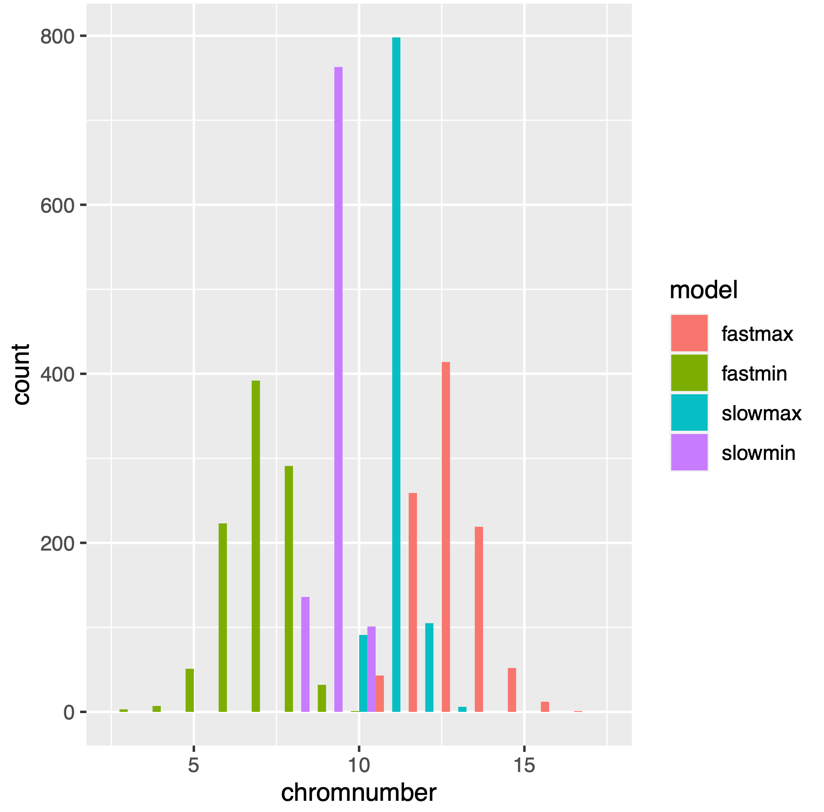
rep(c("fastmax","fastmin","slowmax","slowmin"),

each=1000))

colnames(df) <- c("chromnumber","model")

ggplot(data = df, aes(x = chromnumber, fill = model)) +

geom\_histogram(position = "dodge", binwidth = 1)



What this plot shows us is that when we have a higher rate of evolution (peach and green) we see both a higher maximum and minimum chromosome number than when we have a lower rate of evolution (blue and purple). This is akin to the behavior of a Brownian motion model where though the expected value does not change with rate the variance is proportional to time and rate. The chromosome model implemented here is identical in this respect.

Furthermore, much of the previous literature surrounding holocentricity has focused on the ability to segregate fragmented chromosomes. Therefore, we focus in this paragraph on explanations for high chromosome number including one of our previous analyses. Finally, we agree that we might expect both rates to increase and that these rates are really much closer to what it is we are interested in. Indeed being able to answer this question was one of several motivating desires in the development of this software that we are now applying to this question.

Line 101: Perhaps remind the reader what rate you are looking at.

Corrected

Line 106: Clarify that you refer to the tree of Misof et al.

Corrected

**Reviewer 2**

First, the title conclusion is too broad given the scope of the paper. The title should be changed to:  
  
Chromosome number evolves at equal rates in holocentric and monocentric insects.

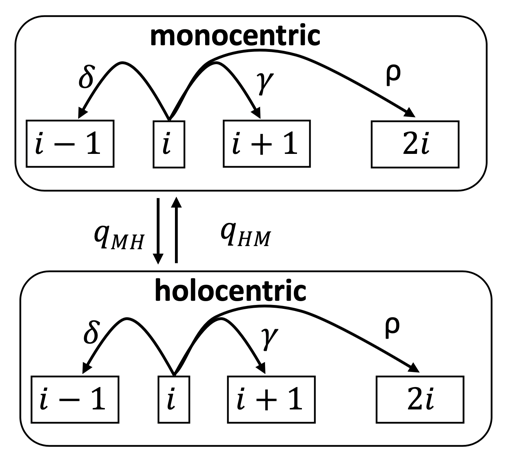
We agree that the title was a bit broad originally. We have changed it to

*Chromosome number evolves at equal rates in holocentric and monocentric clades*

We do not believe that it is necessary to specify insects in the title. The information that has informed our understanding of the differences in monocentric and holocentric chromosomes comes from both plants, insects, and non-insect animals and the segregation behavior in all of these groups is largely similar.

Second, the machinery for rate estimation is too opaque. In particular, the authors should provide a simplified version of the core likelihood equation that is being used.

Because this method has already been published, rather than including a likelihood equation, we decided to include a graphical depiction of the Markov model being used (Supplemental figure 2).



*Supplemental Figure 2. Model for the evolution of chromosome number in monocentric and holocentric lineages. At an instance in time a lineage will have chromosome and either monocentric or holocentric chromosomes. A lineage can make four possible transitions: the fusion of two chromosomes, the fission of a chromosome, a whole genome duplication, and a transition in centromere type (i.e. transition from monocentric to holocentric qMH or transition from holocentric to monocentric qHM).*

Some clarification should be provided on how the polyploidization rate is estimated. In particular, would a polyploidization even be inferred when there is a doubling of chromosome number inferred in an ancestor? Presumably there is no other karyotypic signal identified. What signal in the phylogeny provides the support for these estimates? Some brief discussion would be helpful.

As depicted in the figure above polyploidy provides a route for an immediate doubling of chromosome number. Because of this, anytime the data suggests a very rapid transition from i to 2i chromosomes it will support a higher rate of polyploidy. This approach to inferring polyploidy has been quite well studied [1,3-6]. We believe that the inclusion of the Markov model figure from above will make this easier for readers to understand.

Finally, since the machinery for estimating fusion, fission and polyploidization rates is novel, it would be worth running some simulations on the tree that is being used, with fixed rates on subclades, followed by rate estimation on the simulated tip values for chromosome number. The degree to which the machinery used is good at estimating rates from simulated data is important to know.

The model for estimating these rates is not novel and is already published [1]. This paper contains the simulations that are being suggested here.

**Reviewer 3**

My main concern about the paper is that not enough attention is devoted to the underlying  
HMM model, and to the differences in results between models with and without polyploidy.  
Although the model is explained in an earlier publication (Blackmon et al. 2019), the current  
manuscript would perhaps be easier to understand for the reader if at least a minimal  
conceptual schematic was provided, showing how the eight parameters of the full model are  
related to each other.

First, we note that as implemented there is no unobservable state in our model so it isn’t actually an HMM it is instead a standard Markov model. Second, we agree that it might be easier for people to understand if we provided a schematic and have included supplemental figure 2 in response to this comment and one from reviewer 2.

In the order-level analysis, the exclusion of the polyploidy parameter  
leads to an opposite conclusion. The basic reason for it seems obvious and there is some  
discussion in lines 185-193. However, what is the relation between size of a clade, rate of  
polyploidy and number of chromosomes?

The authors have added a sentence into the discussion to reinforce the take home that while inclusion of this parameter is important and we see difference when running the model with or without this parameter, it does not alter the fact that the credible interval of our statistic still overlaps zero.

*The striking differences that we see in rate estimates under our two models is a clear example of the importance of evaluating the impact of polyploidy. However, we note that in our analysis the credible interval of our test statistic overlapped zero using both approaches. This suggests that the inclusion or exclusion of polyploidy in this particular analysis has no impact on our interpretation of the results (Figure 2).*

Lines XXX

Lines 188: “Even a small number of polyploidy events  
[…] could lead to much higher variance […]” – how small and how much higher?

This is a general statement. How small and how much will depend on the specific phylogeny that is being analyzed for the starting chromosome number and the rate of other types of chromosomal mutations. It is hard to make a more specific statement in this context.

Furthermore, as multiple alternative explanations are provided in the Discussion, how difficult  
would it be to incorporate them into the current model? Since the authors offer “a way  
forward” (line 223 how close are we to evaluating other factors discussed? Are the data there?

The software that we are using was created with this goal in mind and for binary traits like the one studied here we can move forward (we have multiple projects in the lab doing just this). However, for some traits that are truly continuous traits such as population size the answer is a bit more difficult; it is possible to discretize a continuous variables and use existing software lacking this we need new methods (which we are also working to develop).

4,393 are many species, but not that many among insects. How well are different clades  
represented? Is the distribution of sampled species relatively even across the  
phylogeny? Just something to clarify.

The number of species for which genetic and cytogenetic data are available generally show some correlation with number of extant species.

*The number of orders with sufficient data is small (10) in comparison to the size of this group (24). However, this is largely due to insufficient data present for the remaining orders.*

Lines XXX.

line 101: what is the extent of difference between the two phylogenies? “Some clades”  
is vague.

We understand that this wording does vaguely describe the differences that could be seen among these two sets of trees. Therefore, the authors have updated the manuscript to include the clarification on how the trees differ.

The difference between these two trees is in the age estimates in older nodes. The Misof tree favors more recent branching events while the Rainford tree ages the nodes much later.

Lines XXX.

Line 132-144: when discussing “intermdiate”, “lower” and so on rates, why not report  
the means and/or confidence intervals of those?

To maintain the ease with which the manuscript can be read we are maintaining this verbiage in the text. We believe that it is the relative comparisons that are most important in evaluating the results. However, these values should certainly be easy for the reader to access if they want to dig into it. In that vein beside plotting them in figures 2 and 3 we have added a supplemental table that reports all of the credible intervals for each of the analyses discussed in this section and plotted in these two figures

*Insert Table here*

Line 133: why is 20 the cutoff?

There is no generally accepted rule as to the number of taxa required to fit this type of model we chose 20 because past experience has shown that trying to estimate rates for clades smaller than this will often though not always lead to extremely high variance in rate estimates that are largely uninformative.

line 143 takeN

Corrected

line 236: a 100 trees is not many for a posterior. How much variability Is there among  
them?

In response to this and a comment from another reviewer we have performed a bootstrapping study that is mentioned briefly in the body of the manuscript and described fully in the supplemental materials.

Lines XXX, XXXX, XXXX

Lines 260-265: if high rates are not “biologically realistic” (BTW – is there a reference for  
this assumption?), why not limit the uniform prior to low rate values?

An exponential prior has the advantage of still allowing much higher rates if the data strongly suggest that higher rates are favored (it simply puts more probability on lower rates). This is favorable because it does not require the user to pick a cutoff value and decide what an acceptable maximum rate will be. This is also the standard approach that is implemented in analyses with diversitree and chromePlus.

With regard to biologically realistic, when Markov models are applied to discrete characters one way that model fitting can fail is when the optimization algorithm begins using very high rates. Effectively allowing for 1000s of state changes on every branch of the tree. In these regions of parameter space there is a ridge in the likelihood surface that corresponds to sets of rates that are proportional to tip states. Here is a hypothetical that might make this more concrete. Imagine we have a two state (A,B) with equal rates of transition (rate = 0.1). The extant species are observed 60% state A and 40% state B. There will normally be a global optimum where both of the rates are approximately 0.1. However, at very high rates say in the tens of thousands any pairing of rates where the transition rate into A is 1.5 times the transition rate into B will lie on a likelihood ridge that forms a local optimum. Though this behavior is well known in the field and often taught we are not familiar with any publication that discusses this behavior in detail.

Line 291: perhaps should be: “…statistic, where …”

Corrected

Line 293 – monocentric clades evolve slower?

Corrected

The authors should be commended for reproducibility of the analysis. However, everything  
hinges on the R package chromePlus. Yet its repository states “This package is in the early stages of development and should not be used for any analysis at this point.” Not encouraging!

Thank you for pointing this out! The readme for the GitHub repository had not been updated despite publication of the package last year. The package readme file has now been updated to:

If you have questions or problems please let me know [coleoguy@gmail.com](mailto:coleoguy@gmail.com).

chromePlus should be cited as:

Blackmon, H., Justison, J., Mayrose, I. and Goldberg, E.E., 2019. Meiotic drive shapes rates of karyotype evolution in mammals. Evolution, 73(3), pp.511-523.

In terms of methods and data, the paper relies heavily on previous work by the same authors  
(references 5, 6). This is fine, but especially when it comes to the Markov model, a little more  
information on the guiding principles behind the model would be helpful to the reader.  
Incidentally, the website for the R package containing the statistical model states “This  
package is in the early stages of development and should not be used for any analysis  
at this point.”, which is unhelpful for anyone interested in further developments suggested at  
the end of the Discussion.

The authors have added clarification to the Markov model that is being used with the addition of Supplemental Figure 2. In addition, as stated above the readme file has been updated to reflect the publication of the package and confidence in future use.

1. Blackmon, H.; Justison, J.; Mayrose, I.; Goldberg, E.E. Meiotic drive shapes rates of karyotype evolution in mammals. *Evolution* **2019**, *73*, 511-523.

2. Garagna, S.; Broccoli, D.; Redi, C.A.; Searle, J.B.; Cooke, H.J.; Capanna, E. Robertsonian metacentrics of the house mouse lose telomeric sequences but retain some minor satellite DNA in the pericentromeric area. *Chromosoma* **1995**, *103*, 685-692.

3. Freyman, W.A.; Höhna, S. Cladogenetic and anagenetic models of chromosome number evolution: a Bayesian model averaging approach. *Systematic Biology* **2018**, *67*, 195-215.

4. Glick, L.; Mayrose, I. ChromEvol: assessing the pattern of chromosome number evolution and the inference of polyploidy along a phylogeny. *Molecular Biology and Evolution* **2014**, *31*, 1914-1922.

5. Mayrose, I.; Barker, M.S.; Otto, S.P. Probabilistic models of chromosome number evolution and the inference of polyploidy. *Systematic biology* **2010**, *59*, 132-144.

6. Zenil‐Ferguson, R.; Burleigh, J.G.; Ponciano, J.M. chromploid: An R package for chromosome number evolution across the plant tree of life. *Applications in plant sciences* **2018**, *6*.