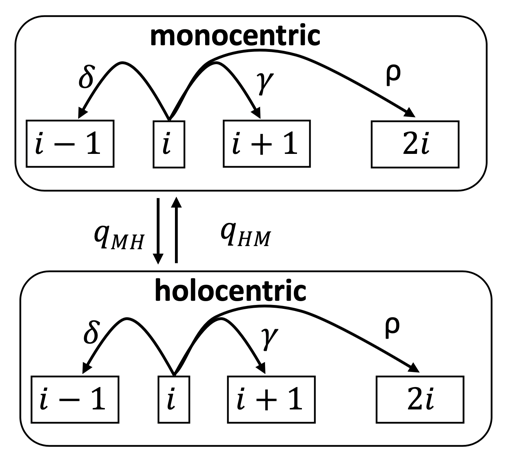
**Supplement**

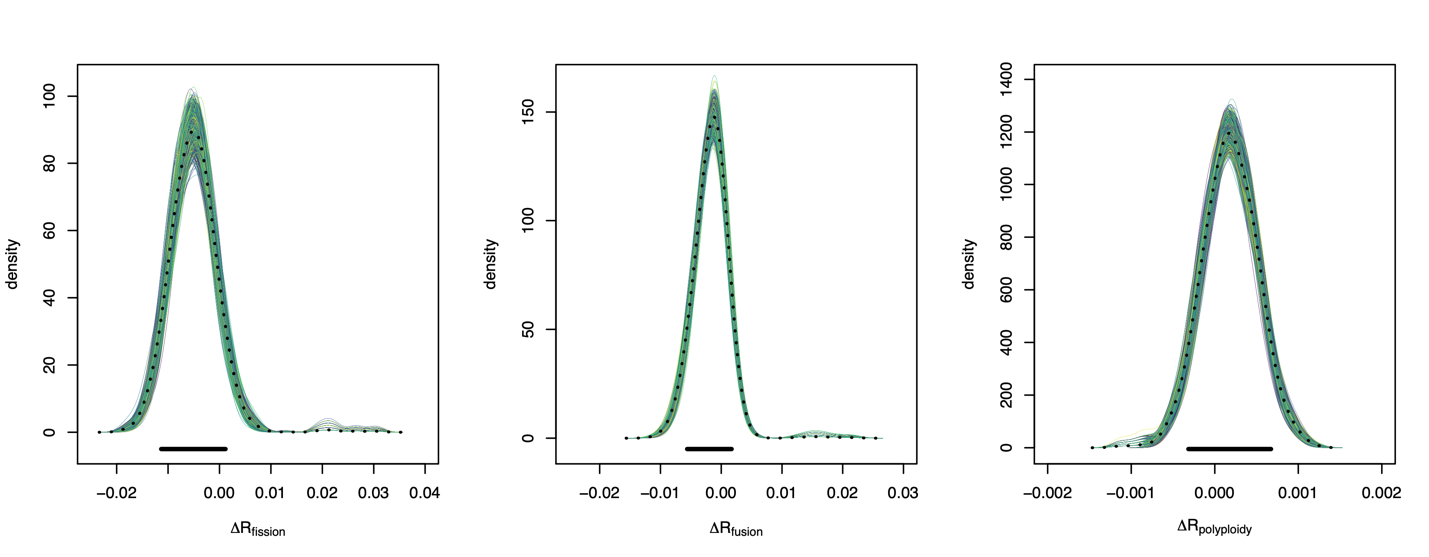


**Supplemental Figure 1. Comparison of inferences under alternative backbones**. In each plot we show the statistic for the three parameters of interest in our model. We find that regardless of the backbone phylogeny the resulting statistic has a largely similar distribution. Black lines represent the statistic estimate using the Misof backbone while red lines represent the statistic estimate using the Rainford backbone.



Supplemental Figure 2. Model for the evolution of chromosome number in monocentric and holocentric lineages. At an instance in time a lineage will have chromosomes 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).

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.



**A**

**C**

**B**

**Supplemental Figure 3. Comparison of bootstrap and empirical estimates**. In each plot we show the statistic for one of the parameters of interest in our model A) fissions, B) fusions, and C) polyploidy. 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.