**Title:** Not all centromeres are equal or are they?

**Authors:** Sarah N. Ruckman12+, Michelle Jonika13+, and Heath Blackmon123

1Department of Biology, Texas A&M University, College Station, Texas, USA

2Ecology and Evolutionary Biology Interdisciplinary Program, Texas A&M University, College Station, Texas, USA

3Genetics Interdisciplinary Program, Texas A&M University, College Station, Texas, USA

+Authors contributed equally and are co-first authors order was decided based on an R script simulating a coin toss

**Keywords:** Fusions, fissions, chromosome evolution, holocentric, monocentric

**Abstract**

Chromosome number is a central aspect of genome organization. Eukaryotes show a wide distribution of both type and number of chromosomes within their genomes. However, this diversity in chromosomes has made finding generalizations that explain this variation to be quite difficult. Holocentric chromosomes have centromeres that are diffuse and spindle fibers attach along the entire length of the chromosome, while monocentric chromosomes have a single, localized centromere. This difference in arrangement has led to the hypothesis that species with holocentric chromosomes can tolerate higher rates of fusions and fissions as compared to monocentric chromosomes, which may generate chromosomal fragments lacking centromeres. To test for differences in the rates of fusions and fissions, we analyzed data from 12,412 species of insects in both a taxonomic and phylogenetic framework. We found XXX.

**Introduction**

Chromosome number is a fundamental aspect of genome organization and is available for tens of thousands of species (Blackmon *et al.* 2017). Eukaryotes show a wide diversity both in the type and number of chromosomes within their genome. The evolution of chromosome number has been recalcitrant to the formation of rules or generalizations that can explain variation in patterns and rates across large clades. What is clear is that within clades fusions and fissions are two of the dominant forces in reshaping karyotypes (Lucek 2018). We use these terms for simplicity to describe single chromosome number changes. However, in reality fusions decreasing chromosome number capture two different processes at the molecular level. First, Robertsonian translocations followed by the loss of nonessential DNA can decrease chromosome number, and second, fusion of telomeres from two chromosomes followed by inactivation of one of the centromeres (Miga 2017). In contrast, fission increasing chromosome number can happen in just the way we might imagine, through fissions in the centromere region and the gaining of new telomeric sequences (Moretti and Sabato 1984; Garagna *et al.* 1995).

Fissions and fusions are assumed to be deleterious or underdominant and as such they should only fix in a population if there is low effective population size. However, centromeric structure may modulate the fitness effect of fusions and fissions. Since holocentric centromeres are diffuse and spindle fibers attach along the entire length of the chromosome it has been hypothesized that species with this type of centromere should have little difficulty segregating chromosomes that have experienced fusions or fissions (Malheiros-Garde and Gardé 1950; Greilhuber 1995; Luceño and Guerra 1996). In contrast, species with monocentric chromosomes have a single, localized centromere and chromosomal fragments generated from fusions or fissions may lack centromeres. These chromosomal fragments will not be able to segregate normally and will thus be deleterious. Therefore, holocentricity has potential to reduce or eliminate selective pressure against and underdominance of chromosome rearrangements. This could allow for a higher rate of fixation (Escudero *et al.* 2012).

In this study, we used all available chromosome data for insects to test whether holocentric chromosomes have a higher rate of fusions and fissions. Our hypothesis is that clades with holocentric chromosomes will tolerate fusions because the centromere is diffuse across the entire length of the chromosome, therefore each chromosome fragment will be more likely to properly segregate during meiosis. We found that XXX.

**Methods**

We downloaded all available chromosome data for insects from a prior study (Blackmon *et al.* 2017). This dataset is composed of 12,412 species comprising 376 families and 3,872 genera. The minimum haploid chromosome number is 2 while the maximum chromosome number is 141. There are 3,465 species with holocentric chromosomes and 8,946 species with monocentric chromosomes. For this dataset we collected the haploid chromosome number for each of the species. We additionally have obtained trees from a previous study that can be used for comparative analyses (Church *et al.* 2019). We have downloaded two sets of phylogenetic trees based on different backbone trees that will be used for our comparative analysis. Each phylogeny includes 1,726 genera and contains a sample of 100 trees from a posterior distribution. Both of these distributions of trees are to the genera-level and matching this data to our trait dataset we have an overlap of 602 tips. We are fitting our model on each tree from the posterior distribution and we randomly sample trait data when more than one species is available for a genus. This approach allows us to account for uncertainty in phylogeny and tip states.

Using the trait data and the posterior distribution trees, we implemented a chromosome number evolution model using chromePlus (Blackmon *et al.* 2019). This model allows us to determine if the rate of chromosome number evolution is significantly different in clades with holocentric and monocentric chromosomes. We obtained estimates of six parameters: rates of chromosome number increase, fissions, (γ1 and γ2), rates of chromosome number decrease, fusions, (δ1 and δ2), and rates of change in karyotype state, monocentric vs. holocentric (*q*12 and *q*21). We then used an uninformative, unbounded improper prior that assumed that all non-negative values are equally likely for all of the parameters. The Markov Chain Monte Carlo (MCMC) was initialized with parameter values drawn from a uniform distribution from 0 to 8, which is broad but biologically reasonable. Preliminary analysis indicated that MCMC chains reached convergence, however some were sampling non-biologically relevant regions of parameter space. To fix this problem, we added a prior that drew from an exponential distribution with a shape parameter of 0.5. This prior tightened our sampled parameter space and ensured that values that were outside of a biologically relevant region were penalized. We repeated the MCMC with all 100 trees at 50 generations each. We removed the first ten samples as our burnin for each run. We then calculated the Δ*r* statistic or the mean rate difference. For each postburnin sample we calculated Δ*r* as

.

**Results**

**Discussion**

Due to the size of the transition rate matrix sampling from the posterior and even simple calculations of the likelihood for a given parameter set is computationally expensive. As a result, we have limited our MCMC chains to 50 generations. However, we have 100 tree replicates at 50 generations each which increases our sample size and as a result are confident in our results.

**References**

**Weird you have refs in text but nothing is showing up down here for me.**