BROAD-SCALE STRUCTURAL EVOLUTION IN INVERTEBRATE GENOMES AND THE POPULATION GENOMICS OF JEWEL SCARABS IN THE SOUTHWESTERN US

A Dissertation

by

TERRENCE PRADAKSHANA SYLVESTER

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Heath Blackmon
Committee Members,	Alan Pepper
	Charles Criscione
	Claudio Casola
Head of Department,	Alex Keene

December 2022

Major Subject: Biology

ABSTRACT

Genome structure, at a fundamental level, can be described by the division of the genome into a discrete number of chromosomes and further divided into autosomes and sex chromosomes. An array of mechanisms or selection pressures can lead to changes in both of these genome characteristics. Meiotic drive, segregation mechanisms, sexual antagonism, epistasis, benefits of higher or lower recombination, and drift have all been invoked to explain changes in the number of chromosomes and the proportion of the genome that is sex-linked through sex chromosomes. Despite over a century of work, this level of genome organization has been resistant to broad generalizations that can explain the striking variation we observe among species. I use comparative phylogenetic approaches to determine the degree to which rates of chromosome number and sex chromosome system evolution vary among orders of insects. I also use these approaches to infer whether mutations that have led to divergence in chromosome number are deleterious, neutral, or beneficial.

The second part of my dissertation is focused on population genomics. As climate changes, many species develop discontinuous distributions. When a species is separated into many isolated demes, the risk of local extirpation increases. *Chrysina gloriosa* is a jewel scarab restricted to high elevations in west Texas, southern New Mexico, southern Arizona, and northern Mexico, where it feeds on several species of trees in the genus *Juniperus*. This beetle is highly sought after by collectors and is one of the most charismatic insects in North America. Despite this, there is currently no population genetic data that would allow for estimates of the health or resiliency of populations. Using population genomic data, I determine the degree of gene flow

among populations of the scarab jewel beetle *C. gloriosa* across the southwestern United States and determine the landscape characteristics that best predict the isolation of demes.

DEDICATION

This dissertation is dedicated to Kasuni Daundasekara, my beloved wife, whose support throughout the years was immense in achieving this final goal.

ACKNOWLEDGEMENTS

I would like to thank my committee chair Dr. Blackmon and my committee members, Dr. Pepper, Dr. Criscione and Dr. Casola for their guidance and support throughout the course of this research.

Thanks also go to my friends and colleagues and the department faculty and staff for making my time at Texas A&M University a great experience.

Finally, thanks to my mother and father for their encouragement and to my wife for her patience and love.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Dr. Blackmon [advisor] and Dr. Pepper [Home Department], Dr. Criscione [Home Department] and Dr. Casola of the Department of Ecology and Conservation Biology.

The data analyzed for Chapters 2 and 4 was provided by Dr Blackmon and Dr Johnston of the Department of Entomology and Dr Hjelmen at Utah Valley University. The analyses depicted in Chapters 3 and 4 were conducted partly by Chandler Kassel and Zach Hoover of the Department of Biology. Michelle Jonika, James Alfieri, L.T. Blackmon, Carl Hjelmen, assisted in fieldwork for Chapter 4.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by Graduate Teaching Assistantships and Startup funds from the Department of Biology, T3 Triad grant from Texas A&M University. This work was also made possible in part by the NIH under Grant Number R35GM138098. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

1. INTRODUCTION

1.1. Broadscale structural evolution of the genome

At a fundamental level, genome structure can be described by the division of the genome into a discrete number and the types of chromosomes a genome is divided among (e.g., autosomes, sex chromosomes, microchromosomes). This division of the genome into chromosomes is one of the first pieces of information we gained about genomes which predates even the Boveri-Sutton chromosome theory of heredity by two decades (Boveri, 1904; Flemming, 1882, 1887; Sutton, 1902). Since then, we have accumulated over 185000 karyotypic records, which includes insects, amphibians, fish and plants (Blackmon & Demuth, 2015a; Blackmon et al., 2017; Degrandi et al., 2020; Morelli et al., 2022; Nagpure et al., 2016; Perkins et al., 2019; Rice et al., 2015; Sylvester et al., 2020). Although we have accumulated data on chromosome numbers for more than a century, we lack a proper understanding of the evolution of this trait.

We think that the lack of information about the evolution of chromosome number could stem from the absence of biologically realistic models that describes chromosome number evolution and the absence of large-scale phylogenies which would allow us to make broad generalizations on karyotype evolution across the tree of life. However, recent developments on both of these fronts enable a variety of new analyses. For example, we see the development of the first model of chromosome number evolution, which uses a likelihood-based approach in 2010. This chromosome number evolution model is known as the ChromEvol model (Glick & Mayrose, 2014; Mayrose et al., 2010). Until then, most studies evaluating chromosome number evolution used a parsimony-based approach or failed to correct for phylogenetic history. After the

emergence of the ChromEvol model, we see the development of many chromosome number evolution models addressing various questions related to the evolution of chromosome number. Recent models like ChromePlus allow for rate variation across a phylogeny and can embrace a fully Bayesian approach (Blackmon et al., 2019).

Apart from the development of models that can be used to infer the evolution of chromosome numbers, we also see an influx of genetic data for many species aiding the reconstruction of phylogenetic relationships that capture a broad array of taxa. For instance, we now have phylogenetic information for 7238 amphibians, 5911 mammals, 9993 birds, 36101 plants and 69000 insects (Chesters, 2020; Janssens et al., 2020; W. Jetz et al., 2012; Walter Jetz & Pyron, 2018; Upham et al., 2019). With this information, we are now beginning to understand what drives chromosome number evolution across large clades. For instance, in mammals, female meiotic drive leads to rapid remodeling of karyotypes in many clades, and in Hymenoptera, transitions to eusociality have driven increases in chromosome number (Blackmon et al., 2019; Ross et al., 2015). Although we are beginning to understand these drivers of chromosome number evolution, we still lack a synthesis across the tree of life that would allow us to recognize general rules that are consistent across clades.

In this study we use data from two clades, Polyenoptera and Amphibia, to understand the dynamics of chromosome number evolution. Each dataset aims to answer a specific set of questions about structural evolution of genomes. We use Polyneoptera dataset to address three questions. First, to what degree are differences in variation in chromosome number among orders due to differences in the mode or tempo of evolution versus phylogenetic history. Second, how

do transitions to asexuality impact rates of chromosome evolution. Third, does repeat content predict rates of chromosome evolution. We leverage data from Amphibia to infer whether mutations that lead to changes in chromosome number are on average deleterious, neutral, or beneficial.

1.2. Genome assembly and Population genetics of *Chrysina gloriosa*

Sky Islands are similar to islands that occur in water, but it is a special type of inland terrain which consists of a sequence of high elevation mountains (Figure 1.1). There are at least 20 Sky Island complexes around the world with some examples being Drakensberg mountains in South Africa, Pentapui mountains in South America and Western Ghats mountains in India (Knowles et al., 2009; Mathien, 1995). Three Sky Island complexes have been identified in the United States, one of which is located in the Southwestern region of Arizona (Mathien, 1995). The Sky Islands of Arizona are a set of mountains largely contained in Coronado National Forest. This complex of mountains consists of 17 ranges including Huachuca, Patagonia, Peloncillo, and Santa Rita mountains. In addition to these widely recognized sky islands several additional mountain ranges are widely spread across the southwest with the Chisos, Davis, and Guadalupe ranges in Texas and the Animas, Big Hatchet, and Organ ranges in New Mexico. In each of these mountain ranges, as elevation increases the vegetation changes from either Sonoran or Chihuahua desert to Madrean forest with a mix of juniper and oak. In the higher and more northern ranges the Madrean forest transitions to one dominated by pine (DeBano, 1995). The intervening desert can act as a potential barrier to migration of individuals among these isolated mountain ranges. Due to lack of gene flow, populations can develop high degrees of genetic differentiation (Bech et al., 2009; Sekar & Karanth, 2013).

The jewel beetle *Chrysina gloriosa* is one of the four species in the genus *Chrysina* with a range that extends into the United States (Hawks, 2002; Young, 1957). *C. gloriosa* was described in 1854 by John L. LeConte (LeConte, 1854). The type locality of the initially described species is uncertain as the specimen is labeled simply "Copper mines and at Camp No 6" (LeConte, 1854). At present, *C. gloriosa*'s range includes Texas, New Mexico, Arizona, and northern Mexico (Morón, 1990). However, they are restricted to high elevation regions in sky islands due to their dependence on *Juniperus sp.* as an adult food source (Morón, 1990; Young, 1957). *C. gloriosa* has an annual life cycle with adults pupating in the soil and emerging during the monsoon rain season in late July and early August (Morón, 1990). The elevation that *C. gloriosa* occurs at is dictated by the distribution of *Juniperus* and so in more northern ranges like the White mountains *C. gloriosa* occurs between 1200-1500 meters while in more southern ranges like Pinaleno mountains it occurs at elevations from 1800-2100. The center of diversity for the genus *Chrysina* is central American and the four species, including *C. gloriosa*, are likely only a sample of the species that once occurred throughout the southwest during the last ice age (Young, 1957). Each of the species now present have had to adapt to conditions that are drier and cooler than those that they experience in regions to the south in the Sierra Madre Occidentalis range in Mexico. This history sets up the species of *Chrysina* in the southwest as a natural experiment in which multiple populations have been evolving in relative isolation for the last 10-20k years. These species are also highly charismatic and as such are targets of amateur collectors, and thus are of significant conservation concern (A. V. Evans, 2010).

Despite the popularity of jewel beetles with both scientists and amateur collectors, to date there have been no population genetic analyses of any of the U.S. species. Therefore, we cannot answer many basic questions about this species. Are populations present in one mountain range truly isolated from adjacent mountain ranges? Why is this species found only in a handful of the apparently suitable canyons within some mountain ranges? Should any populations of this species be protected? The proposed research will allow us to begin to answer these questions. Furthermore, this study will act as a baseline for future studies allowing us to assess the impact of continuing climate change and human encroachment on the habitat of this species.

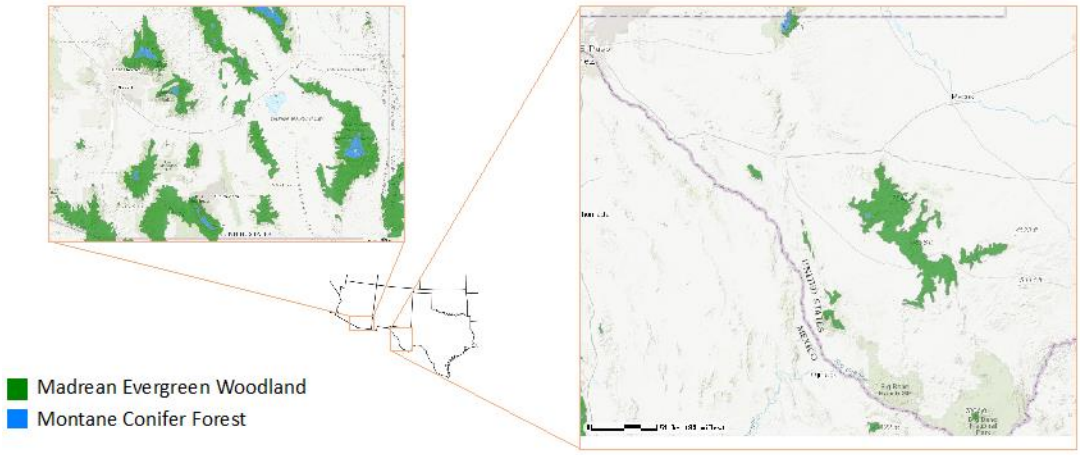


Figure 1.1 Map of Sky Islands spanning Arizona, New Mexico, and Texas. Areas with vegetation that can support *Chrysina* species are shown in green (Madrean forest) and blue (conifer forest).

2. LINEAGE-SPECIFIC PATTERNS OF CHROMOSOME EVOLUTION ARE THE RULE NOT THE EXCEPTION IN POLYNEOPTERA INSECTS[*]

2.1. Overview

The structure of a genome can be described at its simplest by the number of chromosomes and the sex chromosome system it contains. Despite over a century of study, the evolution of genome structure on this scale remains recalcitrant to broad generalizations that can be applied across clades. To address this issue, we have assembled a dataset of 823 karyotypes from the insect group Polyneoptera. This group contains orders with a range of variations in chromosome number and offer the opportunity to explore the possible causes of these differences. We have analyzed these data using both phylogenetic and taxonomic approaches. Our analysis allows us to assess the importance of rates of evolution, phylogenetic history, sex chromosome systems, parthenogenesis, and genome size on variation in chromosome number within clades. We find that fusions play a key role in the origin of new sex chromosomes, and that orders exhibit striking differences in rates of fusions, fissions, and polyploidy. Our results suggest that the difficulty in finding consistent rules that govern evolution at this scale may be due to the presence of many interacting forces that can lead to variation among groups.

2.2. Introduction

Chromosome number is one of the fundamental characteristics of a genome. It is also the first information collected about most genomes. In fact, the first chromosome counts were recorded

---

[*] Sylvester, T., Hjelmen, C. E., Hanrahan, S. J., Lenhart, P. A., Johnston, J. S., & Blackmon, H. (2020). Lineage-specific patterns of chromosome evolution are the rule not the exception in Polyneoptera insects. *Proceedings of the Royal Society B*, *287*(1935), 20201388.

prior to the development of the chromosome theory of inheritance (Flemming, 1882). Despite this early start, consistent rules governing the evolution of chromosome number across large clades remain elusive.

Changes in chromosome number can happen due to several mechanisms. We use the term fusion and fission to describe a decrease or an increase of one in chromosome number, respectively. However, these terms are simplifications and may represent multiple processes at the molecular level. Reduction in chromosome number can happen through Robertsonian translocations with the loss of non-essential DNA (Garagna *et al.*, 1995) or happen through the fusion of two chromosomes at the telomeres followed by loss of one of the centromeres (Gordon et al., 2011; Miga, 2017). By contrast, increases in chromosome number can occur due to simple chromosome fission in the centromere region (Moretti & Sabato, 1984) or due to the duplication of an entire chromosome. Changes in chromosome number of more than one can also occur. Although rare in most animal groups, demiploidy describes an increase chromosome number by one-half. Demiploidy events can occur by the joining of haploid gamete with an unreduced diploid gamete (Hornsey, 1973). Finally, whole-genome duplication can lead to a doubling of chromosome number (Beçak et al., 1970).

These changes in chromosome number can have broad impacts on gene transcription, recombination rates and sex chromosome evolution. The presence of an extra copy of a chromosome can lead to both increases or decreases in gene transcription (Lockstone et al., 2007; Sun et al., 2013; Williams et al., 2008). It has long been recognized that chromosome number should positively correlate with genome-wide recombination rates (Stebbins, 1958). The

frequency of recombination events and the proper segregation of chromosomes into gametes is dependent on crossing over in meiosis. The lower limit of the number of crossing over events is controlled by the number of chromosome arms in most species and by the number of chromosomes in some species (Dumont, 2017). This relationship between chromosome number and recombination has been suggested as a source of indirect selection on chromosome number in Hymenoptera though, recent analysis suggest this may be only a weak force (Ross et al., 2015; Sherman, 1979). Changes in chromosome number may also impact the evolution and behavior of sex chromosomes. For instance, if chromosomes are broken into smaller chromosomes while keeping all else equal (e.g., genome size), the average chromosome size should be negatively correlated with the number of chromosomes. This can have important impacts on the fate of sex chromosomes. A comparative study of Coleoptera has shown that species are more likely to lose the Y chromosome and transition from XY to XO if they have many small chromosomes rather than a few larger chromosomes (Blackmon & Demuth, 2015b).

In sexual species, it is often assumed that changes in chromosome number are under dominant–heterozygotes have reduced fitness (White, 1954). Chromosomal heterozygosity occurs when the chromosome complement from the parents differs (e.g., if one parent contributes a fused chromosome). Perhaps the most widely known example of this is hybridization between horses and donkeys where the offspring carries 32 chromosomes from the mother and 31 chromosomes from the father and is sterile (Wodsedalek, 1916). However, in wild mice which are heterozygous for a single fusion between chromosomes 16 and 17, there is no significant reduction in fertility and thus no reduction in fitness (Britton-Davidian *et al.*, 1990). A large number of crosses in lemurs (where the chromosome number ranged from 44 to 60) exhibit a full

range of fitness effects in crosses with parents that have different chromosome numbers (Ratomponirina *et al.,* 1988). By contrast, one can hypothesize that changes in chromosome number might be less deleterious in asexual species since they do not have to pair with any other genome in the population. Consistent with this, many asexual species have considerable variation in chromosome number (Lachowska & Holecova, 1998).

To better understand the dynamics of chromosome evolution, we have chosen to work with the insect clade Polyneoptera, which includes the orders Blattodea (roaches and termites), Dermaptera, Embiidina, Mantodea, Notoptera, Orthoptera (grasshoppers), Phasmatodea and Plecoptera. Polyneoptera show striking differences in chromosome number variation among orders. One of the central goals of this work was to determine if these differences are due to idiosyncratic rates and patterns of evolution in each order, or due simply to differences in phylogenetic history. Polyneoptera also have variation in sex chromosome systems and, include sexual and asexual lineages allowing us to explore interactions between these characteristics and chromosome number. We assembled a trait dataset of chromosome number, sex chromosome system (SCS), genome size and reproductive mode. We analyzed these data in both a taxonomic and a phylogenetic framework to determine the impact of the sexual system on rates of chromosome number evolution, the source of transitions in SCSs, and identify differences in patterns of chromosome number evolution among orders.

2.3. Methods

2.3.1. Data collection and phylogenetic inference

We downloaded all available chromosome data for the insect clade Polyneoptera from the Tree of Sex database and supplemented this with extensive literature searches (Blackmon et al., 2017; Tree of Sex Consortium, 2014). All these data are available at www.karyotype.org. We also downloaded genome size data from the animal genome size database (Gregory, 2002) and supplemented this dataset with 60 new genome size estimates to yield a final dataset of 185 genome size estimates (Appendix A: Table 6.1).

Using PyPHLAWD and Phylota, we assembled two sequence datasets one for all Polyneoptera and one focused on the order Phasmatodea (Appendix A: Table 6.2) (Sanderson et al., 2008; Smith & Brown, 2018). Sequences were aligned and checked for quality using MAFFT v. 7 and Gblocks v. 0.91b, respectively (Castresana, 2000; Katoh et al., 2019). Rogue taxa were identified (Sup. Figure 2.3) and removed with Mesquite v. 3.51 based on preliminary trees built with RAxML v 8.2.10 (Maddison & Maddison, 2019; Miller et al., 2010; Stamatakis, 2014). Our final alignment for Polyneoptera contained 232 taxonomic units with 73% missing data, while the final alignment for the insect order Phasmatodea contained 41 taxonomic units with 57% missing data. We conducted two independent BEAST v. 2.5 runs of 100 million generations to infer time-calibrated phylogenies under a relaxed lognormal clock, a birth–death model, GTR + G as the nucleotide substitution model, and priors on the age of seven nodes (Sup. Table 2.2) (Bouckaert *et al.*, 2014). The initial 50% of each MCMC run was discarded based on evaluation with Tracer v. 1.7 (Rambaut *et al.*, 2018), and 50 phylogenetic trees were randomly sampled from the post-burnin period of each run. The 100 sampled trees form the posterior distribution

used for the analyses described below. This approach was repeated to build the phylogenies for Phasmatodea.

2.3.2. Modeling chromosome number evolution

We used the R package chromePlus to estimate rates of chromosome number evolution (Blackmon *et al.*, 2019). We tested two versions of our model, a simple model with chromosome fission and fusion and a complex model which included fission, fusion, and polyploidy. Although we use the terms fusion and fission for convenience, it should be noted that these are simply changes (decreases and increases, respectively) in the haploid number by one. Based on the likelihood ratio test results (discussed below), we used the complex model to estimate the rates of chromosome number evolution. To get reliable estimates for the rates of chromosome number evolution, we only analyzed the four orders with at least 20 representatives.

To account for uncertainty in chromosome number (e.g., when there were reports of multiple values for a tip in our phylogeny), we randomly sampled among the possible values and repeated for each tree. To account for uncertainty in phylogeny, we ran an MCMC of 1000 generations for each of the 100 trees in the posterior distribution. Inspection of the parameter estimates revealed that our MCMC runs converged by 50 generations in most cases. We conservatively discarded the initial 25% as burnin and randomly sampled 100 states from the post-burnin portion of the run. This process yielded 10000 point-estimates that define the posterior distribution of the parameters in our model. We tested for differences in rates of chromosome number evolution by comparing the 95% credible interval of the posterior distribution for each parameter in our model. Rates were inferred with branch lengths transformed to make trees unit length. However,

all rates reported have been back transformed so they represent transition rates in units of millions of years. As is customary for Markov models, the reported rates are lambda parameters for exponential distributions that describe the expected waiting time for a transition to occur. Since our tree has branch lengths in units of millions of years these are reported in units of per MY.

2.3.3. Genome size

We tested whether genome size as a proxy for repetitive content might explain variation in rates of chromosome number evolution. Expansions in genome size are largely due to repetitive content, especially transposable elements (Bennetzen et al., 2005; Kidwell, 2002). We reasoned that increased transposon activity could lead to a greater frequency of fusion and fission mutations and result in higher rates of chromosome number evolution in larger genomes. We also tested for a correlation between genome size and chromosome number, reasoning that recent whole-genome duplications should lead to an increase in both values.

We first tested whether genome size was predictive of chromosome number. For this analysis, we fit a linear model where genome size was the predictor variable and chromosome number was the response variable using all species with both chromosome number and genome size estimates (n = 55) (R Core Team, 2016). This was repeated for a reduced dataset using a phylogenetically corrected linear model including only those taxa present on our phylogeny (n = 23) (Ho & Ané, 2014). To test whether the genome size for a species predicted its rate of chromosome number evolution, we first calculated tip rates for all species on our phylogeny that also had genome size estimates (n = 20). This tip rate was calculated as the difference between

the tip value and the most probable chromosome number of the immediate ancestor of a given tip (see below for ancestral state reconstructions of chromosome number) divided by the branch length between the ancestor and the tip. We use the mean of this value calculated across the posterior distribution of trees. We evaluated both an absolute tip rate and a directional tip rate (accounting for whether the change is an increase or decrease in chromosome number). We then fit both standard and phylogenetically corrected linear models where genome size predicted either the absolute rate or the directional rate (Ho & Ané, 2014; R Core Team, 2016).

2.3.4. Ancestral state reconstructions

We estimated the ancestral states of chromosome number and the sex chromosome system (SCS). We estimated the ancestral states of the chromosome number at the root of each order using chromEvol v. 2.0. (Glick & Mayrose, 2014; Mayrose et al., 2010). We used a fixed parameter model which included chromosome gains, losses, and whole-genome duplication—matching the model used in chromePlus. For each tree from our posterior distribution, we took the mean of each parameter estimate from the corresponding chromePlus analysis described above and supplied these to infer ancestral states in chromEvol. We combined the estimates from the analysis of all trees in the posterior.

The estimate of ancestral states for SCSs was done using the ARD model in the function ACE in the R package APE (Paradis & Schliep, 2019). We classified multi-XY SCS as XY which resulted in two states (XO and XY). To estimate the number of transitions in SCSs, we created the same model and performed stochastic mappings in the R package phytools (Revell, 2012).

Data and all R code for analyses are provided in a GitHub repository: https://github.com/Tsylvester8/Polyneoptera.

2.4. Results

2.4.1. Evolution of sex chromosome systems

In our dataset, 23 genera (182 taxa) contain species with at least two types of SCSs (i.e., XO, XY, or multi-XY). In each of these genera, we compared the mean haploid autosome number for all species with a given SCS. By comparing these means within genera, we can determine if differences are consistent with fusions or fissions as a source of transitions among SCSs. Briefly, if transitions from XO to XY are generated by the fusion of an autosome to a sex chromosome, we would expect a lower mean autosome number for XY species. Likewise, if transitions from XO or XY to multi-XY are generated by the fusion of an autosome to a sex chromosome, we would expect a lower mean autosome number in multi-XY species. By contrast, if transitions from XY to multi-XY are generated by the fission of an existing sex chromosome, we would expect the mean autosome number to be unchanged in the multi-XY species. We find strong support for fusions as a source of transitions from XO to XY SCSs. Of the 17 genera with both XO and XY species, 94% (16/17) show a lower mean number of autosomes in XY species (Table 2.1). However, we find support for both fusions and fissions leading to transitions from XY to multi-XY. Of the 10 genera evaluated 40% (4/10) have higher or unchanged mean number of autosomes in multi-XY species (suggestive of fissions). By contrast, 60% (6/10) of the genera have a lower mean autosome number in the multi-XY species (suggestive of fusions).

Table 2.1 Chromosome number and sex chromosome systems. Within each genus, we report the mean number of autosomes for all species having XO, XY, or complex sex chromosome systems. The last two columns state whether the given chromosome numbers support fusion or fission as an important process in transition between these sex chromosome systems. Negative (–) symbol indicates a distribution of chromosome number that does not support either mechanism.

Order	Genus	Mean number of autosomes			XO to XY	XO or XY to multi-XY
		XO	XY	multi-XY		
Blattodea	*Cryptotermes* (6)		23	16.4		fusion
Dermaptera	*Forficula* (14)		11	10.8		fusion
	Nala (3)		17.5	17		fusion
	Nesogaster (2)	10		9		fusion
Mantodea	*Deiphobe* (2)	9		12		fission
Orthoptera	*Aleuas* (6)	9	9.2		-	
	Dichroplus (35)	10.74	8.71	9	fusion	fission
	Diponthus (7)	10.5	10		fusion	
	Eurotettix (5)		10	9		fusion
	Gryllotalpa (5)	9.67	5		fusion	
	Isophya (25)	15	14		fusion	
	Leiotettix (10)	11	8	5.5	fusion	fusion
	Scotussa (8)	10.6	8.5	9	fusion	fission
	Scyllina (3)	11	10		fusion	
	Tetrixocephalus (5)	11	10		fusion	
	Xyleus (9)	11	10		fusion	
	Zoniopoda (6)	11	10		fusion	
Phasmatodea	*Didymuria* (11)	17.6	14.67		fusion	
	Isagoras (3)	18	16		fusion	
	Leptynia (5)	18.25	17		fusion	
	Podacanthus (3)	17	13		fusion	
	Prisopus (2)	24	13		fusion	
Plecoptera	*Perla* (7)	9.5	4	11.75	fusion	fission

2.4.2. Ancestral states and rates of sex chromosome evolution

We find that the ancestral state for SCS in Polyneoptera clade was XO, with a probability of 90.3%. Similarly, the most probable ancestral state for each order was also XO, except for Isoptera and Dermaptera, where XY is more probable (Figure 2.1). We find the credible intervals

of the transition rates from XO to XY and XY to XO to be largely overlapping, with means of 0.00202 and 0.00200, respectively. However, transitions from XO to XY were more common (mean = 15.3), while transitions from XY to XO were relatively rare (mean = 6.7) (Sup. Table 2.4).

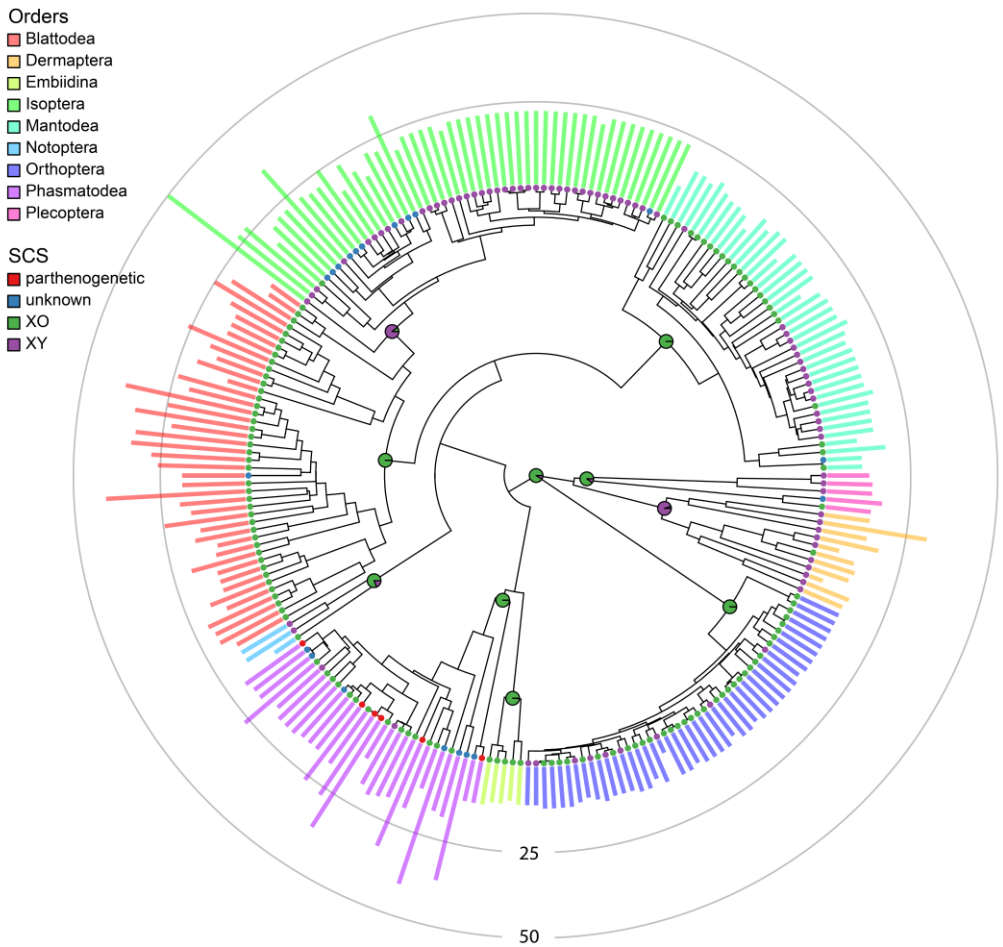


Figure 2.1 One of the 100 trees from the posterior distribution with chromosome number and sex chromosome system displayed. Bar heights represent the haploid chromosome number of each taxon. Tips are labelled according to the sex chromosome system/reproductive mode of the taxa. Tips that are marked as being XY includes species with XY and multi-XY sex chromosome systems. The pie charts at the roots of each order and at the root of the tree represent the probability of that node being either XO or XY, averaged across the posterior distribution of 100 phylogenies (in this analysis, we discarded the tips that are parthenogenetic and that did not have data for SCS). The rings represent the 25 and 50 chromosome number margins.

2.4.3. Rates of chromosome number evolution

Some Polyneoptera orders exhibit little variation in chromosome number while others are highly variable (Figure 2.1). This could be because some orders are evolving more quickly, or it could be because their phylogenetic history has allowed for a greater period of divergence. To draw any rigorous conclusions, we must explicitly control for this history. We began by applying a base model that includes only fusions and fissions and compared this via a likelihood ratio test with a model that included polyploidy. This was repeated for each of the 100 trees from our posterior distribution for each order. All orders showed some support for the model including polyploidy and overall, 77.6% of our likelihood ratio tests supported the more complex model that included polyploidy. For this reason, all analyses were done with the model with fusions, fissions, and polyploidy, allowing us to compare the same set of rates across all orders. In Blattodea (including Isoptera), we estimate a mean fusion rate of 0.128, a fission rate of 0.150 and a polyploidy rate of 0.003 (Sup. Table 2.3). By contrast, if we remove the subclade Isoptera from Blattodea we find that parameter estimates increase to 0.420, 0.385 and 0.004 for fusions, fissions, and polyploidy, respectively. This is consistent with rate estimates for Isoptera in isolation, where we infer rates of 0.044, 0.063 and 0.003 for fusions, fissions, and polyploidy, respectively. Mantodea rate estimates exhibited high uncertainty, overlapping rate estimates in most orders for most parameters (Figure 2.2, Sup. Figure 2.4). Some of the lowest rates we estimated were in Orthoptera where the fusion rate was 0.003 and the fission rate was 0.024. However, we do find the polyploidy rate to be relatively high with a mean rate of 0.101.

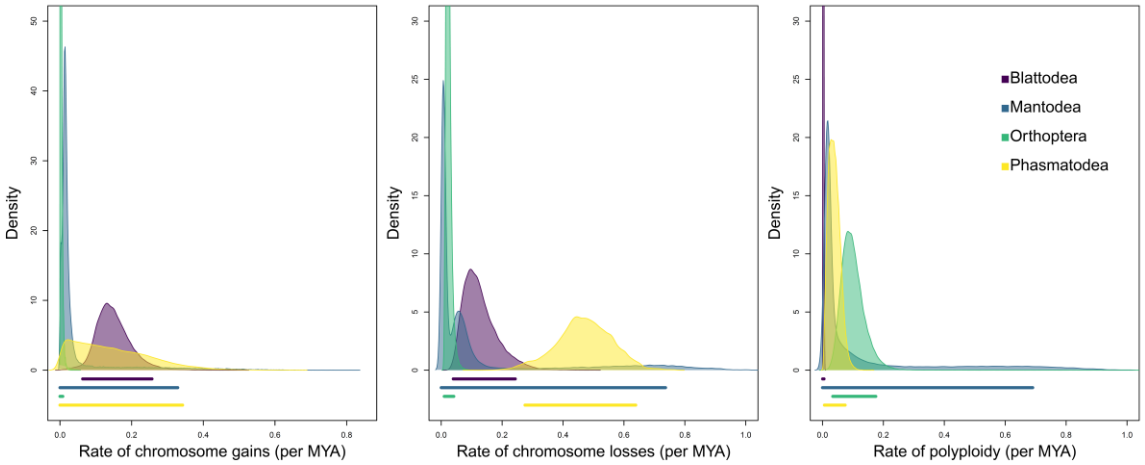


Figure 2.2 Rates of chromosome (a) fission, (b) fusion and (c) polyploidy of four orders in the insect clade Polyneoptera. Bars below each distribution indicates the 95% HPD interval. Orders are indicated by the fill color.

2.4.4. Chromosome number evolution versus genome size

We performed genome size estimation of 60 Polyneoptera species from the orders Blattodea, Mantodea, Orthoptera and Phasmatodea (Appendix A: Table 6.1). The largest of these was 18051.1 Mbp in *Hadrotettix trifasciatus* (Orthoptera), while the smallest was 2071 Mbp measured in *Thesprotia graminis* (Mantodea). These data, along with 125 publicly available genome size estimates, had an overlap of 55 species with our chromosome dataset and 23 species with our phylogenetic dataset.

We found a significant trend towards lower chromosome numbers with higher genome sizes (p-value = 0.01), but after correcting for phylogeny this was not significant (p-value = 0.75). This difference appears to be due to the largest genome sizes and some of the smallest chromosome numbers both occurring in Orthoptera (Sup. Figure 2.5A). We found no significant relationship

between directional tip rates and genome size (Sup. Figure 2.5B), or absolute tip rates (Sup. Figure 2.5C).

2.4.5. Asexuality and rates of chromosome number evolution

Our dataset contains 13 parthenogenetic Phasmatodea species. We tested whether the rates of chromosome number evolution are contingent on the reproductive mode (phylogeny and trait data is shown in Sup. Figure 2.7). We found that there is no significant difference in the rates of chromosome fusion and fission between sexually and asexually reproducing lineages (Sup. Figure 2.8A and Sup. Figure 2.8B). However, we find that rates of polyploidy are significantly higher in asexually reproducing lineages than in sexually reproducing lineages (Sup. Figure 2.8C).

2.5. Discussion

The evolution of chromosome number across large clades and long timespans is fundamental to the diversity of genomes we observe across the tree of life. Despite this, we are only beginning to understand how chromosome number evolves. In this study, we have focused on the dynamics of chromosome number evolution in Polyneoptera.

2.5.1. Sex chromosomes and chromosome number evolution

The transition between SCSs from XO to XY can occur through the fusion of the X chromosome and an autosome or by sex chromosome turnover with fixation of the ancestral X as an autosome (Charlesworth & Charlesworth, 1980; Vicoso & Bachtrog, 2015; White, 1954). Transition via fusion will lead to a reduction in the total number of autosomes, while turnover should lead to no

change in the total number of autosomes. Our data show a clear pattern of reduced autosome number in taxa with likely XO to XY transitions supporting fusions as a dominant pathway for this change (Table 2.1). Transition from XY to multi-XY can occur through the fusion of a sex chromosome and an autosome or by sex chromosome fission. Transition via fusion will lead to a reduction in the mean number of autosomes while fissions will lead to no change in the number of autosomes. Our data show a mixed pattern with multi-XY species having both increased and decreased autosome number (Table 2.1). We interpret this as evidence for both fusions and fissions as an important source of multi-XY systems. Combining our results for XO to XY and XY to multi-XY transitions, we find that fusions are a dominant route for changes in SCSs a pattern consistent with common sexual antagonistic variation on autosomes leading to selection for fusions (Charlesworth & Charlesworth, 1980; Kitano & Peichel, 2012). The recent development of a null model for fusions that facilitates testing for an excess of sex chromosome autosome fusions is an obvious next step that could be applied in groups like Orthoptera and Phasmatodea (Anderson et al., 2020).

Our genus-level analysis of autosome number and SCS is limited to only those genera with more than one SCS and thus omits much of our data. When autosome number among orders is parsed by SCS, differences in the mean autosome number suggest that in some groups the origin of transitions may differ from our genus-level analysis. For instance, the mean autosome number of all XY species in both Blattodea and Dermaptera is higher than the mean autosome number of XO species (Sup. Figure 2.9). This pattern is not expected if fusions are the primary source of transitions from XO to XY (Table 2.1). However, without modeling both SCS and chromosome number jointly on a phylogeny these patterns are at best difficult to interpret.

Our assembled data can also help us understand the fate of the Y chromosomes. It has been suggested that Y chromosomes may be destined to decay and loss given the inevitability of mutation accumulation and reductions in the area that undergoes recombination and allows for segregation (Blackmon & Brandvain, 2017; Steinemann & Steinemann, 2005). Alternatively, they may be retained through cycles of rejuvenation or even transitions into alternative forms of meiosis (Blackmon & Demuth, 2014, 2015b). In our inference of SCS evolution, the rate of Y chromosome gains and Y chromosome losses are both approximately 0.002 (Sup. Figure 2.8). However, we find that Y chromosome gains are more common with a mean of 15.3 across our entire phylogeny while losses are relatively rare with a mean of 6.7 across the entire phylogeny. This pattern is intuitive when we consider that the ancestor of this group was likely XO and thus there has been relatively little time for the gain of the Y chromosome to then be followed by its decay and loss.

2.5.2. Constraints on chromosome number evolution

In many clades, chromosome number is likely to change primarily through fusions and fissions of existing chromosomes (Blackmon et al., 2019; Sved et al., 2016). However, chromosome number could also change due to aneuploidy or whole-genome duplication events that fix in a population, creating duplicate copies of one or more chromosomes. In fact, a recent analysis of 28 transcriptomes from Polyneoptera species revealed evidence for at least four independent whole-genome duplication events and two independent partial genome duplication events (Z. Li et al., 2018). Parsing the relative contribution of fissions and aneuploidy to increases in chromosome number is not possible with our dataset, but could be tested with cross-species chromosome painting via fluorescence in situ hybridization (Liehr *et al.*, 2017). The converse

chromosome number decrease due to aneuploidy is likely to be exceedingly rare since all genes on the chromosome would have to be dispensable. However, these processes (fusion, fission, whole-genome duplication and aneuploidy) could all lead to sterile offspring if two populations (one with the chromosome duplication and one without) hybridize, since the heterozygous offspring may have difficulty segregating unmatched chromosomes during meiosis or gametes may carry an incomplete set of genes (White, 1978b). In both sexual and asexual species, chromosome increase due to aneuploidy may be rare due to the impact of aneuploidy on dosage which may lead to stoichiometric imbalances in gene networks. However, asexually reproducing species should be immune to the problem of proper segregation since they cannot outcross. For these reasons, we expected to see a higher rate of chromosome number increase and decrease in asexual species. Our Phasmatodea dataset has a mean of 9.3 transitions from sexual to asexual reproduction and offers a chance to test this hypothesis (Sup. Table 2.5). To our surprise, our analysis illustrates that rate of chromosome increase, and decrease are equal in sexual and asexual Phasmatodea (Sup. Figure 2.8). We interpret this as evidence that the constraints on chromosome number change via fusions, fissions and aneuploidy are largely similar in sexual and asexual Phasmatodea. The most parsimonious explanation seems to be that the changes observed are largely neutral and that individuals that are heterozygous for chromosomal rearrangements do not typically have difficulty properly segregating chromosomes during meiosis. By contrast, the constraints on polyploidy appear to be lifted in asexual lineages, suggesting these changes may be deleterious in sexual species but neutral or nearly neutral in asexual species.

2.5.3. Variation in rates of chromosome number evolution

Most studies of chromosome number evolution have been done on small clades in isolation (De Oliveira & Moraes, 2015; McCann et al., 2016; Rockman & Rowell, 2002). This creates a challenge in understanding variation in rates of chromosome number evolution across the tree of life since rates are fundamentally influenced by the time constraints and branch lengths inferred in a study (but see: Blackmon et al., 2019; Zenil-Ferguson et al., 2017). By inferring rates in four orders all using a common tree, we are able to make a more valid comparison among clades and determine whether some groups are evolving at significantly different rates. We found many examples of significantly different rates of chromosome number evolution among orders. Blattodea have a higher rate of fissions than Orthoptera (Figure 2.2a; Sup. Table 2.3). Blattodea (excluding Isoptera) have a higher rate of fusions than Isoptera and Orthoptera (Figure 2.2b; Sup. Table 2.3). Fusions are also higher in Phasmatodea than Blattodea, Isoptera and Orthoptera (Sup. Table 2.3). Polyploidy is higher in Orthoptera than Blattodea and, Phasmatodea is higher than Blattodea (Figure 2.2c; Sup. Table 2.3). In line with existing evidence (Blackmon et al., 2017; Lokki & Saura, 1979), polyploidy is higher in asexual than sexual species (Sup. Figure 2.8).

One possible explanation for variation in rates of chromosome number evolution is fundamental differences in the repeat content of the genome. For instance, large numbers or recent expansions of transposable elements may lead to more frequent chromosome breakage or other structural rearrangements that change chromosome number (Carbone *et al.*, 2014). If transposable elements have expanded in the genomes of a clade, we might expect to see a signature of this in increased genome sizes (Kidwell, 2002). However, we found no association between genome size and absolute rates of chromosome number evolution suggesting that repetitive content is not a

driving force in the stability of large-scale genome structure across Polyneoptera. We also investigated the relationship between chromosome number and genome size. In particular, we hypothesized that if recent polyploidy events were present, we should expect to find increases in both measures. Indeed, a linear model with chromosome number as the response variable and genome size as the predictor variable is significant. However, this pattern is in the opposite direction from what we would expect due to polyploidy (smaller chromosome numbers are found in larger genomes). However, this pattern is driven largely by the low chromosome number and large genome size in Orthoptera, and once corrected for phylogeny, we find no significant relationship between these variables. We interpret this result as evidence that our dataset lacks any very recent polyploidization. An additional expectation for recent polyploids would be that they would exhibit a large positive tip rate (large increases in chromosome number since the most recent common ancestor) and a large genome size. However, we do not find any significant relationship between tip rates and genome size with or without correction for phylogeny (Sup. Figure 2.5). Moving forward, the recent development of multiple probabilistic models of chromosome number evolution that allow for associations with speciation or binary characters offers a way forward to further tease apart the determinants of rates of chromosome number evolution (Blackmon et al., 2019; Freyman & Höhna, 2018; Zenil-Ferguson et al., 2018).

2.5.4. Reconciling a century of chromosome research

We find that many of our results confirm previous hypotheses on chromosome number evolution and SCS evolution in Polyneoptera. With respect to SCSs, Mantodea, Orthoptera and Phasmatodea had all previously been hypothesized to originate from XO ancestors and our results confirm these hypotheses (Blackman et al., 1995; Blackmon et al., 2017; Sally Hughes-

Schrader, 1959). Additionally, evidence that the X chromosome of the German cockroach, Blattella germanica (Blattodea), is homologous to the X chromosome in most Diptera is consistent with a shared XO SCS in the ancestor of all Polyneoptera (Meisel et al., 2019). However, other work has shown that the X chromosome in *Drosophila melanogaster* is not homologous to the Z in *Bombyx mori* or the X in *Tribolium castaneum*, or even the X in many flies, suggesting that insects may have frequent turnover in SCSs (Kitano & Peichel, 2012; Pease & Hahn, 2012). Even if early polyneopterans shared a common XO SCS, the high frequency of sex chromosome autosome fusions and Y chromosome losses that we document suggest that the gene content of sex chromosomes of extant species is likely variable.

In other cases, the application of probabilistic models to our expanded dataset challenges previous hypotheses. For instance, it has been hypothesized that the most recent common ancestor of Blattodea (including Isoptera) was XY, and the Y chromosome had been rapidly lost in Blattodea (excluding Isoptera). Our results support an XO ancestor with a probability of 99.04% (Bergamaschi et al., 2007; Luykx, 1990). In Isoptera, it has been hypothesized that the ancestral SCS was XO, but our results suggest that ancestral SCS of Isoptera was in fact XY (with a probability of 89.44%) (Blackman et al., 1995; Luykx, 1990).

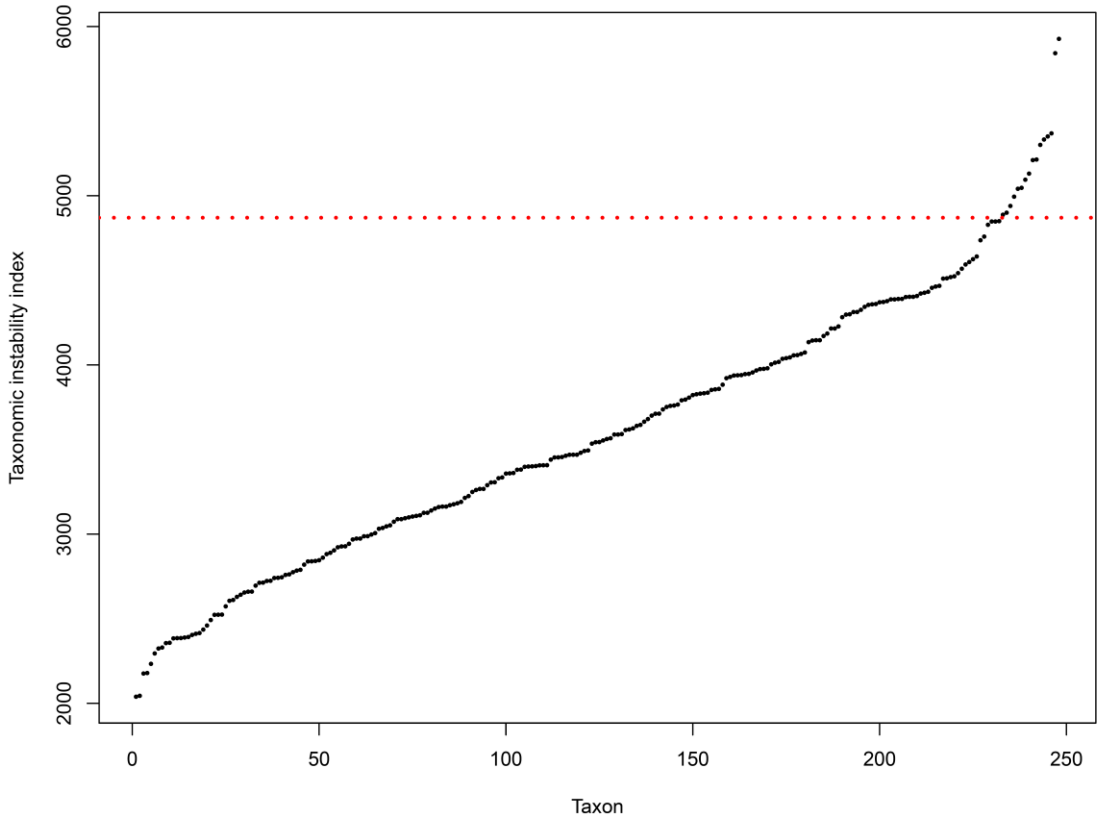
Even without model-based analyses, some authors have suggested that fusions or fissions were more important in some groups. For instance, in Isoptera, it has been previously hypothesized that fusions are more common than fissions (Bergamaschi et al., 2007; Luykx, 1990). Although we do not find a significant difference between the rates of fusion and rates of fission (Sup. Table 2.3), our ancestral state analysis for chromosome number finds that the average number of

fusion events are significantly higher than number of fission events (33.32 fusion events and 21.94 fission events, t-test p-value less than 0.05). Our results depart most strongly from previous work in estimates of ancestral states. For instance, in Mantodea, we inferred 8 and 7 as the most probable ancestral haploid number for the group, while previous work predicted ancestral haploid number of 14 (S. Hughes-Schrader, 1950). In Orthoptera, we inferred six as the most probable ancestral haploid number, while previous work predicted ancestral haploid number of 12 (White, 1954). Finally, in Phasmatodea, we inferred 9 and 10 as the most probable ancestral haploid number, while previous work predicted ancestral haploid number of 18 (Sup. Figure 2.6) (Sally Hughes-Schrader, 1959). We note however, that our ancestral state estimates are dependent on the model applied. In our study, we used a model that allowed for a possibility of polyploidy in all orders. This has the impact of increasing the probability of low ancestral state estimates that may not be realistic if growing genome evidence finds reduced support for whole-genome duplication events in these orders. However, we note that our preliminary analyses showed that our finding of differences in rates among orders is not impacted by the inclusion or exclusion of polyploidy in our model.

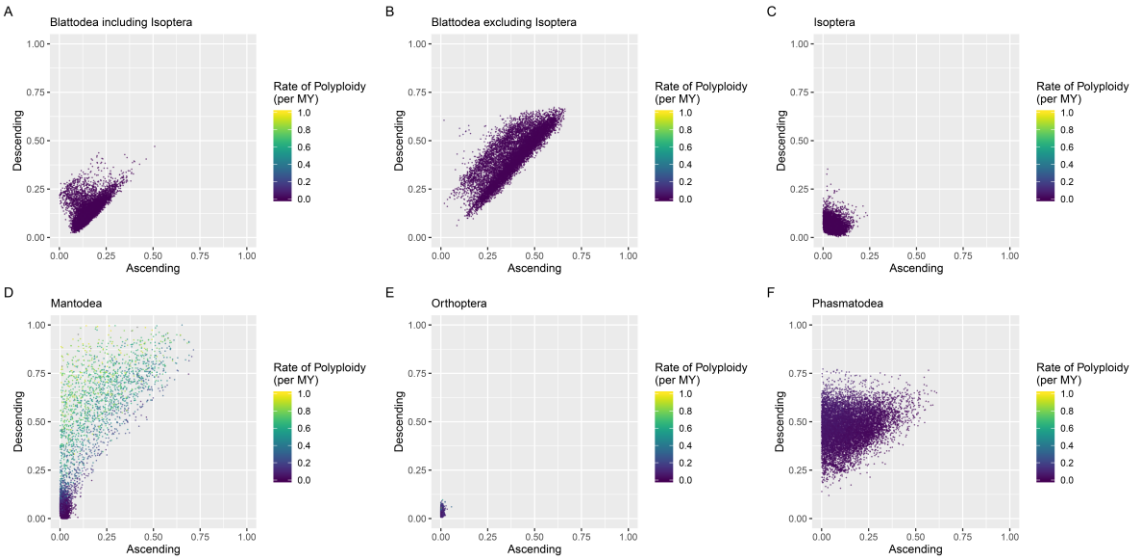
2.6. Conclusion

Our analyses and synthesis point to numerous clades that should be targeted with future whole-genome sequencing projects. For instance, chromosome level genome assemblies of XY orthopterans would be a powerful tool to discover whether the same chromosome is repeatedly co-opted into new XY systems. Similarly, we have identified sister species in Dermaptera, like *Labidura riparia* and *Nala lividipes*, that, though closely related, have 12–14 and 34–40 chromosomes, respectively. Whole-genome sequencing and comparative genomics would allow

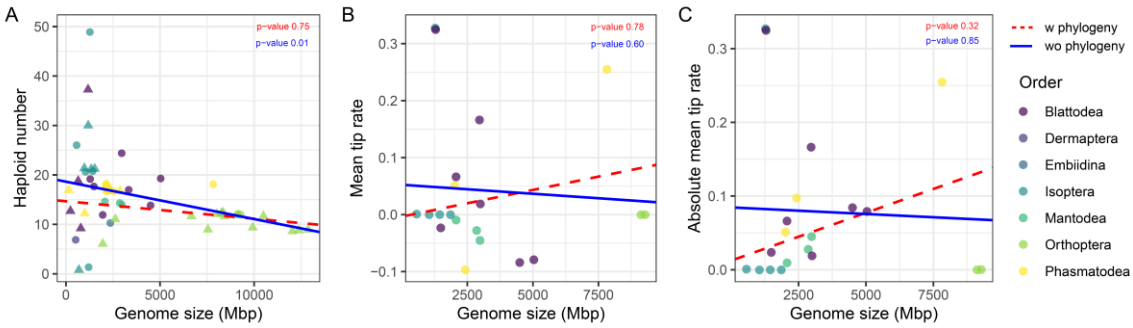
us to better understand how these dramatic restructuring of karyotypes have occurred. Taken as a whole, our results illustrate that the striking differences in chromosome number variation among orders is due to differences in rates and patterns of chromosome number evolution within orders and not due simply to sampling or the age of different clades. With the exception of Mantodea, all investigated orders had at least one transition rate that was different from one or more other orders. This has important implications for our understanding of the speciation processes. For instance, while many chromosomal speciation models (Baker & Bickham, 1986; White, 1978a) have been thought to be unimportant in recent times, it may be that in groups like Blattodea (excluding Isoptera), that exhibit high rates of chromosomal evolution, these models may explain an important source of extant diversity. Even if these older models do not represent a primary source of reproductive isolation, groups with higher rates of chromosome number evolution may be more likely to experience speciation. For instance, speciation may be facilitated under newer models of chromosomal speciation through sheltering of portions of the genome from admixture allowing incipient species to diverge and build-up genetic incompatibilities (Rieseberg, 2001). More broadly, depending on the importance of epistatic relationships, the reorganization of the genome through fusions and fissions may be important in determining the ability of organisms to adapt to novel environments (Stebbins, 1971).



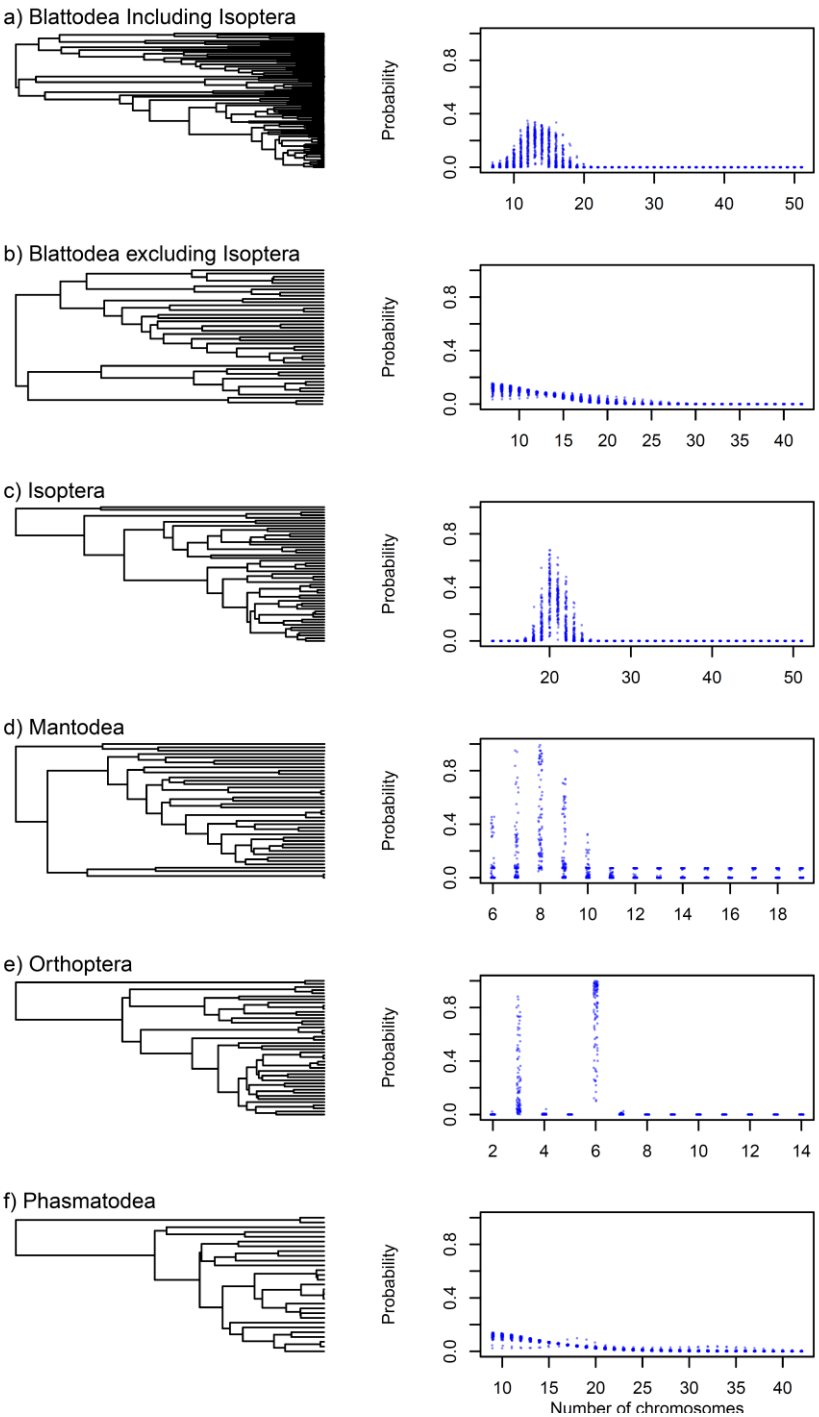
Sup. Figure 2.3 The red dotted line represents the cutoff point of 4780. approximately 94% of the taxa fall below this cutoff point.



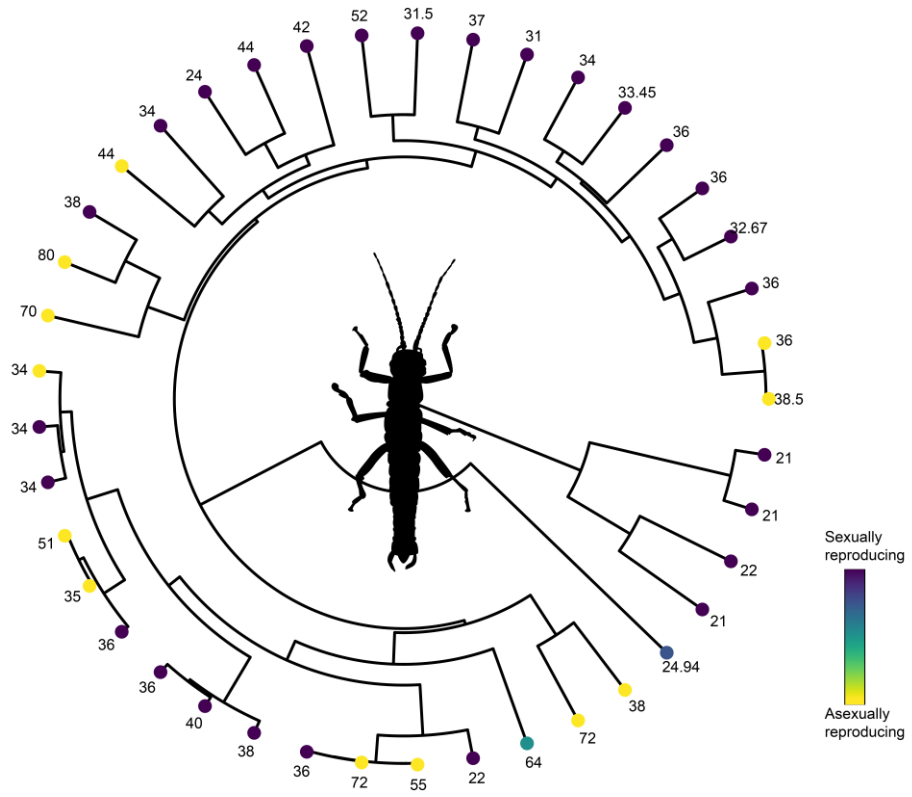
Sup. Figure 2.4 Rate of chromosome gain (ascending) and loss (descending) in five orders of Polyneoptera. Each point is colored based on the inferred rate of polyploidy.



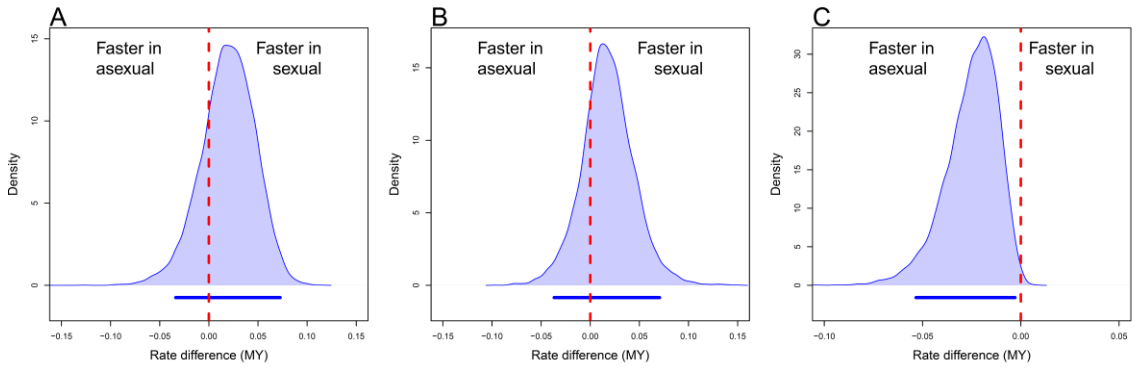
Sup. Figure 2.5 Impacts of genome size. Blue lines indicate regression line without phylogenetic correction and red dashed line indicates regression line with phylogenetic correction. P-values are printed in the same color in each plot. Circles are taxa that are present in the phylogeny and triangles are taxa that are not present in the phylogeny. A) Haploid chromosome number and genome size B) Mean tip rate and genome size C) Absolute tip rate and genome size.



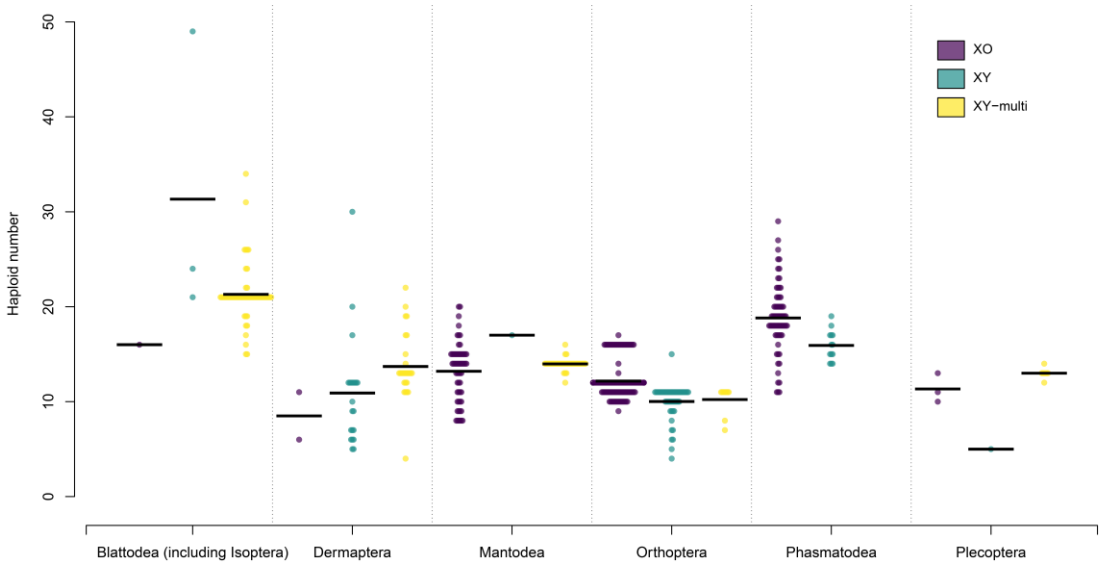
Sup. Figure 2.6 Ancestral state inference of chromosome number. Each point is the probability associated with a particular chromosome number on one of the 100 trees from the posterior distribution.



Sup. Figure 2.7 Phylogeny of Phasmatodea with reproductive mode and chromosome number. Tips are colored according to the mode of reproduction (sexual or asexual). Some lineages show intermediate colors. These are lineages which have both sexual and asexual populations. The shade of color indicates the probability of observing either reproductive mode in these lineages. The numbers indicate the mean chromosome number for each species.



Sup. Figure 2.8 Rates of chromosome number evolution in sexual and asexual lineages in Phasmatodea. A) fission, B) fusion, and C) polyploidy. Bars below the plot indicates the 95% HPD interval.



Sup. Figure 2.9 Variation in chromosome numbers parsed by sex chromosome system in six Polyneoptera orders. The vertical axis indicates the haploid chromosome count.

Sup. Table 2.2 The mean age and standard deviations applied to specified nodes in our beast analysis.

Node	Mean	Standard Deviation
Blattodea (including Isoptera)	197	28
Isoptera	136	4
Phasmatodea + Embiidina	164	31
Plecoptera	269	40
Dermaptera	302	46
Orthoptera	248	36
Notoptera	204	34

Sup. Table 2.3 95% Highest Posterior Density distribution for chromosome fissions, fusions, and polyploidy of the six analyses.

Order	Fission	Fusion	Polyploidy
Blattodea	0.063 - 0.257	0.04 - 0.243	0 - 0.005
Blattodea sensu stricto	0.173 - 0.604	0.202 - 0.619	0 - 0.01
Isoptera	0.002 - 0.1	0.019 - 0.114	0 - 0.006
Mantodea	0 - 0.329	0 - 0.737	0 - 0.69
Orthoptera	0 - 0.008	0.01 - 0.041	0.034 - 0.175
Phasmatodea	0 - 0.342	0.275 - 0.639	0.006 - 0.074

Sup. Table 2.4 Transition rates and mean number of transitions obtained from the stochastic mapping of sex chromosome evolution model.

Transition	Mean rate (95% credible interval)	Mean number of transitions
XO to XY	0.002 (0.0015 - 0.0026)	15.3
XY to XO	0.0021 (0.001 - 0.0036)	6.7

Sup. Table 2.5 Transition rates and the mean number of transitions from sexual to asexual reproduction from stochastic mapping. The transition rate of parthenogenesis to sexual reproducing was set to zero.

Mean rate (95% Credible Interval)	Mean number of transitions
0.0063 (0.0052 - 0.0078)	9.3

3. VARIATION IN RATES OF CHROMOSOME EVOLUTION IN AMPHIBIANS AND THE IMPACTS OF LIFE HISTORY ON RATES OF CHROMOSOME EVOLUTION IN ANURA AND CAUDATA

3.1. Overview

The division of genomes into chromosomes was first documented in the late 19th century. Since then, we have accumulated over a century worth of data for chromosome number. However, the evolutionary dynamics of chromosome number have been studied in only a handful of clades and often without modern comparative methods. Here we use data from Amphibians to understand the dynamics of chromosome number evolution across a large clade using a biologically realistic model. This group contains three orders with striking differences in the distribution of chromosome number. Furthermore, Amphibians show a remarkable diversity of reproductive modes, which we can use as a proxy for effective population size to understand the fitness effects of the mutations that change chromosome number in natural populations. We show that despite the differences in the karyotypic organization among the orders of amphibia, there is no significant differences in the rates of chromosome number evolution between these orders. We also find that Anura and Caudata with lower effective population sizes have faster rates of chromosome number evolution. These results suggest that variation in the range of chromosome number in amphibian orders is due to phylogenetic history, and the mutations that change chromosome number in lineages are mildly deleterious or underdominant.

3.2. Introduction

Division of the genome into discrete chromosomes is a fundamental aspect of genome architecture. The number of chromosomes a genome is divided into is often one of the first pieces of information we know about genomes. For example, the work of Walther Flemming in the 1880s identified a diploid chromosome number of 24 in *Salamandra infraimmaculata* (Flemming, 1882, 1887), which predates the Boveri-Sutton chromosome theory of heredity by two decades (Boveri, 1904; Sutton, 1902). However, despite over 130 years of karyological research, systematic studies of karyotype evolution have been taxonomically limited or lacked the use of modern comparative methods (Morescalchi, 1975; Voss *et al.*, 2011).

Changes in chromosome number can arise through various mechanisms and have a broad range of impacts on individuals and their fitness. The most straightforward changes in the chromosome number, where the chromosome number is decreased or increased by one unit, can arise through chromosome fusions and fissions, respectively. Mechanisms for chromosome fusion include Robertsonian translocation followed by the loss of non-essential DNA and telomere fusion followed by the random loss or inactivation of a centromere. Chromosome fission occurs when a chromosome splits at the centromere and generates new telomeric sequences on the broken end (Imai et al., 1988; Perry et al., 2004; Zhao et al., 2015). Changes in chromosome number can also occur through a polyploidy event (e.g. whole genome duplication) (Adams & Wendel, 2005). In this case, the haploid chromosome number is increased by a factor of two.
Changes in chromosome number are often considered to be mildly deleterious (White, 1954). For example, contributions of an unequal number of chromosomes from parents can cause fertility issues in their offspring due to improper segregation during meiosis (Harton & Tempest,

2012). Chromosomal rearrangements can also have various effects on genomic functionality beyond reproduction. In humans, for example, an extra copy on chromosome 21 is observed in individuals with Down syndrome, and monosomy of the X chromosome is seen in individuals with Turner syndrome (Shaffer & Theisen, 2010). Both conditions can result in physical abnormalities and are associated with variations in gene expression that accompany aneuploidy (Dürrbaum & Storchová, 2016; Epstein, 1989; Lippe, 1991). In addition, somatic aneuploidy is also highly correlated with cancer, possibly due to the loss of tumor suppressor genes or overexpression of oncogenes (Liu *et al.*, 2016). However, changes in chromosome number have also been observed as having no impact on fitness in some specific lizard populations (Porter & Sites, 1985). However, our understanding of the fitness effects of chromosome change are largely limited to either segregating variation within a population, disease-causing mutations, or the forced hybridization of diverged lineages. Comparative phylogenetic approaches to studying chromosome number change are particularly powerful because they offer the opportunity to understand the fitness effects of chromosomal mutations that lead to extant diversity in genome structure.

Amphibians exhibit a remarkable variation in chromosome numbers. For example, the lowest reported diploid chromosome number for amphibians is 14, seen in several frogs belonging to the genera *Arthroleptis*. The highest reported chromosome number is 108 chromosomes, seen in polyploid (12n) *Xenopus* species (B. J. Evans et al., 2015; Green et al., 1991). In addition to the broad variation in chromosome number, amphibian karyotypes exhibit the presence of gonochoristic polyploids, which strongly contrast to amniotes and insects where polyploidy is almost exclusively restricted to hybrid or parthenogenetic taxa (Blackmon et al., 2017; Fujita &

Moritz, 2009; but see: Z. Li et al., 2018). For example, at least five families of anurans have gonochoristic species with tetraploid or higher ploidy levels (Bogart, 1980).

The distribution of the number of chromosomes within each Amphibian order is diverse. Anurans exhibit one extreme where most species have comparatively few chromosomes that are relatively homogenous in morphology. Except for the enigmatic *Ascaphus truei*, all currently-sampled anurans lack micro-chromosomes (Green et al., 1980). In contrast, Gymnophiona represents the other extreme, with most species' karyotypes reminiscent of reptiles and comprised of a mixture of macro and micro chromosomes. As might be expected, this group also has the highest mean chromosome number within Amphibians (2N=32). Caudata is an intermediate between these two and is characterized by heterogeneous, clade-specific patterns of chromosome number and morphology. For example, some groups of Caudata appear more similar to Gymnophiona, such as Hynobiidae, with a mean diploid chromosome number of 56. All Hynobiidae also possess micro chromosomes, with *Hynobius retardatus* having just six micro chromosomes and *Ranodon sibiricusin* having the most with 19 (Kuro-o et al., 1987; Morescalchi et al., 1979). In comparison, most other Caudata groups have fewer chromosomes, similar to Anura with largely metacentric chromosomes and the absence of micro chromosomes (Sessions, 2008). Thus, the highly variable and clade-specific patterns of micro chromosomes and chromosome numbers among extant Amphibians have led to conflicting hypotheses describing amphibian karyotype evolution.

In our study, we used amphibian chromosome data to explore the dynamics of chromosome number evolution further. Specifically, we were interested in knowing whether the variation

among orders is due to different rates of fusion, fission, and polyploidy or simply due to varying time for divergence among clades. To test this, we evaluated whether there are significant differences in the rates of chromosome number evolution between the amphibian orders Anura (frogs), Caudata (salamanders), and Gymnophiona (caecilians).

We were also interested in the selection coefficient of mutations that change chromosome number among lineages. We used development type as a proxy for the effective population size. Amphibians show four major developmental types: aquatic breeding, gel nesting, foam nesting and terrestrial direct development. In aquatic breeding, gel, and foam nesting reproductive modes, which we identify as indirect development, there is an aquatic larval stage and a terrestrial adult stage. However, in the terrestrial direct development, we do not see the aquatic larval stage. Instead, a froglet emerges out of the eggs which are deposited on moist soil or carried on the back of the female frog. We assumed that those taxa that undergo direct development would have a small population size due for several reasons (Dubois, 2004; Duellman, 1989). Indirect developing amphibians release large amounts of eggs compared to direct developing amphibians. We also see high species richness and small range sizes in direct developing amphibians compared to their indirect developing counterparts, suggesting that direct developing amphibians have smaller effective population sizes (Dubois, 2004; Hanken, 1999). Although, this pattern is consistent with Anurans, we see that direct developing Salamanders typically have larger effective population sizes than their indirect developing counterparts (Funk et al., 1999; Gill, 1978; Hernández-Pacheco et al., 2019; Kazitsa et al., 2018)

We hypothesized that if mutations that change chromosome number are deleterious, we should see higher rates of chromosome number evolution in clades likely to have smaller effective population sizes. In contrast if these mutations are beneficial, we should see lower rates in clades likely to have smaller effective population size. Finally, if these mutations are neutral, we should see no significant differences in rates between clades with small and large effective population size.

3.3. Methods

3.3.1. Data Collection and curation

We used the amphibian karyotype database to collect chromosome number data (Accessed on: August 28th, 2021) (Perkins *et al.*, 2019) and the AmphiBIO database v1 (Accessed on: September 1st, 2021) (Oliveira *et al.*, 2017) to collect the development type (direct or indirect) for the suborder Anura. The phylogeny used to analyze these data was taken from a previous study (Liedtke *et al.*, 2018). Our phylogenetic dataset included a posterior distribution of 100 phylogenetic trees.

In order to maximize the overlap between the trait and phylogenetic data, we searched for species-level and genus-level matches within our phylogenetic and chromosome number datasets using the approach discussed in Blackmon and Demuth (2014). Briefly, we kept all instances of species-level matches between the two datasets and searched for genus-level matches when there were no species-level matches for a given genus. In cases of genus-level matches where one species had trait data, but no phylogenetic data and the other had phylogenetic data but no trait

data, the species were combined into a single chimeric taxonomic unit representing the given genus. Our final dataset consisted of 1091 species level matches and 15 genus level matches.

3.3.2. Modeling chromosome number evolution

We fit a single model of chromosome number evolution for each suborder of amphibians. We used the R packages chromePlus and diversitree to build the likelihood model (Blackmon et al., 2019; FitzJohn, 2012). First, we transformed the dataset into a diversitree compatible matrix using the 'datatoMatrix' function in the R package chromePlus. This transformation generates a probability matrix that describes the probability of being in each state for each taxonomic unit. Using this probability matrix, we generated the initial likelihood model using the R function 'make.musse' in the package 'diversitree'. Then constrained the likelihood model using the function 'constrain.musse' in the package 'chromePlus' to make the final model of chromosome number evolution. Our model included parameters for chromosome fusion, fission, and whole-genome duplications.

Additionally, since we used an SSE model to generate our likelihood function, we had two additional parameters for speciation and extinction. However, we were not interested in these two parameters. The presence of speciation and extinction parameters will not constrain our inference of the rate estimations for chromosome fusions fissions and whole-genome duplications.

Once we built the likelihood model, we ran an MCMC for 100 generations for each tree in our posterior distribution of phylogenetic trees. We used an exponential distribution with a rate

parameter of 5 as the probability distribution for MCMC sampling (prior). We visually examined the likelihood of each MCMC and discarded 50% of our generations as burn-in. The post-burnin portion of each MCMC was combined to produce a total of 5000 samples for each parameter.

3.3.3. Analysis of direct development

chromePlus model of chromosome number evolution can utilize the information of a binary trait, and we can re-configure the chromosome number evolution model to estimate the impact of the binary trait on chromosome number evolution. We repeated our analyses as described previously, using development type (direct or indirect) as our binary trait of interest. With this new data, our model of chromosome number evolution consists of eight parameters. Three parameters describe the evolution of the chromosome number at each state, and the last two describe the transition between direct and indirect development. However, based on our dataset's broad agreement that direct development is a derived state with no reversions, we set the transition rate from direct development to indirect development to zero (But see Chippindale et al., 2004; Wake & Hanken, 2004).

3.3.4. Model adequacy

In order to evaluate the adequacy of our model, we simulated chromosome numbers and binary traits using the rate estimates from our previous step. We then calculated summary statistics that described the distribution of the chromosome numbers and the binary trait and compared them with the summary statistics of the empirical dataset. We used the proportion of the indirect developing taxa and the coefficient of variance in the chromosome number as our summary statistics.

3.4. Results

3.4.1. Distribution of the chromosome numbers and development type within Amphibia

Our chromosome dataset contained karyotype data for 2124 amphibian species. Our trait dataset had data for 6777 species, out of which 3997 are direct developing and 1700 indirect developing. Upon searching both datasets for overlapping species, we found that we had trait and chromosome data for 1107 species. Among the three amphibian orders, Anura had the largest range for chromosome numbers, followed by Caudata and Gymnophiona. Anuran species had a minimum haploid number of 7 in the genus *Arthroleptis* and a maximum of 54 in the genus *Xenopus*. In Caudata, the minimum was 11 in the genus *Notophthalmus*, and the maximum was 39 in the genus *Onychodactylus*. Gymnophiona, with the smallest range and upper limit, had a minimum of 10 in the genus *Chthonerpeton* and a maximum of 22 in the genus *Ichthyophis*.

Within Anura, a single genus, *Xenopus*, had the highest chromosome numbers, and this clade was characterized by having short branch lengths. Therefore, our chromosome number evolution analysis was performed using two phylogenetic datasets, one including the genus *Xenopus* and one excluding the genus, to control for our results being driven by this one clade (figure 3.1).

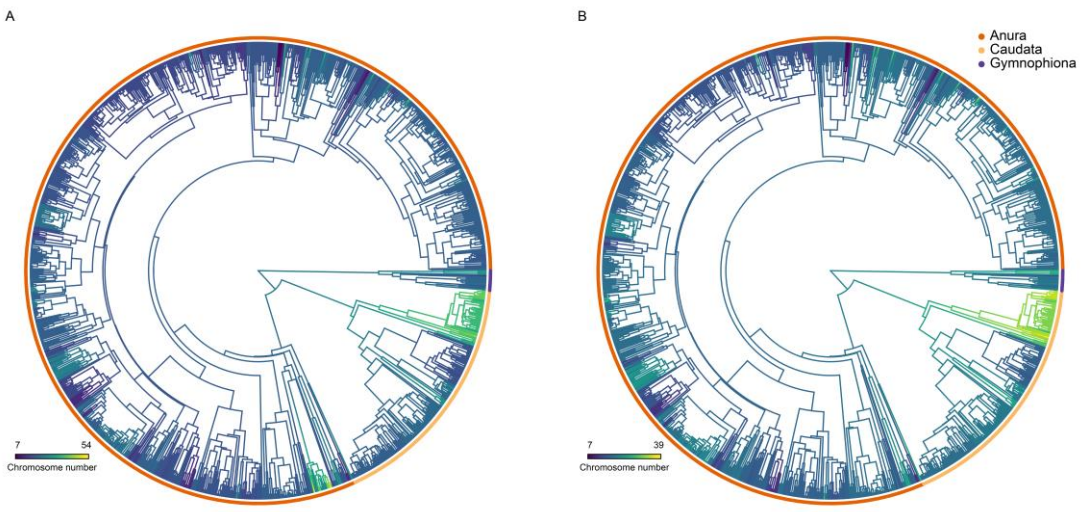


Figure 3.1 Phylogeny of Amphibia. The branches are colored based on the chromosome numbers observed at the tips. Outer ring in each phylogeny indicates the amphibian orders. A) Phylogeny with the Anuran genus *Xenopus* included. B) Phylogeny with the Anuran genus *Xenopus* excluded.

3.4.2. Rates of chromosome number evolution within orders

We first tested whether the variation in chromosome number that we see in amphibian orders is due to phylogenetic history or due to these orders having distinct rates of chromosome number evolution. We used a Markov model that included chromosome fusions, fissions, and polyploidy. We applied this model to each of the 100 trees in our posterior distribution and calculated the 95% Highest Posterior Density (HPD) interval for each rate parameter. We find that the HPD interval overlaps in all three orders for chromosome fusions, fissions, and polyploidy, suggesting no significant difference in the rates of chromosome number evolution between the orders (figure 3.2). We further tested if the removal of the clade *Xenopus* had an impact on the rates of chromosome number evolution. With the removal of the genus *Xenopus*, we see a reduction in the rates of polyploidy in Anurans. However, we do not see a significant difference between orders when we remove *Xenopus*.

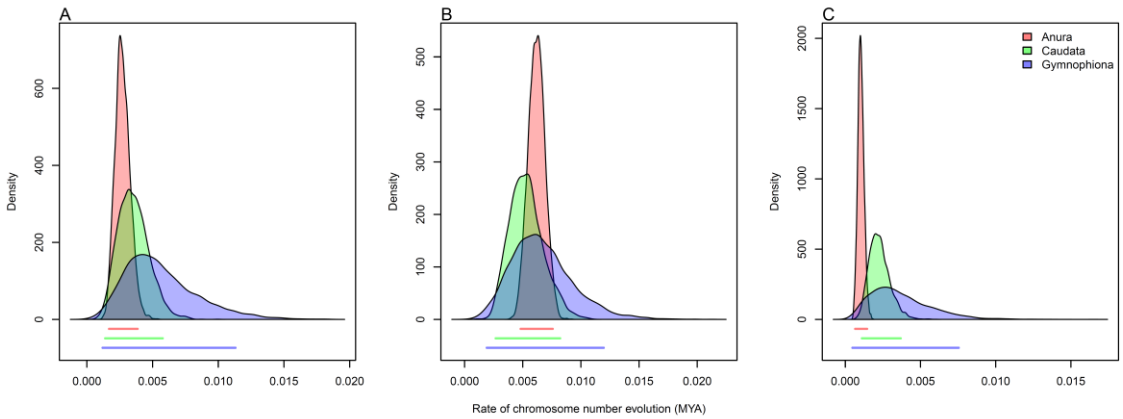


Figure 3.2 Rate of A) fusion B) fission and C) polyploidy among Amphibian orders. Colors represent the three orders and the bars underneath the distributions represent the 95% HPD interval. The genus *Xenopus* (Anura) is included in this analysis.

3.4.3. Impact of direct development on the rates of chromosome number evolution

We inferred the impact of a binary trait on the chromosome number evolution using the chromPlus model of chromosome number evolution. Our choice of the binary trait was direct development. The two states in this binary trait are direct-developing taxa (those taxa that do not have a free-living tadpole stage) and indirect developing taxa (those taxa that have a free-living tadpole stage). We estimated the rates of chromosome number evolution for each state of the binary trait and calculated the Delta R (ΔR) statistic for each rate parameter and the 95% highest posterior density interval (HPD) of each ΔR statistic. The ΔR statistic is defined as the difference between the rate of chromosome number evolution between two states. For consistency, we subtracted the rate estimates of direct-developing taxa from that of the indirect developing taxa.

If the HPD interval does not include zero, then we identify that one state has a higher rate for the parameter of interest compared to the other state. The state that has a higher rate can be identified

by the sign of the HPD interval (whether it is positive or negative). If the HPD interval spans zero, then we identify that neither state have higher rates.

In our analysis of chromosome number evolution for all amphibians, we find that the rate of chromosome fusion is significantly higher in the direct-developing taxa compared to the indirect developing taxa (Figure 3.3A). There is no significant difference in the rate of fission between the two developmental modes. However, when we exclude the genus *Xenopus* (which predominantly has indirect developing taxa) we do not see any changes in the observed pattern (Figure 3.3B). The presence or absence of the genus *Xenopus* had a significant impact on the polyploidy rate. In our analysis including the genus *Xenopus*, we see that the rate of polyploidy is significantly high in indirect developing taxa compared to direct developing taxa. However, once the genus *Xenopus* is excluded, the difference in the rate of polyploidy becomes non-significant (Figure 3.3).

We further tested the rates of chromosome number evolution with respect to the development type within Anura and Caudata. We excluded Gymnophiona as the number of datapoints we have is inadequate to get reliable parameter estimates. In Our analysis of Anura, we see that the rate of chromosome fusion is significantly higher in direct-developing taxa compared to that of indirect developing taxa (Figure 3.4A). However, we do not see any significant differences in the rates of chromosome fission and polyploidy. However, in our analysis within Caudata we see the opposite pattern where all three rates (fusion, fission, and polyploidy) are significantly higher in indirect developing taxa compared to that of direct developing taxa (Figure 3.4B).

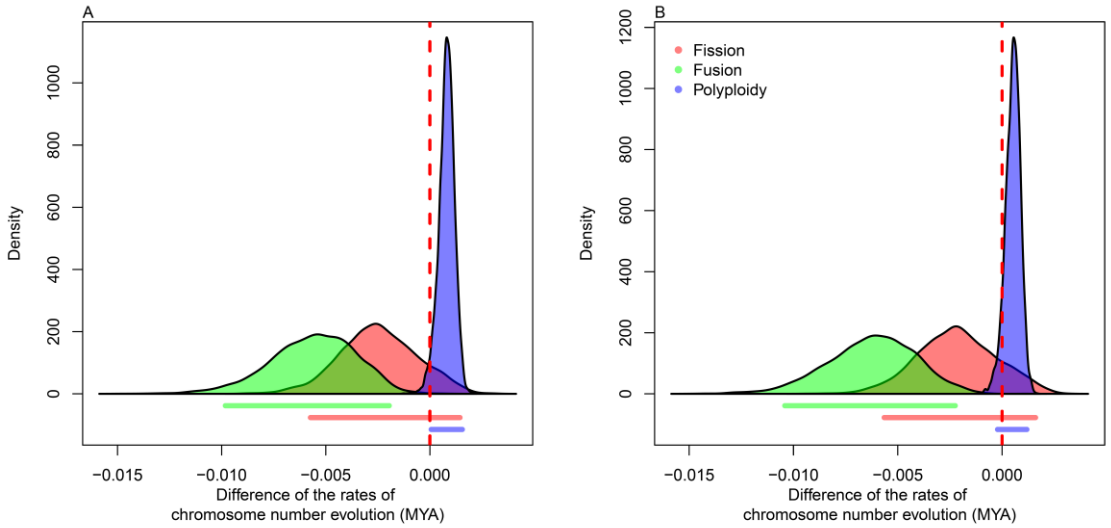


Figure 3.3 Distribution of the ΔR Statistic between direct developing and indirect developing Amphibians. We calculated the ΔR by subtracting the rate estimates of indirect development from direct development. A) Including and B) excluding the Anuran genus *Xenopus*.

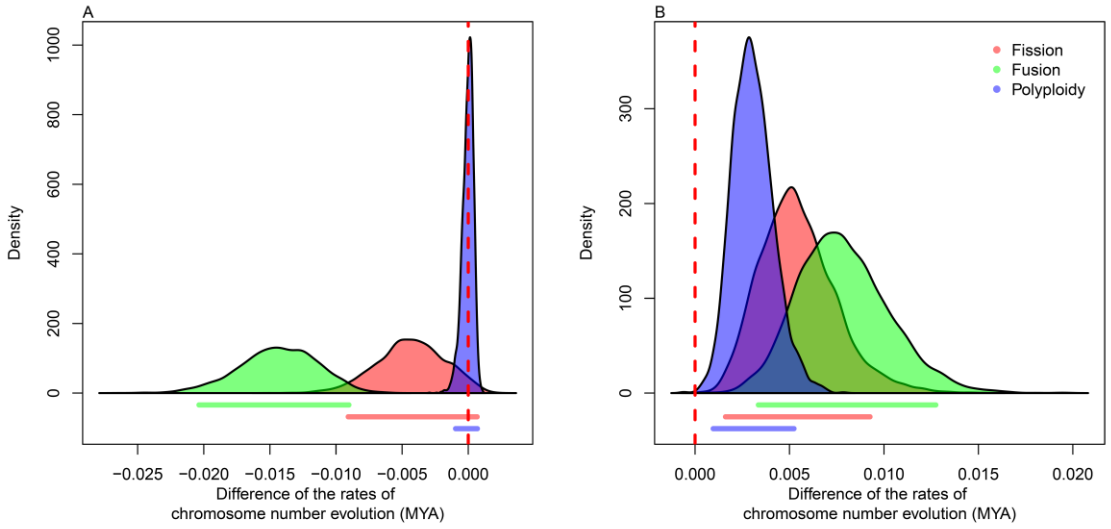


Figure 3.4 Distribution of the ΔR Statistic between direct developing and indirect developing Amphibians. We calculated the ΔR by subtracting the rate estimates of indirect development from direct development. A) within Anura and B) within Caudata

We also evaluated our model to see if the model was able to capture the underlying dynamics of the chromosome number evolution within amphibia. We performed model evaluation by simulating chromosome numbers and binary traits using the rate estimates we obtained from our Markov model. We then calculated summary statistics for each simulated dataset and obtained a distribution for the given summary statistics. We used the proportion of indirect developing taxa and the Coefficient of Variance in chromosome number as the summary statistics for our model evaluation representing the two traits used in this evaluation. We then compared the same summary statistics for our empirical dataset with the distributions from our simulations. We find that the simulated datasets infer a marginally higher proportion of indirect developing taxa compared to the empirical dataset (Figure 3.5A). However, we see that the coefficient of variance of the simulated dataset is centered around the coefficient of variance of the empirical dataset (Figure 3.5B).

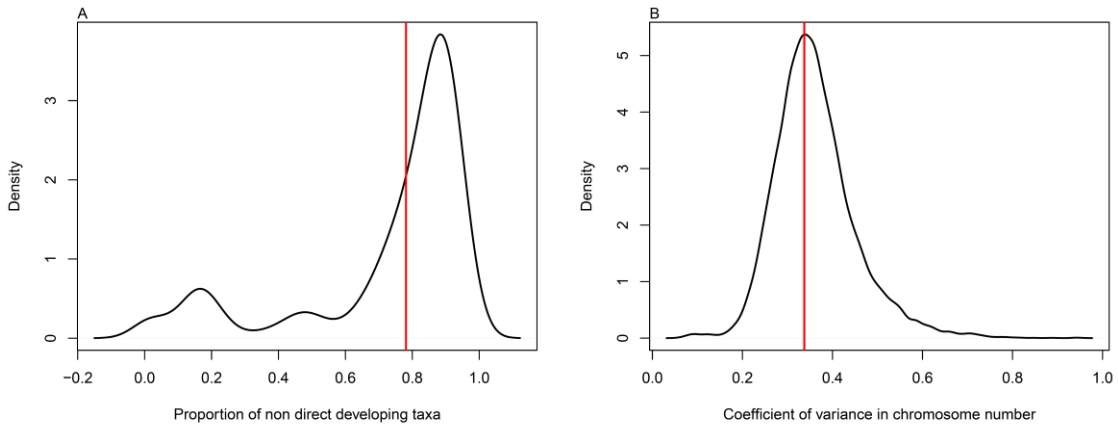


Figure 3.5 Model adequacy testing. A) proportion of indirect developing taxa and B) Coefficient of variance in the chromosome number. Red vertical bar is the empirical value, and the black line is the values from simulated datasets.

3.5. Discussion

Chromosome number is a fundamental component of the genome. Yet, after more than a century of study, it remains difficult to make broad generalizations about chromosome number evolution that apply across the tree of life. Even with datasets having information on chromosome numbers for many species, there was no clear answer to why some clades have small variation in chromosome number and others have a larger variation in chromosome number. However, we see a recent uptick in studies that examine chromosome number evolution at broad taxonomic scales, using model-based approaches, to answer the question of chromosome number variation in the tree of life, to make broad generalizations of chromosome number evolution and identify traits that have an impact on the chromosome number evolution (Blackmon et al., 2019; Ruckman et al., 2020; Sylvester et al., 2020).

Despite having a rich history of cytogenetics, we do not have a clear understanding of what factors lead to rapid or slow evolution of Amphibian karyotypes (Flemming, 1882, 1887). A recent synthesis of Amphibian chromosome numbers has paved way to understand the evolution of chromosome number across Amphibia by collecting all available chromosome number data scattered across many journals and locked behind paywalls into a single publicly available dataset (Perkins et al., 2019). Here we utilize data from this Amphibian dataset to investigate broad scale patterns of chromosome number evolution in Amphibia.

3.5.1. Do amphibian orders differ in the rates of chromosome number evolution?

The distribution of chromosome numbers within amphibians is characterized by two extremes. In one extreme, represented by Anura, we see species with few macro chromosomes. On the other

extreme, represented by Gymnophiona, we see species with many chromosomes and a mixture of micro and macro chromosomes. The order Caudata has species that represent karyotypes of both extremes. We see many studies that address chromosome number evolution at a fine scale within these orders (e.g., genera level or family level), however, to our knowledge, there are no studies that explore the dynamics of the chromosome number evolution on the order level scale for amphibians.

In our amphibian dataset, we have found that the rates at which fusion, fission, or polyploidy events occur did not vary significantly across the three orders despite their differences in the range of their karyotype sizes. Our results suggest that the variation in chromosome number we see in these orders is not due to differences in the rates of chromosome number evolution but rather due to phylogenetic history.

3.5.2. Fitness impact of mutations that change chromosome number

Changes in chromosome number can be neutral, beneficial, or detrimental. For example, we see the full range of fitness effects on changes in chromosome number in lemurs in crosses with parents having different chromosome numbers (Ratomponirina et al., 1988). However, the general acceptance is that these changes in chromosome number are neutral or mildly deleterious (White, 1954). If they are mildly deleterious, we should see less changes in chromosome number in populations that have a large effective population size compared to populations with smaller effective population size. In contrast, if the changes are beneficial, we should see higher rates of chromosome number evolution in populations with high effective population size. In populations

with low effective population size, the efficacy of selection becomes much weaker and therefore genetic drift dominates the fate of mutations.

We used direct development as a proxy for population size. We hypothesize that those taxa that underdo direct development have a low effective population size compared to those taxa that do not undergo direct development. For instance, we see that the clutch size of direct developing taxa are on average smaller than indirect developing taxa (Furness et al., 2022; Gould et al., 2022). Therefore, if changes in chromosome number are beneficial, we expect to see high rate of chromosome number in indirect developing taxa. If changes in chromosome number are detrimental, we expect to see the opposite pattern where indirect developing taxa having a low rate of chromosome number evolution. Finally, if the changes in chromosome number are neutral, we expect to see similar rates of chromosome number evolution in both direct and indirect developing taxa.

3.5.3. Do rates of chromosome number evolution vary between direct and indirect developing taxa?

Our results show that in amphibians, direct developing species have a higher rate of chromosome fusion than indirect developing species. These results align with our original hypothesis, which was predicated on the idea that direct-developing taxa would be of a small effective population. A smaller effective population leads to low efficacy of selection and, therefore, higher rates of chromosome number evolution. Thus, our data support the notion that direct-developing amphibian taxa would exhibit higher rates of chromosome number evolution.

Interestingly, rates of polyploidy did not differ significantly between direct-developing and indirect developing taxa except when we include the Anuran genus *Xenopus*. Our analysis with *Xenopus* included shows that the rate of polyploidy is significantly high in indirect developing taxa than direct developing taxa. The impact of the removal of the genus *Xenopus* on the rate of polyploidy suggests this rate parameter is driven by the karyotypes of *Xenopus* species and the short branch lengths of this clade. However, polyploidy events in *Xenopus* are characterized by allopolyploids where we see hybridization followed by genome duplication. In contrast most amphibian polyploidy events are characterized by autopolyploids (Kobel & Du Pasquier, 1986; Tymowska, 1991).

Our analysis on Caudata karyotype evolution implies a complete opposite pattern of chromosome number evolution with respect to the developmental mode compared to Amphibians. We see that those taxa that undergo direct development having low variation in chromosome number (haploid chromosome number range from 13 to 14) while indirect developing taxa with a much higher variation in chromosome number (haploid chromosome number range from 11 to 39). Our analysis shows higher rates of chromosome number evolution (fission, fusion, and polyploidy) in indirect developing taxa compared to direct developing taxa. Based on our hypothesis on the relationship of development mode and the population size, these results suggest that the changes in chromosome number within salamanders are beneficial. However, we see that some of these salamanders, especially those that are indirect developing species have a smaller effective population size in the range of 20 to 200 (Funk et al., 1999; Gill, 1978; Kazitsa et al., 2018). Therefore, we think that our results still support the idea that

chromosome number changes are deleterious. However, our hypothesis on direct development and population size does not hold true with respect to Caudates.

3.5.4. Conclusions

Here we examine the patterns of chromosome number evolution in Amphibians and how development type impacts the chromosome number evolution. Our results show that there is no significant difference in the rates of chromosome number evolution between amphibian orders despite having many differences in the general karyotype pattern observed in these orders. We also show that, in general, taxa that undergo direct development having higher rates of chromosome number evolution suggesting that these changes in chromosome number are deleterious. Although we see an opposite pattern in Caudates where indirect developing species having higher rates, upon further examination, we see that these indirect developing species have low effective population size.

4. GENOME ASSEMBLY AND POPULATION GENETICS OF THE SCARAB BEETLE *CHRYSINA GLORIOSA*

4.1. Overview

The Jewel scarab *Chrysina gloriosa* is one of the most charismatic beetles in the United States. Additionally, *C. gloriosa* is highly sought after by professional and amateur collectors worldwide. However, we lack information about the population structure, current population size, or the demographic history of this beetle. Without this data it is impossible to make well informed conservation decisions about this beetle. To address these questions, we reconstructed the *C. gloriosa* genome using a combination of long-read sequencing and Omni-C data. We further sampled *C. gloriosa* specimens from five mountain ranges and performed low-coverage skim sequencing. Also, using publicly available data, we modeled the species distribution under several climatic models. Our results show that *C. gloriosa* has highly structured populations, with only two populations showing evidence for recent gene flow. We also show that *C. gloriosa* experienced a population reduction coinciding with the changes in the climatic conditions occurred between the late Pleistocene and mid-Holocene epochs. Our predictions on habitat size distribution show that changing climatic conditions will severely impact the suitable habitat size. *C. gloriosa* populations are more likely to become extinct from their habitats due to extensive collection, harsh climatic conditions, and lack of gene flow between populations. Our results will be helpful in implementing reasonable conservation practices for this species and other species with similar distribution among sky islands.

4.2. Introduction

The scarab beetle *Chrysina gloriosa* LeConte 1854 (previously known as *Plusiotus gloriosa*) is one of the most charismatic beetles found in the continental US and one of the four beetles in the genus *Chrysina* with a range that extends into the United States (Cazier, 1951; LeConte, 1854; Young, 1957). Commonly known as the glorious beetle or the glorious scarab, this beetle has a metallic green body with silver stripes and blue eyes. Adults of *C. gloriosa* depend on juniper trees as a food source, and larval forms depend on decaying logs (Ritcher, 1966). Consequently, the distribution of *C. gloriosa* is limited by the distribution of the juniper trees (Young, 1957). In addition to the dependence on the Juniper trees as a food source, *C. gloriosa* plays an essential role as a pollinator of Juniper trees because *C. gloriosa* utilizes the juniper flowers as a mating ground (Graf, 2019). Currently, *C. gloriosa* is limited to higher elevation mountains of West Texas and Southeastern Arizona, an area commonly known as the sky islands. High elevation regions of these sky islands act as refugia for *C. gloriosa*, along with the three other *Chrysina* species that reach the United States and are thought to represent the remnants of a more widespread distribution that occurred during the Pleistocene epoch (Young, 1957). These species of the genus *Chrysina* are a relic of the cooler and wetter era of the Pleistocene epoch (Young, 1957).

Sky islands, in general, are a set of mountains that are separated from each other. The surrounding low land of these mountains contains vegetation utterly different from what is seen in these mountains, resembling a set of islands separated by a vast ocean. Hence, this geographic region is often described as sky islands with dessert seas. The Madrean sky islands of Arizona and the sky islands of Texas are separated by the Sonoran or the Chiricahua dessert, respectively.

This geographic formation provides an excellent opportunity to study speciation, adaptation, and genome evolution across short and long-time scales because populations of species seen in these isolated mountain ranges act as biological replicates that experience similar environmental pressures but are isolated from each other for centuries or even millennia. A condition that is immensely difficult to recreate in laboratory environments.

Due to its colorful nature, *C. gloriosa* is equally highly sought after by professional and amateur collectors worldwide (Genoways & Baker, 1979). With the current trend of rising global temperatures and the aggressive collection of these beetles, we think that extreme care may be justified to preserve the genetic diversity of these beetles seen in each mountain range. However, there is an absence of biological datasets and a lack of published literature related to the biology of not only *C. gloriosa* but also other beetles in this genus, making it difficult to make informed decisions about the conservation of these beetles. Most of the published literature on *Chrysina* species, including *C. gloriosa*, focuses on the physical aspects of the body of the beetle, such as polarizing properties of the cuticle (Brady & Cummings, 2010; Sharma et al., 2009).

The goal of this study was twofold. First, to reconstruct the genome of *C. gloriosa*. We used a single female specimen collected in 2019 and performed Nanopore long-read sequencing for genome reconstruction. In addition, we used Omni-C data to scaffold the genome at the chromosome level. Second, we reveal the population genomic characteristics (population structure and demographic history) of *C. gloriosa* populations in the United States. We ask the following questions. Does *C. gloriosa* migrate between and within mountain ranges? Is the demographic history shaped by the climate history in the southwestern United States? To answer these questions, we collected *C. gloriosa* samples from four mountain ranges in the sky islands

of Arizona and Davis Mountain in West Texas. We performed low coverage whole genome sequencing using Illumina short-read sequencing. We estimated population structure using model-based and non-model-based approaches and used sequential Markovian coalescent (SMC) models to estimate the demographic history. In addition, we used species distribution models to estimate the past, current, and future predictions of species distribution.

4.3. Methods

4.3.1. Sample collection

We collected *C. gloriosa* samples from West Texas and Southeast Arizona during July and August in 2017, 2018 and 2019 (Figure 4.1). All specimens were collected at night between 7.00 p.m. and 11.00 p.m. We used a combination of Mercury vapor and UV lamps placed in front of a vertical white sheet to attract beetles. All samples collected were stored in 100% ethanol at -20°C until DNA extraction.

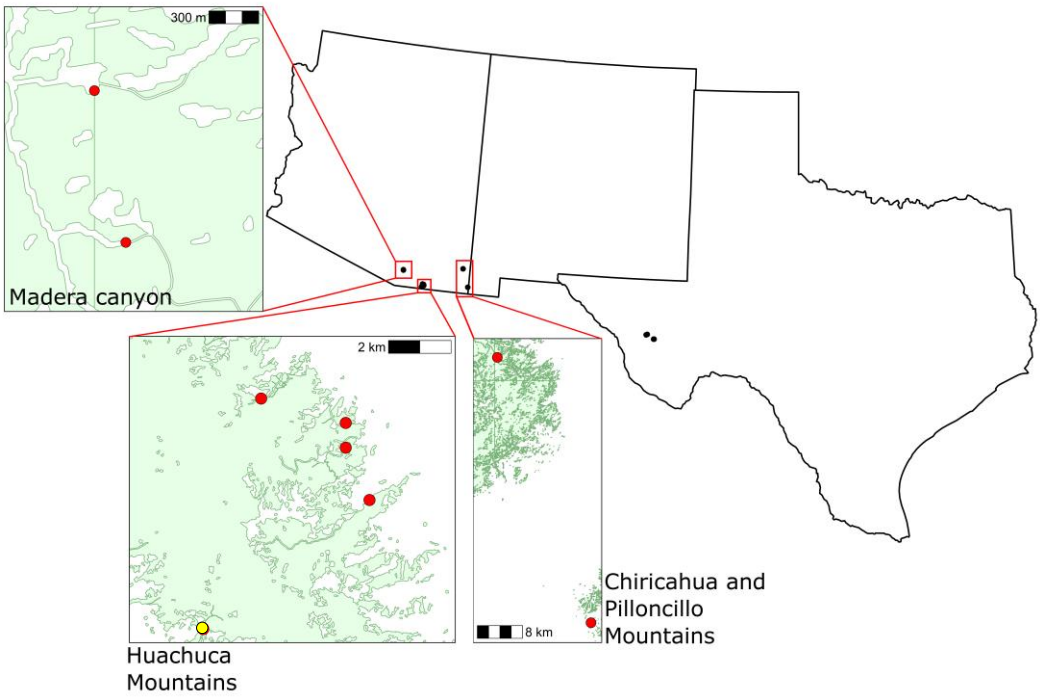


Figure 4.1 Sampling locations of *Chrysina gloriosa* across West Texas and Southeastern Arizona. The sampling location which we used a single female specimen to assemble the genome is indicated in yellow color.

4.3.2. Genome reconstruction

4.3.2.1. DNA extraction and sequencing

We used a single female specimen for DNA extraction. The sample we used for DNA extraction came from our collections at the Ida Canyon, Huachuca Mountains on the 28th of July 2019. To determine the sex, we dissected the abdomen and examined reproductive structures. DNA extraction and sequencing were done at the Texas A&M Institute for Genome Sciences and Society Core facility. We dissected the muscle tissue from the legs for DNA extraction. We used the Nanobind insect BIG DNA kit v 0.18 (Circulomics) for DNA extraction and followed the Circulomics high-molecular-weight insect DNA extraction protocol. The integrity of the extracted DNA was assessed using a Genomic DNA ScreenTape on a TapeStation (Agilent). We

used the Nanopore sequencing platform to generate long-read sequencing. The sequencing libraries were prepared following the manufacturer's protocol using the SQK-RAD004 rapid sequencing library. Finally, we used six R9.4.1 MinION flow cells to generate the sequencing data. The final data output consisted of 45.56 GB of data at an estimated coverage of 53x (we estimated the genome size [next section] to be approximately 850 MB using flow cytometry).

4.3.2.2. Estimation of *Chrysina gloriosa* genome size

Genome size of *C. gloriosa* was determined using flow cytometric methods per Johnston et al. (2019). Neural tissue from individual frozen samples of *Chrysina gloriosia* was dissected and deposited into 1 mL of Galbraith buffer. All samples were co-prepared with a standard (lab stock of *Drosophila virilis*, genome size = 328 Mbp). Samples were then gently ground with a Kontes "A" pestle approximately 15 times to release nuclei. After passing samples through 41 micrometer mesh filters, samples were stained with 25 µl of 1mg/µl propidium iodide and incubated in the dark. Samples were then run on a Beckman Coulter CytoFlex flow cytometer with a 488 nm blue laser. Means of 2C nuclei fluorescence peaks were measured for both the sample and the standard using gating methods supplied with the instrument's software before calculating the estimated genome size.

4.3.2.3. Dovetail Omni-C library preparation and sequencing

The preparation of Omni-C libraries and sequencing were performed by Dovetail Genomics. For each Dovetail Omni-C library, chromatin was fixed in place with formaldehyde. Fixed chromatin was digested with DNaseI and then extracted. Chromatin ends were repaired and ligated to a biotinylated bridge adapter, followed by proximity ligation of adapter-containing ends. After

proximity ligation, crosslinks were reversed, and the DNA was purified. Purified DNA was treated to remove biotin that was not internal to ligated fragments. Sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adapters. Biotin-containing fragments were isolated using streptavidin beads before PCR enrichment of each library. The library was sequenced on an Illumina HiSeqX platform to produce approximately 30x sequence coverage.

4.3.2.4. Genome assembly

We assembled the genome using NextDenovo v 2.5.0 (https://github.com/Nextomics/NextDenovo) and polished using NextPolish (Hu *et al.*, 2020). We first concatenated all reads from each flow cell into a single fastq file and filtered reads below 1000 bp using the program filtlong v 0.2.1 (https://github.com/rrwick/Filtlong). We then assembled the genome using NextDenovo with default parameters. We set the genome size to 850 Mb based on our genome size estimates. We then polished the genome using NextPolish with default parameters (Hu *et al.*, 2020).

We used Blobtools v1.1.1 for contaminant screening (Laetsch & Blaxter, 2017). Blobtools is a visualization tool that requires a genome, a map file and a hits file to process and visualize the presence of contaminants. To generate the map file, we mapped all raw reads against the assembled genome using minimap2, converted the resulting sam file to bam format using the SAMtools view module, and sorted and indexed using SAMtools sort and index modules, respectively. To generate the hits file, we first downloaded a preprocessed nucleotide blast database from NCBI (downloaded on 25th October 2021). We then used the blastn tool of NCBI

Blast v2.12.0 to blast the contigs against the preformatted blast database to generate the hits file (Camacho *et al.*, 2009). Finally, we ran blobtools using these input files to generate a blobplot showing the level of contamination in the genome assembly.

4.3.2.5. Assembly and annotation of Mitochondria

We downloaded the mitochondrial assembly of *Tribolium castaneum* through NCBI genome browser (accessed date May 25, 2022) and blasted against the *Chrysina gloriosa* assembly using BLAST v2.11 to filter mitochondrial contigs. We then built a dot plot using LAST v1045, comparing the mitochondrial contig of the *C. gloriosa* assembly with the mitochondrial assembly of the *Tribolium castaneum* for further confirmation. The LAST plots indicated that C. *gloriosa* mitochondria was assembled multiple times (back-to-back assembly). Therefore, we reassembled *C. gloriosa* mitochondria using a circularization tool to prevent back-to-back assembly. All confirmed mitochondrial contigs were removed from the nuclear assembly.

We mapped all reads to the mitochondrial contig using minimap2 v 2.24 and filtered only aligned reads using SAMtools view module (H. Li, 2018; H. Li et al., 2009). We further removed spurious alignments by filtering all mapped reads that have a mapping quality below 30. We assembled the resulting reads using Unicycler under default settings to generate the mitochondrial assembly (Wick *et al.*, 2017). Finally, we annotated the mitochondrial contig of *C. gloriosa* assembly using Mitos 2 web server (Bernt *et al.*, 2013). We used RefSeq 89 Metazoa as the reference sequence and invertebrate as the genetic code. We set the maximum overlap parameter to 100 and selected sensitive only under the ncRNA tab. We kept all other parameters default. Finally we visualized the mitochondrial annotation using OG-DRAW v 1.3.1 (Greiner et al., 2019)

4.3.2.6. Scaffolding Nuclear Assembly

We created the initial contact map using Juicer v 2.0. We ran Juicer with the –assembly flag to generate the input files necessary for the scaffolding program 3d-DNA (Dudchenko et al., 2017; Durand et al., 2016). We then ran 3d-DNA with five rounds of mis-join correction and used sites with map quality of 30 or higher for scaffolding and visualization. Next, we used the assembly tools module in Juicebox v 1.11 (JBAT) (Dudchenko *et al.*, 2018) to curate the assembly output from 3d-DNA, following the recommendations of Howe et al. (2021). However, we only curated large-scale mis-assemblies that are easily noticeable (e.g., combining scaffolds that are separate but should be a single scaffold based on the evidence from the contact matrix). We used the software HiC-Hiker to further correct fine-scale misassembles present within scaffolds. HiC-Hiker uses a probabilistic approach to determine the possible orientation for a given set of contigs. We set the maximum distance for the probability of observing a contact between two loci (-K flag) to calculate automatically. Finally, we assessed the completeness of the genome using BUSCO v 5.2.2 (Simão *et al.*, 2015). In addition, scaffolds were run as a query to a *Tribolium castaneum* reference genome and the scarab beetle *Trypoxylus dichotomus genome*, separately, with minimap2 v2.24 (H. Li, 2018). The resulting pairwise alignment file was used to generate circos plots between the *C. gloriosa* and *Tribolium castaneum* and *Trypoxylus dichotomus* genomes using Circos v0.69-9 (Krzywinski *et al.*, 2009).

4.3.2.7. Annotation of repeats

We used the Extensive *de novo* Transposable Element Annotator (EDTA) v2.0.0 pipeline to identify transposable elements across the *C. gloriosa* genome (Ou *et al.*, 2019). Tandem Repeats Finder (TRF) v4.09.1 was used to identify large tandem repeats (Benson, 1999). The parameters

for TRF were 2, 5, 7, 80, 10, 50, and 2000 with a maximum expected repeat array length of 10 million bases. The R package micRocounter was used to identify SSRs with repeat units of 2-6 bases (Lo et al., 2019). The minimum number of motif repeats to be considered an SSR was 6 for dimers, 4 for trimers, and 3 for tetramers, pentamers, and hexamers. The maximum allowable gap for a repeat array to continue was one base for all SSRs.

4.3.2.8. Genome annotation

We used Liftoff v 1.5.2 to map features from an annotated genome into the *C. gloriosa* genome assembly (Shumate & Salzberg, 2020). We used the annotation of *Trypoxylus dichotomus* and the repeat masked *Chrysina gloriosa* genome as the target (Wang *et al.*, 2022). We ran Liftoff with default settings. However, changed the minimap options within Liftoff (-r and -z parameters set to 2k and 5000, respectively) to account for species divergence as recommended by the developer. Furthermore, we used the 'polish' parameter to re-align coding sequences to account for missing start/stop codons or in frame stop codons.

4.3.3. Population genetic analysis

4.3.3.1. DNA extraction and sequencing

We used a single muscle tissue dissected from one of the hind legs of each beetle to extract DNA. We used the QIAGEN blood and tissue DNA extraction kit (QIAGEN) and followed the manufacturer's protocol. We used the NanoDrop (Thermo Fisher) to examine the quality of the extracted DNA and the Quantus fluorometer (Promega) to quantify the extracted DNA. We sequenced DNA through the Texas A&M AgriLife Genomics and Bioinformatics service center (https://www.txgen.tamu.edu/). We used Illumina short-read sequencing platform and used the

NovaSeq 6000 sequencing system. All specimens were sequenced using the 2 x 150 bp paired-end method at 1-2x coverage.

4.3.3.2. Mapping and variant calling

4.3.3.2.1. Quality control

First, we combined the forward and reverse reads from each lane to generate a single forward read file and a single reverse read file for each specimen. Then we quality-checked the reads using fastQC and combined all reports using multiQC for analysis (Ewels et al., 2016; Simons, 2010). Then, we used fastP to remove adapter sequences and trim reads based on quality (Chen et al., 2018). We removed sequence reads with a score of less than 20 on the Phred scale. We discarded the unpaired reads and only kept paired reads for mapping and variant calling. To ensure that the trimming step removed bad quality regions from the reads, we did an additional quality check using fastQC followed by multiQC. Our quality control step retained 98.7% of the data.

4.3.3.2.2. Mapping reads to the reference genome

First, we combined the nuclear and mitochondrial genome into a single file so that both organelle and nuclear genomic information are in a single fasta file. We then indexed the *C. gloriosa* genome using SAMtools faidx and BWA index modules (H. Li, 2013; H. Li & Durbin, 2009, 2010). We mapped all reads using the BWA-MEM algorithm and converted the resulting SAM files to BAM format using the SAMtools view module. Next, we sorted the mapped files by name and coordinate order using the SAMtools sort module. Then, we used SAMtools fixmate module to correct errors on read-pairing due to the alignment program, followed by SAMtools

markdup module to remove duplicate reads. Finally, we extracted mapped reads using the SAMtools view module. To remove spurious mappings, we only kept sites with a mapping quality greater than 30.

4.3.3.2.3. Variant calling and filtering

We used a combination of BCFtools and VCFtools for variant calling and filtering (Danecek *et al.*, 2011, 2021). First, we generated the initial variant file using the BCFtools mpileup module. We called variants using the BCFtools call module with the multiallelic caller (-m option in the BCFtools call module) as the variant calling option. We kept single nucleotide polymorphisms (SNPs) with a base quality of 20 and a mapping quality of 30. We separated the resulting VCF file into two subsets representing nuclear-only and mitochondrial-only variants. We further filtered the variants based on quality, read depth, missingness and minor allele frequency. We used the following thresholds to filter variants; minor allele frequency (MAF) of 0.05 (5%), maximum missing percentage across all samples as 0.8 (tolerate 20% missing data), minimum read depth per SNP and minimum read depth per SNP across all samples as 1x. We retained only bi-allelic single nucleotide polymorphisms (SNPs) in our variant dataset. Our final SNP dataset comprised 2 million nuclear variants and 577 mitochondrial variants.

4.3.3.3. Determination of sex and the X chromosome scaffold

We used a combination of read depth and Blast to identify the X chromosome scaffolds. First, we generated the mean read depth per scaffold using the SAMtools depth command. Then, we extracted the mean depth of the ten longest scaffolds representing the ten chromosomes of a typical Scarab karyotype. We find that, in different specimens, the smallest of the ten scaffolds

had a similar or half the coverage compared to the rest of the examined scaffolds. Each specimen had varying coverage from 1x to 2x for the longest nine scaffolds. Therefore, we normalized the coverage across all ten scaffolds to compare the coverage between specimens. For a given sample, we divided the coverage of all scaffolds by the mean coverage of the nine longest scaffolds. We then compared the coverage for scaffold ten with the rest of the scaffolds and assigned sex based on the average coverage of scaffold ten. To further validate our findings, we BLAST searched the *C. gloriosa* contigs scaffolds against the *Tribolium castaneum* genome and compared which contigs of *C. gloriosa* match the X chromosome of the *Tribolium castaneum* genome.

4.3.3.4. Assembly of the putative Y chromosome

We extracted all reads that did not map to the nuclear and mitochondrial genome using the SAMtools view module for each individual identified as a putative male. We then pooled all forward and reversed reads into a single forward and reverse fastq files. We also pooled all singleton reads into a single fastq file as well. We assembled these reads using the ABySS with varying kmer sizes (18 to 122 with increments of 8) and under two minimum kmer coverage multiplicity cutoff values (-kc flag set to 2 and 3). We selected the best assembly based on the total size, number of contigs and the N50 value.

4.3.3.5. Population structure of *C. gloriosa*

We filtered the VCF file for population structure analysis to remove variants in linkage disequilibrium using PLINK1.9 (Chang et al., 2015). We used a sliding window of 50 SNPs where we advanced by 10 SNPs. Additionally, we used a correlation coefficient (r2) of 0.1 to

filter SNPs that are in LD. This process resulted in 380,605 SNPs in our filtered VCF file for the nuclear genome. We did not apply the LD filter for the mitochondrial genome since it is a non-recombining single genome, and all SNPs are linked.

To analyze population structure, we used several approaches. First, we used the filtered SNPs to conduct a principal component analysis (PCA). We used PLINK1.9 to perform the PCA and the R package tidyverse to visualize the results (Chang et al., 2015; Wickham et al., 2019). We then used t-SNE, a machine learning approach, on principle components from the PCA to further reduce dimensionality and explore fine-scale population structure (Van der Maaten & Hinton, 2008). We set the 'perplexity' parameter, which defines the number of neighbors for a given point as ten since we had ten or fewer specimens from each location. We set the theta parameter to zero to increase the accuracy. However, selecting the theta parameter to a lower value will harm the speed of the calculations. One disadvantage of the t-SNE approach is that the parameters used (e.g., perplexity – the number of close neighbors to a given data point) will significantly impact the final output. Therefore, we ran the t-SNE analysis with a range of values for the perplexity parameter and used a single value that recapitulate the geographic distribution of data.

We also estimated the individual admixture proportions using Admixture v1.3 in a maximum likelihood framework (Alexander et al., 2015). We used the variant set filtered for linkage (same set of variants used to run PCA) to estimate individual admixture proportions. We set the K value (ancestral populations) to range from 2 to 12 with ten-fold cross-validation.

4.3.3.6. Estimation of effective migration surface

We used the program FEEMS to estimate the migration surface of *Chrysina gloriosa* across Arizona (Marcus *et al.,* 2021). FEEMS require four input files: genotypes, locations, outer boundary, and a dense global grid. We removed samples from Texas from our variant file as we assumed there is no migration between samples from Texas and Arizona. This removal resulted in some variants becoming monomorphic. We further removed these variants from the variant file as the program FEEMS fails in the presence of monomorphic sites. For the outer boundary, we used the border coordinates of the state of Arizona.

We used the R package dggridR to make a discrete triangular grid for Arizona. Each triangle represents an area of approximately 25 km^2. At this resolution, we can capture fine-scale migration patterns (e.g., those within mountains). We used PLINK for the imputation of missing genotypes. For the missing sites, we assigned the mean at each SNP. We then ran cross-validation for a range of values for the smoothness parameter (λ) ranging from 0.1 to 10 and selected the λ with the least cross-validation error. The smoothness parameter determines the strength of the penalty placed on the migration surface, with lower values for λ revealing fine-scale migratory patterns (Marcus *et al.,* 2021). Based on the cross-validation result, we chose 35.93 as the λ value and estimated the migratory surface with all other parameters kept at their default values.

4.3.3.7. Phylogeny reconstruction using mitochondria

We generated consensus sequences for the mitochondria of each sample using a custom R script as done in the previous section. We then used BEAST v2.6to generate the mitochondrial

phylogeny. Here we used GTR+G as the nucleotide substitution model, a relaxed lognormal clock, and a birth-death tree prior. We conducted two independent MCMC runs, each for 10 million generations. We then visually inspected the convergence of each MCMC using Tracer v1.7.2 and discarded the initial 50% of each MCMC as burn-in. We randomly sampled 50 trees from the posterior of each MCMC run to generate a posterior distribution of 100 trees. We then used TreeAnnotator to generate the maximum clade credibility (MCC) tree using the posterior distribution of 100 trees.

4.3.3.8. Demographic history of *C. gloriosa* populations

We used SMC++ v1.15.2 to estimate the demographic history for each population of *C. gloriosa* (Terhorst et al., 2017). First, we subset the multi-sample VCF file into population-level using bcftools view module. Then we converted the VCF file to SMC format using the vcf2SMC module in SMC++. We created a separate SMC file for each individual and scaffold following the composite likelihood approach. The advantage of this approach is we can incorporate information from all the individuals for a given location when we estimate the demographic history instead of a single distinguished individual. We only used the largest 13 contigs, which capture 99.8% of the assembled genome. We then estimated the demographic history using all the SMC files for a given population using the estimate module in SMC++, and we repeated this process ten times for each population. We used a generation time of one year and a mutation rate of 2.1×10^{-9} to scale time and the effective population size. Like other recent and similar analyses of Coleoptera species demographics (Pélissié *et al.*, 2022) we used the mutation rate estimate from a recent publication on the non-biting midge (Oppold & Pfenninger, 2017).

4.3.3.9. Species distribution modeling

We downloaded all available occurrence data for *C. gloriosa* from iNaturalist and GBIF datasets (accessed March 2021). We removed data points that had no coordinates and duplicated data points. We supplemented this dataset with the Texas A&M insect collection datasets and our data from field visits which yielded a final dataset consisting of 776 occurrences.

Next, we downloaded climatic data for current and future climates from the Worldclime database (accessed May 2022). We used Worldclime dataset version 2.0 to download current and future climatic data (Fick & Hijmans, 2017). All data had a resolution of 30 arc seconds (i.e., each grid cell has an area of $1km^2$). For future climatic data we used Shared Socio-economic Pathways (SSP) 245 model. These SSPs represent greenhouse gas emission scenarios and associated climatic conditions. SSP245 represents current levels of greenhouse gas emissions. We downloaded predicted climatic data for the period of 2061 to 2080 (far future). These future climatic data come from the climatic modeling at the Goddard Institute for Space Studies (GISS) (Kelley *et al.*, 2020). Each data set for a given climatic condition consists of 19 bioclimatic variables. We also downloaded elevation data for the current climate through the WorldClime dataset.

We downloaded the individual shapefiles of Texas, New Mexico, and Arizona through the United States census bureau (https://www.census.gov/). We combined them to produce a single shapefile that uses ArcGIS to capture all three states. We then cropped the bioclimatic datasets to the extent of the study area and calculated the Pearson correlation between each bioclimatic variable using the R package virtualspecies (Leroy et al., 2016). We randomly selected a single

bioclimatic variable from those pairs with an absolute Pearson correlation coefficient greater than 0.7 for the species distribution modeling. This process retained only seven bioclimatic variables.

We used MaxEnt to model the species distribution for the current climate and, projected into the future climate (Phillips et al., 2006, 2004). We used the output format as Cloglog. The Cloglog format outputs the probability of occurrence for the given species. We ran for ten replicates with cross-validate as the replicate run type using a random starting seed for each replicate. We kept all other parameters at their default values. Finally, we used the mean from all replicates to analyze the final output.

4.4. Results

4.4.1. Assembly statistics

We build the *C. gloriosa* genome using Nanopore long reads and scaffolded with Omni-C data. The initial assembly using Nanopore data yielded 239 contigs spanning 642 MB at coverage of 44x, contig N50 of 8.7 MB, and the largest contig of 30.1 MB. We assessed the completeness of the assembly using BUSCO v 5.2.2 and used the endopterygota_odb10 dataset as the core set of 2124 single-copy orthologs genes. We find that 95.8% of the single copy orthologs are present in the assembled genome (94.6% complete single copy, 1.2% complete duplicated, 2.9% fragmented and 1.3% missing) (Figure 4.2).

We used blobtools to assess the level of contamination in the assembled genome and remove contigs that we identify as potential contaminants. We are specifically interested in identifying

and eliminating Prokaryotic contigs in this step. We downloaded the preformatted nucleotide database from NCBI and blasted the contigs of the *Chrysina gloriosa* genome against the NCBI nucleotide database. We then inspected the blast output to find contigs matching prokaryotic species. In our analysis, we find that only a single contig has a match with prokaryotic species. We removed this contig from the genome and used the new filtered genome for subsequent assembly processes (figure from blobtools).

We used 3d-DNA to scaffold the genome using Omni-C data and manual and automatic curation using Juicebox assembly tools and HiC-Hiker, respectively. The output from 3d-DNA consisted of 492 scaffolds with 19 scaffolds exceeding 1MB in size. Using JBAT, we placed these 19 scaffolds into ten scaffolds representing the ten chromosomes in a typical Scarab beetle karyotype. We also arranged smaller scaffolds when the connections were clear. This process reduced the total number of scaffolds to 454. The final assembly had an N50 of 72 MB, and the largest fragment size was 109 MB (before curation, the N50 was 37MB and the largest fragment size was 75 MB). The largest ten scaffolds covered 98.3% of the genome. When assessed for the completeness of the scaffolded assembly using BUSCO 5.2.2, we did not observe a significant change in the BUSCO scores in the scaffolded assembly (94.4% complete single copy, 1.1% complete duplicated, 2.9% fragmented and 1.6% missing).

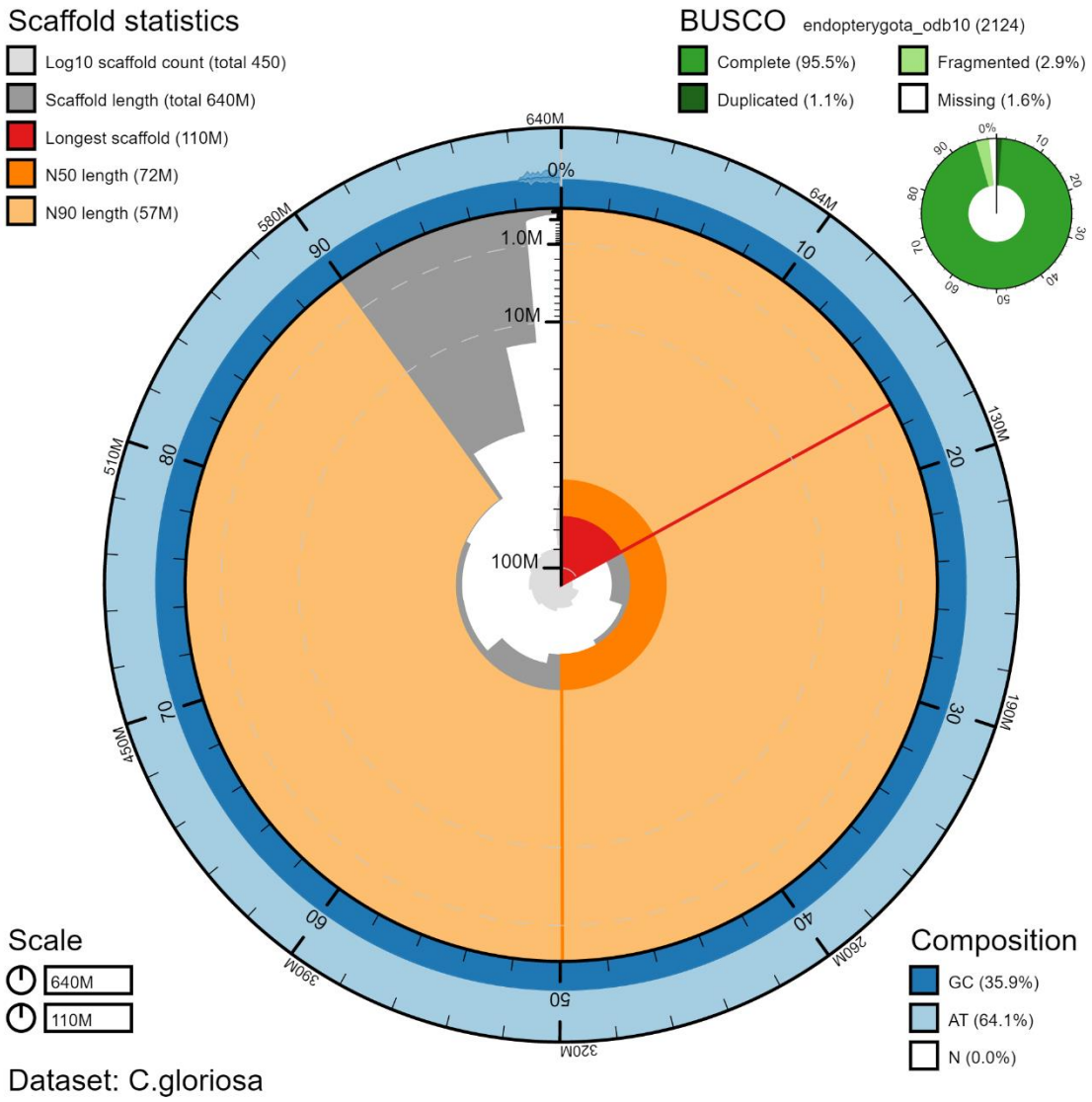


Figure 4.2 Assembly statistics and completeness of the assembled *Chrysina gloriosa* genome. Genome size is 640 Mbp, longest scaffold is 110 Mbp and N50 is 97 Mbp. GC composition is 35.9%. In BUSCO analysis, 95.5% of the orthologs are complete single or duplicated and 3.5% of the orthologs are fragmented or missing.

4.4.2. Distribution of repeats

We used Extensive *de novo* Transposable Element Annotator (EDTA) to annotate transposable elements (TE) and to mask the genome for repeat sequences. EDTA uses RepeatModler and

RepeatMasker internally to identify and mask repeat sequences. The low threshold cutoff resulted in 9.7% of the genome masking for repeats, while the high threshold cutoff masked 48.8% of the genome masking for repeats. For subsequent analysis we used the low threshold cutoff maksed genome as per developers' recommendation. However, for the synteny analysis we used high-threshold mased genome to minimize spurious alignments stemming from repeat regions.

All scaffolds show similar levels of repetitive content except scaffolds nine and ten (Figure 4.3). In both scaffolds nine and ten we see a relatively high percentage of macrosatellites while in scaffold ten we see relatively low distribution of DNA transposons, retrotransposons, and unclassified transposons. In all scaffolds except for scaffold seven and ten, we see high abundance of retrotransposons near the center of the scaffold indicative of the centromeric region (Figure 4.4). However, this signal is less clear in DNA transposons and unclassified transposons. We further see an abundance of macrosatellites near the ends of the scaffolds. In scaffold eight we see localization of retrotransposons towards the ends of the scaffolds. We further see an abundance of macrosatellites near the middle in scaffold eight.

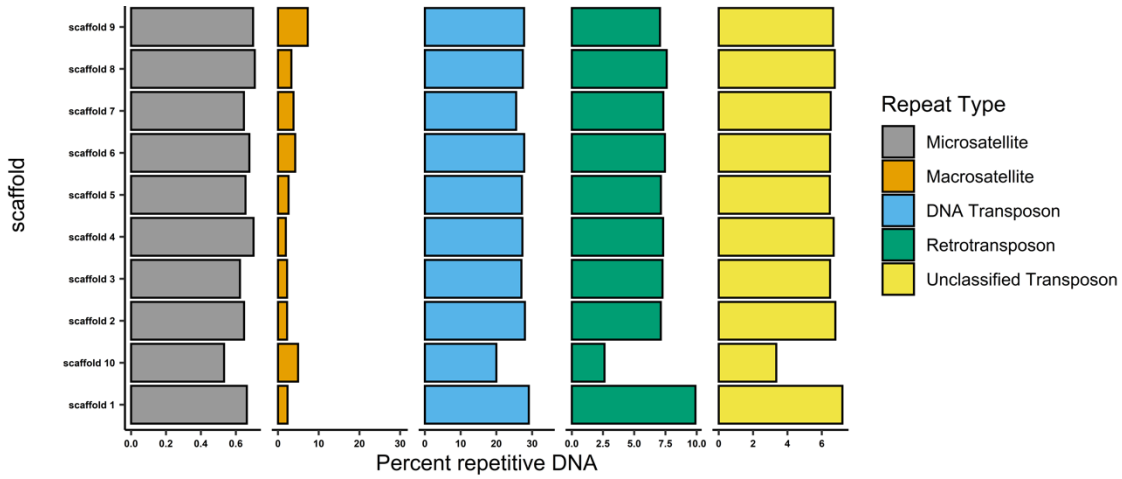


Figure 4.3 Percentage distribution of repeat classes across all scaffolds of the *C. gloriosa* genome.

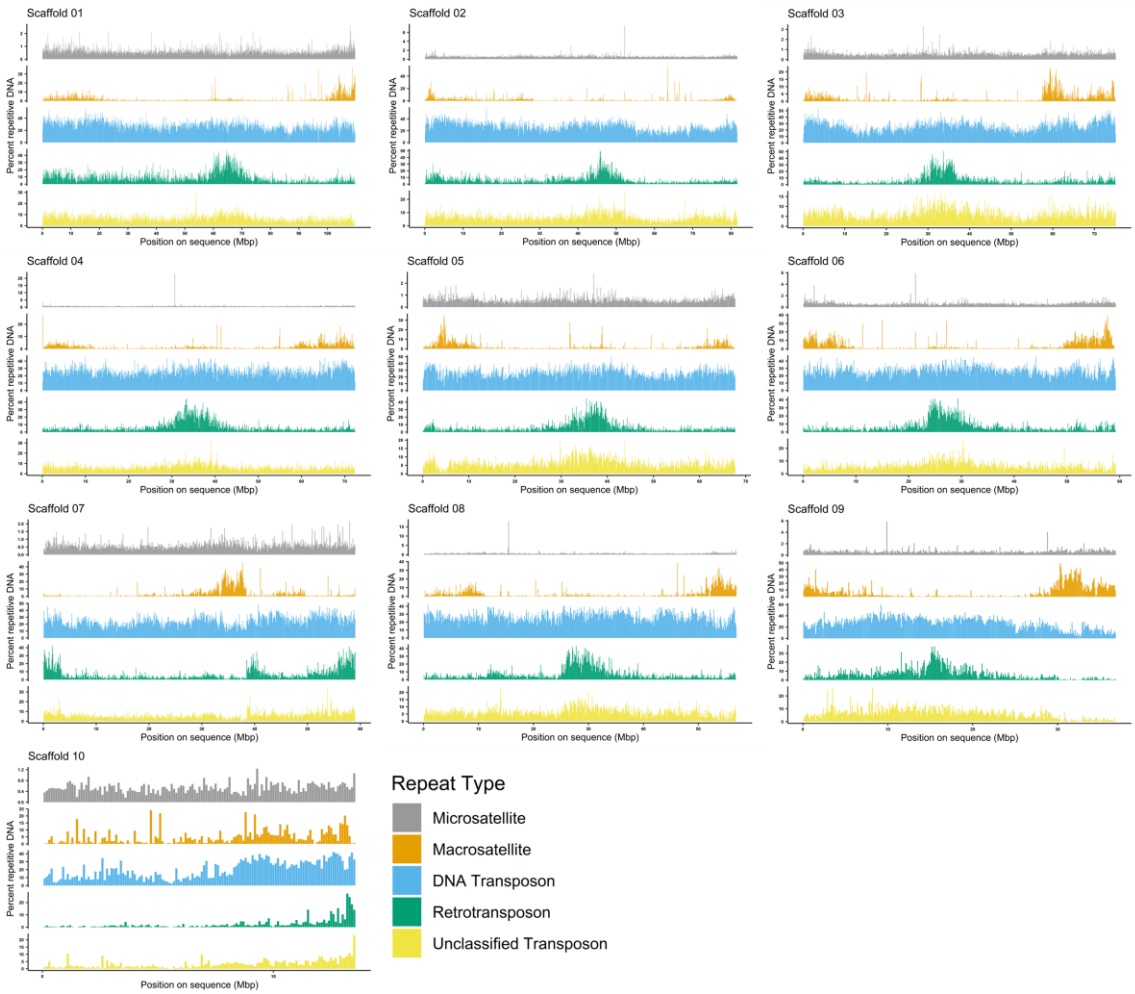


Figure 4.4 Distribution of repetitive content across all scaffolds in the *C. gloriosa* genome. We classified repetitive content into four categories: (from top to bottom of each panel) micro and macro-satellites, DNA transposons, retrotransposons, and unclassified transposons,

4.4.3. Genome synteny

We compared the *C. gloriosa* assembly with the genomes of *Tribolium castaneum* and *Trypoxylus dichotumus* (Herndon et al., 2020; Wang et al., 2022). We chose *T. castaneum* as it is a well curated genome assembly. However, *T. castaneum* and *C. gloriosa* have diverged from each other approximately 246 million of years ago (MYA) (Kumar et al., 2017). On the other hand, *T. dichotumus* is a more closely related species to *C. gloriosa* which have diverged 74

MYA from *C. gloriosa* (Kumar et al., 2017). In our comparisons with *T. castaneum* we see that LG2 of *T. castaneum* shows synteny with two scaffolds of *C. gloriosa* (Figure 4.5A). We further see that LG10 of *T. castaneum* matches with several scaffolds in *C. gloriosa*. In our comparison with *T. dichotumus* we find that each of the *C. gloriosa* scaffolds has a clear one to one orthology with scaffolds in the *T. dichotumus* assembly (Figure 4.5B). Despite this structural conservation we document frequent inversions within chromosomes.

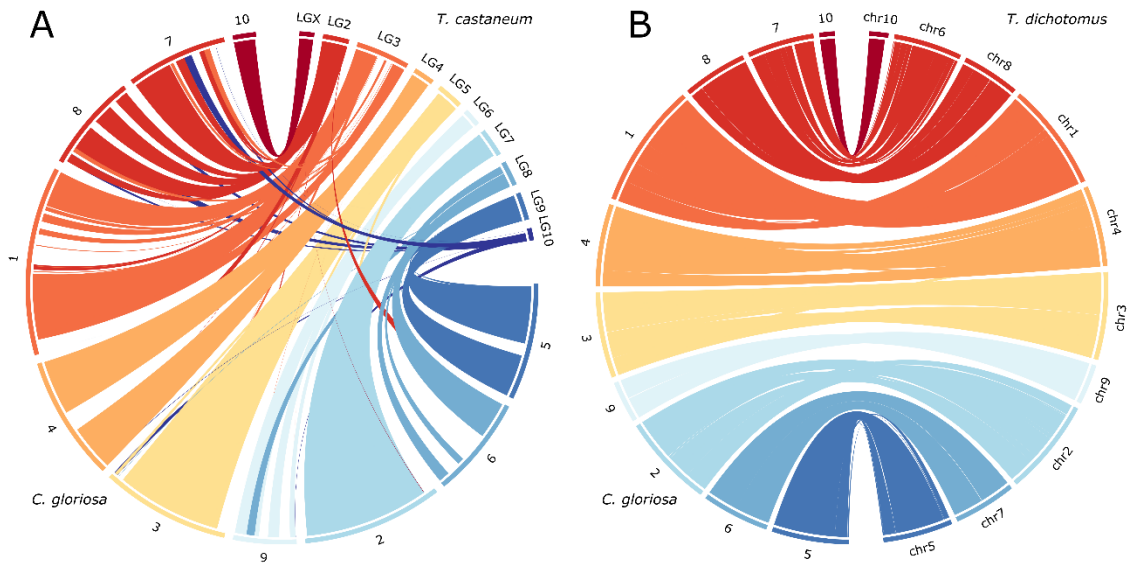


Figure 4.5 Circos plots comparing the *Chrysina gloriosa* genome with A) *Tribolium castaneum* genome and B) *Trypoxylus dichotomus* genome.

4.4.4. Identification of the X chromosome scaffolds and assigning sex

Once we normalized the mean coverage of each scaffold across all samples, we found that scaffold 10 had approximately 1x coverage in some specimens and approximately 0.5x coverage in others. We identify this scaffold as the X chromosome based on the expectation of male and female sequencing coverage of the X chromosome. When we Blast the *C. gloriosa* genome against the *Tribolium castaneum* genome, we find that Scaffold 10 of *C. gloriosa* matches well

with the X chromosome of the *Tribolium castaneum* genome, further confirming that scaffold 10 is the X chromosome of *C. gloriosa*. We could assign sex for all collected specimens based on the X chromosome scaffold coverage. We had 66 males and 15 females in our sample collection. (Figure 4.6). This sex ratio bias is likely a function of collecting specimens via lights as males are more likely to fly in search of females (Rodriguez-Soana & Miller, 2007; Tini et al., 2017). We suspect that slightly higher coverage observed in scaffold ten for females is due to the increased number of macrosatellites on this scaffold (Figure 4.3). If some of these macrosatellites are collapsed in the assembly, it would lead to an overestimation of the coverage relative to other chromosomes.

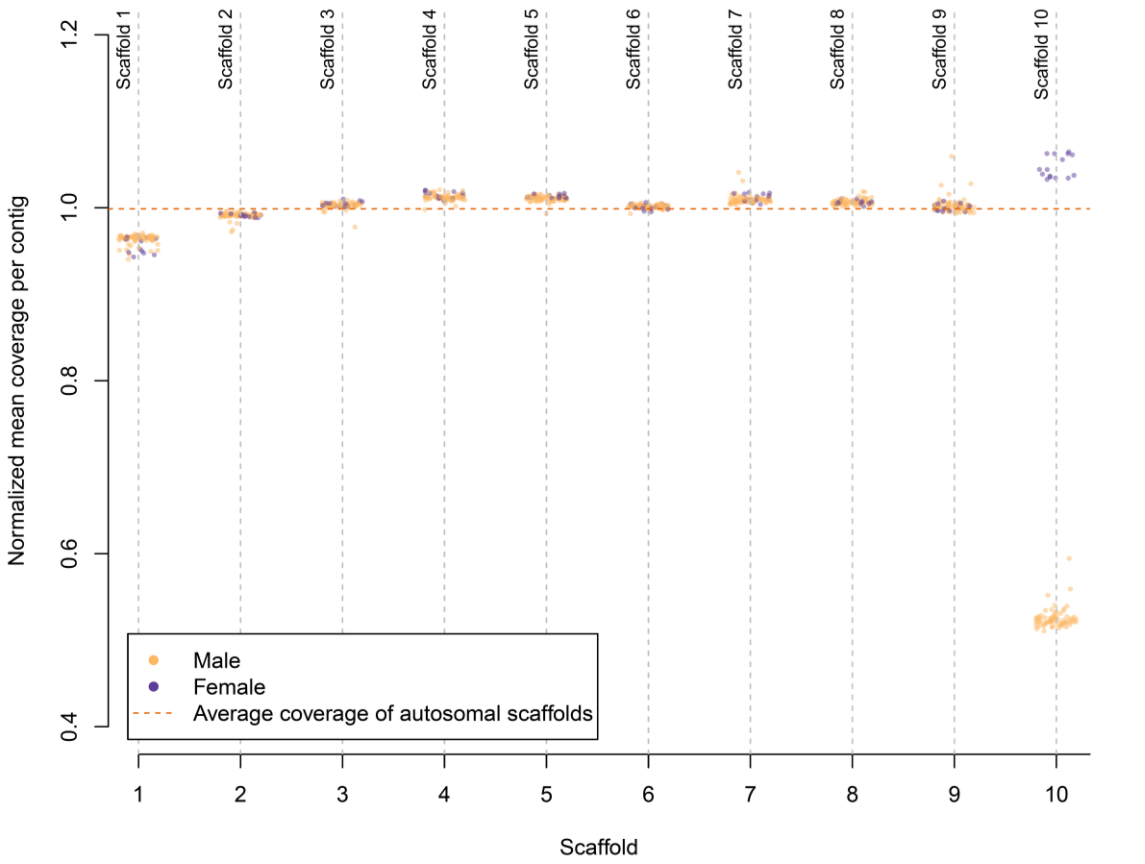


Figure 4.6 Mean normalized depth of each contig (size ordered) across all samples.

4.4.5. Assembly of the Y chromosome

We assembled the Y chromosome using reads from males unmapped to the assembled genome. We pooled all unmapped reads and assembled using ABySS at a range of kmer values and two multiplicity cutoff values and chose the best assembly based on N50, assembly size and the number of contigs. Our putative Y chromosome assembly includes 12307 contigs and 1031 contigs with a length greater than 500 bp. The largest contig size is 6490 bp, the N50 was 1096, and the total size of the assembly was 8.3 MB. The N50 and the total size statistics are for contigs greater than 500 bp (Appendix B: Table 6.3).

4.4.6. Annotation of the genome

We used the low threshold repeat masked genome from the EDTA pipeline for genome annotation. We used Liftoff to lift-over annotation from *Trypoxylus dichotomus*. Our annotation using Liftoff captured 4402 genes consisting of 25216 coding sequences and 18931 exons.

4.4.7. Assembly and annotation of the mitochondrial genome

We assembled the mitochondrial genome using Unicycler and annotated using the MITOS web server. Our assembly of *Chrysina gloriosa* mitochondrial genome was 17813 bases in size. The annotation consisted of 12 protein-coding genes, two ribosomal genes and 16 tRNA coding genes.

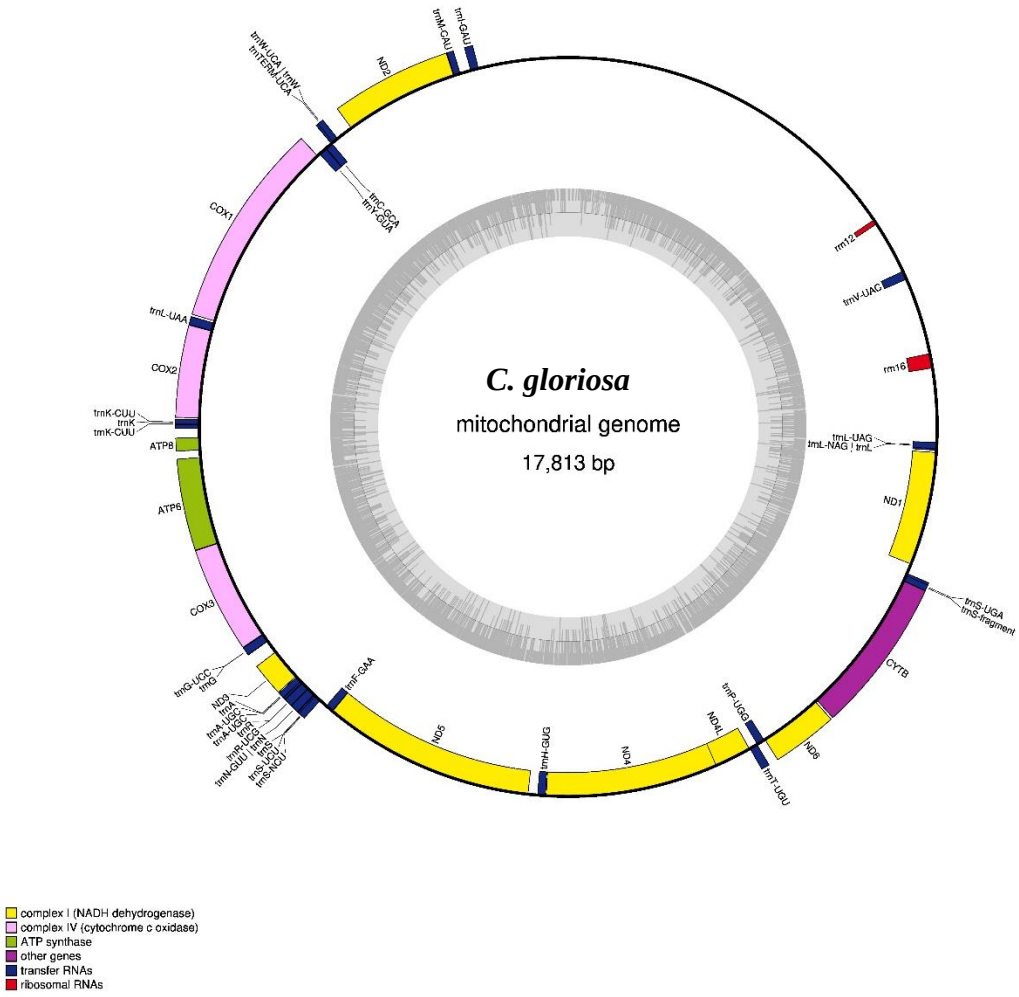


Figure 4.7 Annotation of the mitochondrial genome

4.4.8. Population structure using PCA

We employed both non-model-based and model-based approaches to infer the population structure of *C. gloriosa*. For the non-model-based approach, we used PCA to infer population structure. In addition, we further reduced the dimensionality of the PCA result by using the t-SNE approach.

Our nuclear dataset consisted of approximately 2 million SNPs. After filtering for linkage, the dataset consisted of 380,605 variants. Our inference of population structure using the nuclear dataset shows a clear separation between the Texas and Arizona specimens in the principal component 1 axis, which explains 22.3 % of the genetic variation in all samples (Figure 4.8A). We also find a clear separation of Arizona-only specimens, by the mountain range, along the PC2 axis, which explains 5.26% of the genetic variation within Arizona-only samples. However, we find that the Chiricahua Mountain and Geronimo Trail (Peloncillo Mountains) samples cluster together. We also performed a fine-scale population structure inference using t-SNE using the first five principal components. We see the same pattern observed in the PCA analysis but with a clear separation representing the sampling localities from Arizona. Chiricahua mountain samples and Peloncillo Mountains samples now have separate, though close clusters. We also see a within-mountain separation as well. For instance, we see the Huachuca Mountain specimens in two close clusters representing all samples from the west side of the mountain in one cluster and all samples from the east side of the mountain in the other (Figure 4.8B). This pattern is weakly observable in the results from PCA as well.

When we analyzed the population structure using mitochondria-only variants, we saw patterns similar to what we saw on the nuclear-only variants. Here we see a clear separation between the Arizona and Texas specimens along the PC1 axis, which explain 87.4% of the genetic variation in mitochondria samples (Figure 4.9A). The Arizona-only samples show a tight spread across the PC2 axis with no clear separation among these specimens. When we reduced dimensionality even further with the t-SNE approach, we started to see the separation in Arizona samples. Our results clearly distinguish the Chiricahua Mountain and the Peloncillo Mountains specimens into

two different clusters. We also see the Huachuca mountains and the Madera canyon specimens in a single cluster. However, this cluster consists of two subgroups. In one group, we see a mix of all Madera canyon specimens and Huachuca Mountain specimens. The other group only have specimens from the Huachuca mountains (Figure 4.9B).

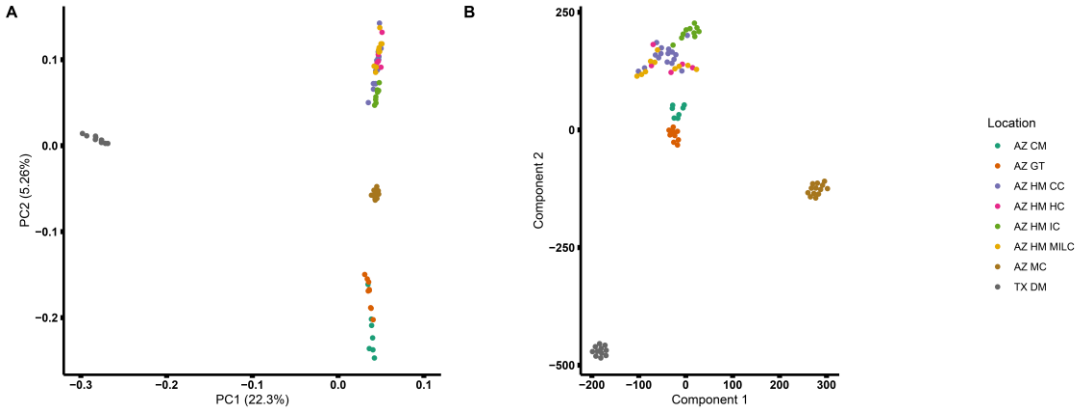


Figure 4.8 Population structure of *Chrysina gloriosa* using nuclear variants. A) Principal Component Analysis. B) t-SNE of first five principal components. AZ CM: Chiricahua Mountains, AZ GT: Geronimo Trail (Peloncillo Mountains), AZ HM CC: Huachuca Mountain Carr Canyon, AZ HM HC: Huachuca Mountain Hunter Canyon, AZ HM MILC: Huachuca Mountain Miller Canyon, AZ HM IC, Huachuca Mountain Ida Canyon, TX DM: Davis Mountain

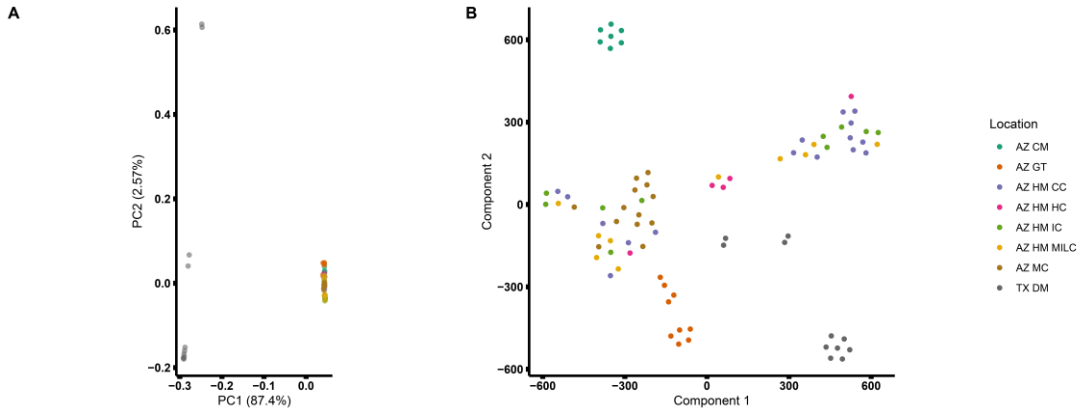


Figure 4.9 Population structure of *Chrysina gloriosa* using mitochondrial variants. A) Principal Component Analysis. B) t-SNE of first five principal components. AZ CM: Chiricahua Mountains, AZ GT: Geronimo Trail (Peloncillo Mountains), AZ HM CC: Huachuca Mountain Carr Canyon, AZ HM HC: Huachuca Mountain Hunter Canyon, AZ HM MILC: Huachuca Mountain Miller Canyon, AZ HM IC, Huachuca Mountain Ida Canyon, TX DM: Davis Mountain

4.4.9. Population structure using Admixture

We further inferred the population structure under a maximum likelihood approach using the program Admixture using nuclear and mitochondrial datasets. In the nuclear dataset, we see a clear separation of samples representing Arizona and Texas at K=2, representing Texas and Arizona populations. This pattern is consistent in the mitochondrial dataset as well (Figure 4.10). At K=4, we see no level of admixture in the Davis Mountain, Madera Canyon and Chiricahua Mountain populations representing three of the four ancestral populations. The Peloncillo Mountains population shows admixture from Chiricahua and Madera canyon populations with a relatively higher admixture from the Chiricahua population. The Huachuca Mountain samples representing the fourth ancestral population show some admixture from both Madera canyon and Chiricahua Mountain populations. Admixture from both these populations is much more evident in the Ida canyon specimens compared to the other specimens in the Huachuca mountains. The specimens from the other three sampling locations (east side of the mountain) in the Huachuca

Mountain have a relatively higher admixture from Madera canyon than from the Chiricahua Mountain population. We observe the same pattern at K=5 as well (Figure 4.10).

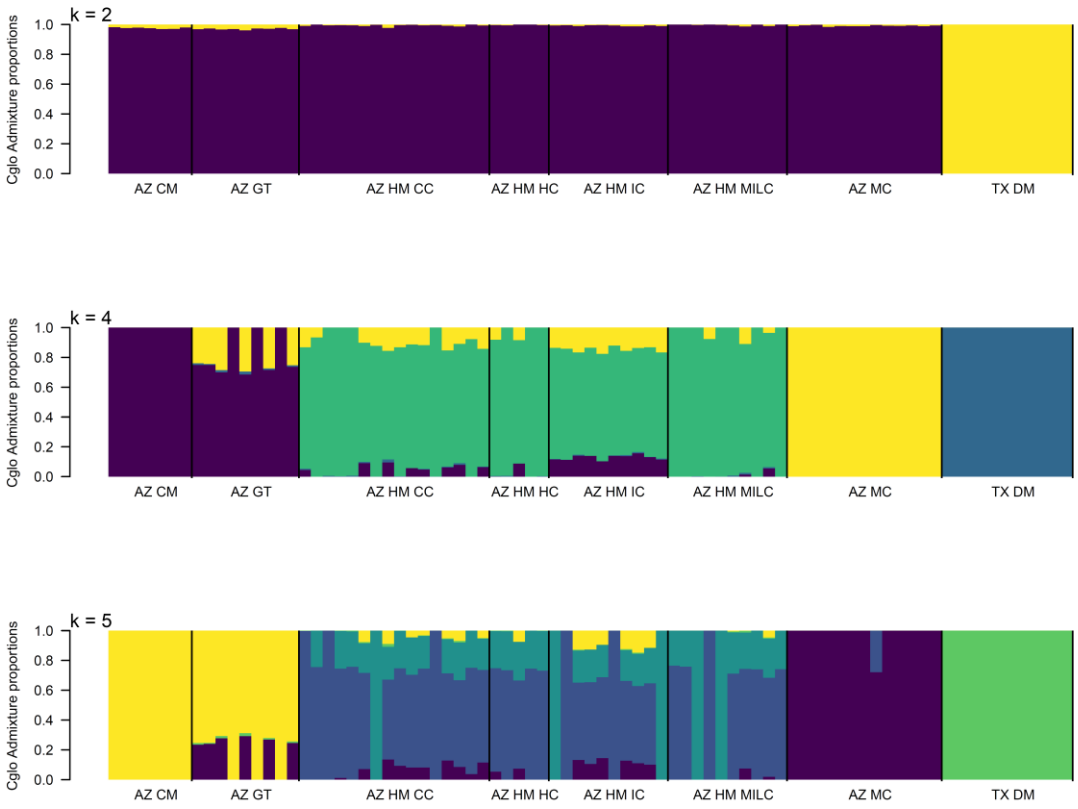


Figure 4.10 Bayesian clustering of *C. gloriosa* populations with ADMIXTURE software using the nuclear variants. From top to bottom bar plots of K = 2,4 and 5.

In the mitochondria sample, we see a different result at k=4 and k=5 compared to the nuclear dataset. At K=4, we still see no level of admixture in the Davis Mountain, Madera Canyon and Chiricahua Mountain populations representing three of the four ancestral populations. However, the Peloncillo Mountains samples show a pronounced admixture level from the Chiricahua Mountain populations. Admixture from Madera canyon is present but rare. In the Huachuca mountains representing the fourth ancestral population, we only see admixture from the Madera canyon population (Figure 4.11). At K=5, this pattern is mostly the same except in the

Chiricahua Mountain and Peloncillo Mountains populations. We now see a unique ancestral population representing the Peloncillo Mountains populations with a minor level of admixture from Madera Canyon and the Chiricahua Mountain populations.

Figure 4.11 Bayesian clustering of C. gloriosa populations with ADMIXTURE software using the mitochondrial variants. From top to bottom bar plots of K = 2,4 and 5.

4.4.10. Effective migration surface

We used FEEMs to estimate the migration surface of *C. gloriosa* across the Sky Islands of Arizona. We see increased probabilities of migration between the Chiricahua Mountains and the Peloncillo Mountains populations. In the Huachuca mountains samples, we estimate decreased rates of migration between the west and east sides of the mountain. We also see lower than

average migration between the Madera canyon populations and the Huachuca mountains population (Figure 4.12).

Figure 4.12 Effective migration surface of *C. gloriosa* specimens from Arizona. Cooler colors (blue) represent increased probability of migration while warmer colors (orange) represent decreased probability of migration.

4.4.11. Phylogeny reconstruction using mitochondria

We used BEAST to reconstruct the mitochondrial phylogeny under a Bayesian framework. We see that there are two clades representing Texas and Arizona. Within the Arizona clade, we see the Peloncillo Mountains samples in a single clade, while Chiricahua Mountain samples cluster together with a couple of samples from the Huachuca mountains. We also see that Madera Canyon samples are mixed with the Huachuca Mountain samples (Figure 4.13).

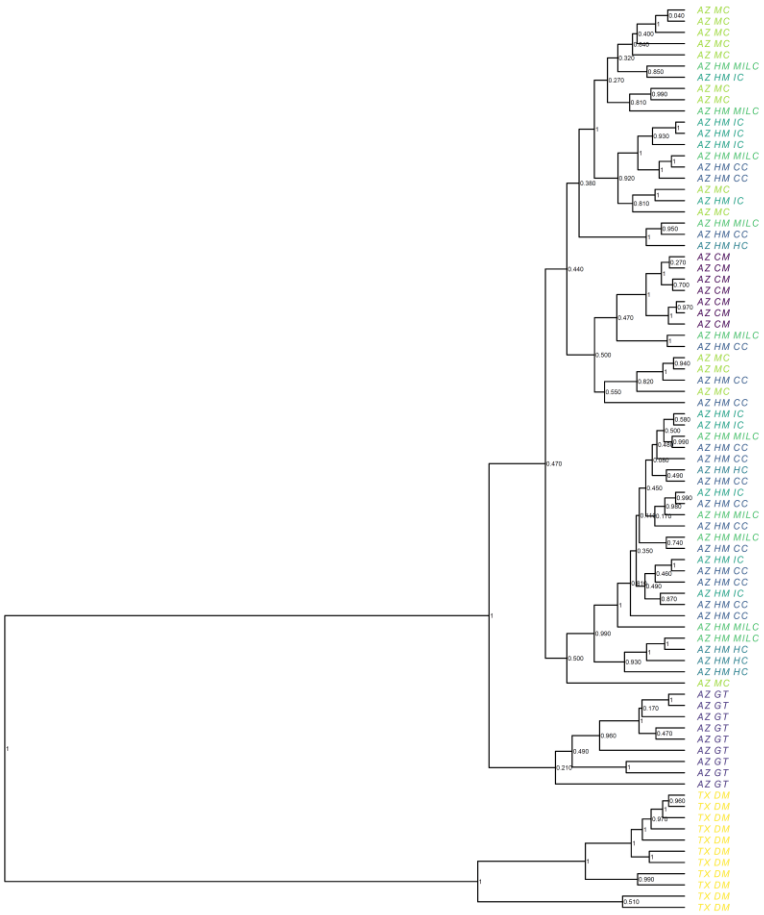


Figure 4.13 Maximum clade credibility tree build using mitochondrial variants. Each sampling location is given a unique color. Node values represent the posterior probability.

4.4.12. Demographic history of *C. gloriosa* populations

Our inference of the demographic history of *C. gloriosa* shows a decline in the effective population size from the late Pleistocene to the mid-Holocene. This pattern is consistent in all *C. gloriosa* populations. We also see that the demographic plots for all Arizona populations of *C. gloriosa* converge just prior to 10k years ago (Figure 4.15). This suggests that at this point all species were contained within a single largely panmictic population. This is consistent with the broader understanding of environmental changes across the southwest at the end of the last ice age (King & Van Devender, 1977). Notably the Texas and Arizona populations do not converge

suggesting that these portions of the range of *C. gloriosa* have been relatively isolated even during the last glacial period.

In our initial analyses we treated all populations separately in this analysis we inferred recent and striking increases in population size for three populations. However, a recent uptick in population size can be caused by recent migration introducing new alleles into a population. This along with other evidence of possible gene flow (previously discussed) suggested that the data used for demographic history of AZ CM and AZ GT should be combined as well as the data from AZ HM and AZ MC. In our final analysis of demographic history, we inferred the largest population size for AZ CM/GT population followed by the AZ HM/MC population. We inferred the smallest effective population size for TX DM samples. This may be a sign of its very long isolation from other *C. gloriosa* populations (Figure 4.15B).

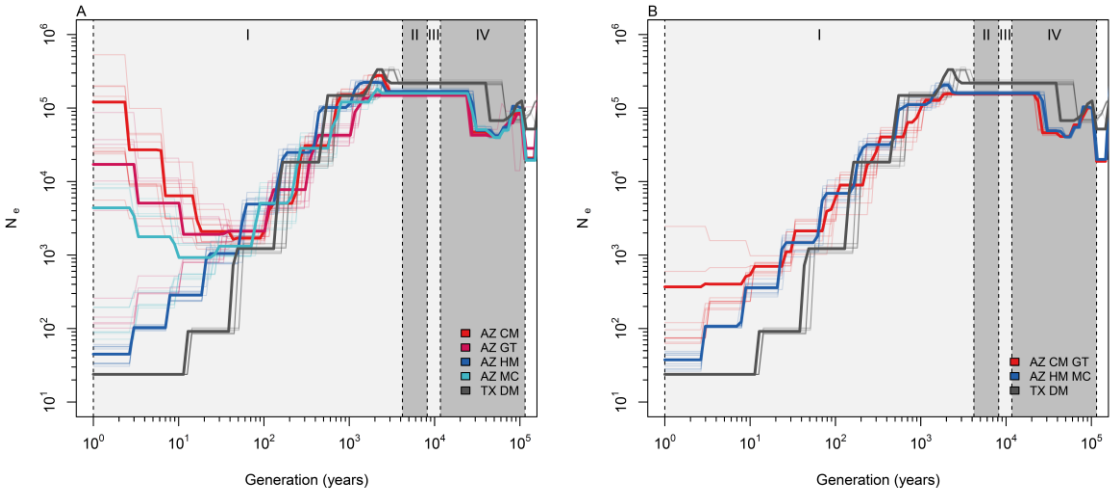


Figure 4.14 Demographic history of Chrysina gloriosa. AZ CM: Arizona Chiricahua Mountains, AZ GT: Arizona Geronimo Trail, AZ HM: Arizona Huachuca Mountains, AZ MC: Arizona Madera Canyon and TX DM: Texas Davis Mountains. A) all populations are considered as independent populations. B) AZ CM and AZ GT populations are considered a single population and, AZ HM and AZ MC populations are considered a single population. Gray shadings represent geological times; I: Late Holocene, II: Mid Holocene, III: Early Holocene, IV: Pleistocene

4.4.13. Species distribution of *C. gloriosa*

We used MaxEnt to build a model that shows the probability of occurrence for *C. gloriosa* for paleo, current and future climatic models obtained from Worldclime.org (Fick & Hijmans, 2017). For the future climatic conditions, we used predicted climatic model for the years 2061 to 2080 using the Shared Socio-economic Pathway 245 (SSP 245) model. Our projections into the future climatic conditions show a 98% reduction in the habitat that is identified as being at least 75% suitable for C. gloriosa at the current climatic conditions (Figure 4.16).

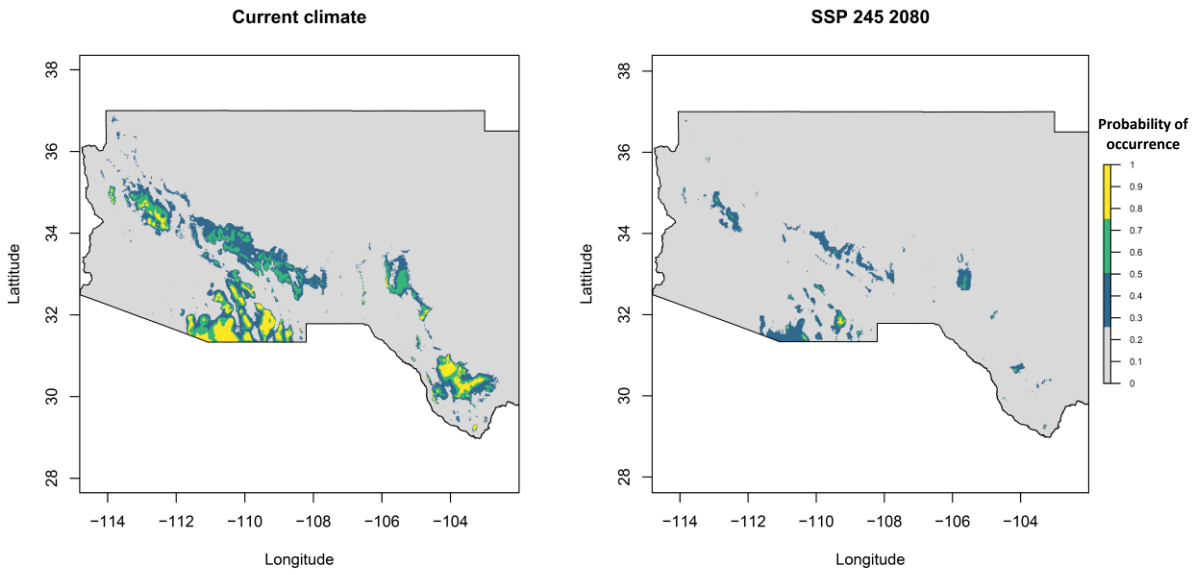


Figure 4.15 Species distribution model of *Chrysina gloriosa* at under the present climatic conditions and future climatic conditions. We chose Shared Socioeconomic Pathway (SSP) 245 as the future climatic model.

4.5. Discussion

4.5.1. Genome assembly of *C. gloriosa*

C. gloriosa is one of only four *Chrysina* species that have a distribution reaching the United States. Until now, no whole-genome data for any *Chrysina* species that reach the US existed, making it difficult to understand the evolution of, and assess the need for conservation of these charismatic beetles. We present the first chromosome-scale genome assembly, built using long-read sequencing and Omni-C data, for the species *C. gloriosa*. Our genome assembly consists of ten large scaffolds that capture 98.3% of the genome, representing a typical Scarabaeoidea's ten chromosomes (9 autosome pairs and an XY chromosome pair). To our knowledge, this is the only genome assembled to chromosome scale for the genus *Chrysina*. The only other genome assembly available in this genus is for *C. resplendens* which is assembled using short-read

sequencing data. The size of the *C. gloriosa* genome is slightly larger at 642 MB compared to the *C. resplendens* genome, which has a genome size of 611MB. However, these species have a smaller genome size than *C. woodii* (another *Chrysina* species that reach the US), whose genome size is estimated to be 856 MB using flow cytometry (Hanrahan & Johnston, 2011).

4.5.2. Genome annotation of *C. gloriosa*

To annotate the genome, we used Liftoff, which compares the genome assembly with an already annotated genome to lift over genome features. We used *Trypoxylus dichotomus* as the reference sequence. The divergence time between *Trypoxylus* and *Chrysina* is approximately 74 million years (MY), making *T. dichotomus* the most closely related species with an annotated genome to *C. gloriosa* (Kumar *et al.*, 2017). However, Liftoff could only match 36% of the *T. dichotomus* annotated genes with the *C. gloriosa* annotation. The low percentage of lift over is likely due to the divergence time of these two species being relatively large as Liftoff is designed to operate with the same or closely related species (Shumate & Salzberg, 2020). For comparison, we also lifted over annotations from two other genomes from GenBank. *Tribolium castaneum*, a well-curated beetle genome (accessed on 25th July 2022), *Onthophagus taurus*, only Scarabaeoidea genome in NCBI GeneBank with an annotation (accessed on 5th August 2022) (Herndon *et al.*, 2020; Thomas *et al.*, 2020). The divergence times between genera *Tribolium* and *Chrysina* is approximately 246 MY, and between genus Onthophagus and *Chrysina* is 183 MY (Kumar *et al.*, 2017). The *Tribolium castaneum* genome consists of 14,322 genes, and the *Onthophagus taurus* consists of 16,538 genes. However, we could lift over 4% and 6% genes from *T. castaneum* and *O. taurus* genomes.

4.5.3. Are the *C. gloriosa* populations isolated?

One of the overarching goals of this study is to find if there is gene flow between the populations of *C. gloriosa*, which have highly fragmented populations. Reduced gene flow coupled with population fragmentation can severely impact the survival of a population (Couvet, 2002). Due to population fragmentation, we expect to see a reduction in the effective population size, affecting the efficacy of natural selection and leading to the fixation of mildly deleterious mutations (Alò & Turner, 2005; Broeck et al., 2017; Lande, 1995). Furthermore, we would also see a reduction in heterozygosity and an increase in effective inbreeding caused by genetic drift leading to a higher risk of population extirpation (Saccheri *et al.*, 1998). Also, a lack of migration means that in the event of local extirpation, recolonization of suitable habitat is unlikely.

Our population structure analyses using the nuclear variant dataset show limited evidence for recent gene flow between the populations, except for Chiricahua and Peloncillo Mountain populations. These two mountains are the closest to each other at the southeastern edge of the Chiricahua Mountain and the northwestern edge of the Peloncillo Mountain separated by just 13km (8 miles). However, the sampling locations between these two mountains are separated by approximately 50km (31 miles).

One possibility for *C. gloriosa* to travel this distance is through the aid of the wind. Other species of beetles have been documented traveling distances greater than 20 km with the aid of wind (Byers, 2000; Chase et al., 2017; Evenden et al., 2014). Furthermore, the mating season of these beetles falls well within the monsoon season, potentially bringing strong wind currents. Therefore, we are not surprised to see gene flow between these two mountains. However, we see

a disagreement when we use the mitochondrial variant dataset. We see these two populations having independent clusters compared to a single cluster. This disagreement between the two datasets suggests that most, if not all, migrants between the two mountain ranges could be males. This male-biased migration is a possibility, given that 81% of our specimens are males. Male-bias in our samples could be due to the active flights of males in locating mates and this may predispose males to long distance dispersal events (as well as collection).

We see some discrepancies between the nuclear and mitochondrial datasets in the Huachuca Mountain and Madera canyon samples suggesting evidence for gene flow. For instance, our nuclear dataset sees specimens from these two mountains having unique clusters. However, in our mitochondrial dataset, we see four specimens from the Huachuca mountains mixed with Madera canyon samples. Additionally, our analysis of the demographic history shows a recent uptick in the effective population size in the Madera canyon, further providing evidence for a recent gene flow between Huachuca Mountain and Madera canyon. Furthermore, admixture in the mitochondrial dataset between Huachuca Mountain and Madera canyon suggest migration of females between these two mountains unlike between Chiricahua and Peloncillo Mountain, where we only see migration of males. In our sampling locations the two closest mountain ranges are the Santa Rita Mountains (Madera Canyon) and the Huachuca Mountains. At the southeastern edge of Santa Rita Mountains and the northwestern edge of the Huachuca Mountains, the two mountain ranges are separated by approximately 3.4km (2.1 miles). Therefore, we think that the proximity of the Santa Rita Mountains (Madera Canyon) and the Huachuca Mountains allow for migration of females.

We have sampled from several locations in the Huachuca mountains and Madera canyon. However, our sampling locations in Madera Canyon are close to each other, so our data does not suggest any gene flow restriction between these two locations. However, our sampling in the Huachuca mountains reflects the east and west distribution of *C. gloriosa* within this mountain range, allowing us to infer gene flow patterns within the mountain range. All our results point out that there is resistance to gene flow even between the west and east regions of Huachuca Mountain. One plausible cause for reduced gene flow between the west and east regions of Huachuca Mountain could be that these two regions are separated by a much higher mountain peak (approximately 3000 m). At these heights, the geo-climatic conditions differ substantially from their current distribution, which would present a barrier to gene flow between these two regions and may make dispersal by flight less likely.

Our results show that the *C. gloriosa* populations are isolated with marginal gene flow between the closest populations. We have a broad level of sampling from different mountain ranges from the sky islands of Arizona and the Davis Mountains of Texas. However, there are many more mountain ranges (e.g., Guadalupe mountains TX, Dragoon mountains AZ, Pene Blanca AZ) where additional sampling will provide a much broader picture of the population structure and gene flow among sky islands. One additional source of information that we are missing is genetic data in *C. gloriosa* populations from Mexico. Sky islands of Arizona are the northern tip of a much broader mountain range that reaches Mexico.

4.5.4. Historical distribution of flora in the Sky islands and the demographic history of *C. gloriosa*

Historical information about the vegetation in the Sonoran and Chihuahuan desert regions in Texas, New Mexico, and Arizona has been well documented by analyzing packrat middens (Holmgren et al., 2003; King & Van Devender, 1977; Van Devender et al., 1978, 1985). A packrat midden, an accumulation of wastes near or within the packrat den, consists of floral and faunal remnants, fecal matter, and small rocks. Due to the cementing action of packrat's urine, these middens have been well preserved for thousands of years (King & Van Devender, 1977). In addition, fauna and flora remnants within these middens indicate the composition of the local environment roughly within a 100 m radius as these packrats do not travel long distances to collect their food. Many middens that date back to the late Pleistocene epoch (~11,000 to ~129,000 years ago) indicate that the now arid low elevation regions in Arizona and New Mexico had abundant juniper, oak, and pine trees during this time (King & Van Devender, 1977). However, middens from similar elevations dated to the mid-Holocene period (5,000 to 7,000 years ago) show a lack of plant material from juniper oak and pine. Similarly, we see the climatic conditions in the Chiricahua desert in Northern Mexico were much wetter than now characterized by the rainfall in the winter in contrast to the tropical summer rainfall we presently see (Metcalfe et al., 2002). Furthermore, using core logging, we also see the disappearance of the pollen record for juniper trees around 11,000 years before present, providing evidence for change in climate conditions during the transition from Pleistocene to Holocene (Metcalfe et al., 2000). All this evidence suggests that plants such as juniper had a more continuous distribution throughout Southwestern Arizona and New Mexico. Lack of plant material from core logs and packrat middens dating to Holocene is evident in the expansion of the desert region to higher

altitudes and climate-driven floral displacement, resulting in a patchy distribution of juniper trees.

Chrysina gloriosa depends on juniper trees as a food source. Furthermore, *C. gloriosa* uses juniper trees as mating grounds. We hypothesize that the historical distribution of *C. gloriosa* was more of a panmictic distribution rather than a patchy distribution, with a much higher effective population size than the current populations (Young, 1957). Our analysis of the demographic history of *C. gloriosa* provides support for this hypothesis. We see a much higher effective population size for *C. gloriosa* during the early Pleistocene epoch, during which the earth's surface was at least 5℃ cooler than now (Burke *et al.*, 2018). With the ending of the last glacial period (~11000 years ago), we see a sudden decline in the population size of *C. gloriosa*, which coincides with the rise in global surface temperatures during the early Holocene period (Burke *et al.*, 2018). In addition, we think that local events such as the prominent wildfires 100 years ago could have shaped the recent demographic history, especially in the Huachuca mountains populations (Bahre, 1985).

Our analysis of population structure, migration surfaces, and demographic history suggests gene flow between the Chiricahua Mountain and Peloncillo Mountains populations. Therefore, we consider these two populations as a single large population. Considering the relatively low-level population connectivity that we document along with ongoing climate change and habitat destruction we would argue that continued surveillance of populations is warranted.

4.5.5. Climate impact on the *C. gloriosa* habitat size

We modeled the distribution of *C. gloriosa* under the current and future climatic conditions. Future climatic conditions are modeled based on greenhouse emission scenarios which are now known as Shared Socio-economic Pathways or SSPs. We chose SSP 245 as it represents the greenhouse emissions of the present world (Hausfather, 2018). Our projections into future climate conditions show a significant reduction in the habitat size. Most of the habitat that is suitable for C. gloriosa are located near the peaks of the Chiricahua mountains. This impact of climate change is not limited to *C. gloriosa*. Many species in these mountains share a similar habitat and we think that many if not all these species are highly impacted by the changes in the climate.

4.6. Conclusion

With the chromosome level genome assembly and population genetics data, we now have population genomic and demographic information for *C. gloriosa*, one of the most charismatic beetles in the United States. With this information, we can now make informed decisions on conserving this beetle.

Our analysis shows that most *C. gloriosa* populations are isolated, with some having limited recent gene flow. We also see evidence for male biased migration between mountains that are further apart. This may decrease the likelihood of recolonization of suitable habitat if most or all possible founders are male.

We also show that climatic conditions have severely impacted the historical population sizes of *C. gloriosa*, with the effective population size of all populations descending to less than 1000 individuals. Our species distribution models also show the severe impact of future global warming, caused by greenhouse gas emissions, on the potential habitat size, with the habitats at current elevations becoming severely reduced or eliminated.

5. CONCLUSIONS

The goal of this compilation is to understand broad scale and fine scale evolution of the genome. To understand the evolution of genomes at broad scale we used chromosome number from Polyneoptera and Amphibia. Both these clades show considerable variation in chromosome number. One of the key questions we asked was is the variation of chromosome number that we see within these clades are due to phylogenetic history or due to difference in the rates of chromosome number evolution. Our results show support for both these aspects. For example, In Polyneoptera, we see that the variation in chromosome number is due to differential rates of chromosome number evolution within these clades and, in Amphibia, we see that the variation in chromosome number is due to phylogenetic history. Our data further shows that there is no significant relationship between the genome size and the rates of chromosome number evolution.

The diversity of life history traits in these two clades further allow us to ask specific questions related to the impact of these life history traits on the evolution of chromosome number. In Polyneoptera, we see both sexual and asexual reproductive modes in the clade Phasmatodea. We see faster rates of Polyploidy with the transition into Parthenogenetic reproduction in Polyneoptera. In Amphibia there is a remarkable diversity of developmental modes. We can use these developmental modes as a proxy for effective population size and infer the fitness impact of the mutations that change chromosome number. Our analysis shows faster rates of chromosome number evolution in species with low effective population size in Amphibia suggesting that these mutations that change the chromosome number are deleterious.

To understand the evolution of genomes at the fine scale we used population genomic data of *Chrysina gloriosa*, a Jewel scarab which is restricted to high elevation forests in the southwestern United States. Furthermore, we put a particular emphasis on the necessity of conservation of *C. gloriosa* using our data. Our results show high degree of population structure in these *C. gloriosa* populations. We further see evidence for gene flow between some populations which are relatively close to each other. In populations with gene flow, we see that long distance migration heavily male biased. This evidence suggests that if a population becomes extirpated, it is highly unlikely for this region to become re colonized from a nearby mountain range. We also show impact of the past climatic changes on the demographic history of *C. gloriosa* as well as how future climatic conditions can have a severe impact on the *C. gloriosa* habitats. All these results show the importance of implementing immediate conservation practices to preserve the evolutionary potential of *C. gloriosa*.

REFERENCES

Adams, K. L., & Wendel, J. F. (2005). Polyploidy and genome evolution in plants. *Current Opinion in Plant Biology*, *8*(2), 135–141.

Alexander, D. H., Shringarpure, S. S., Novembre, J., & Lange, K. (2015). Admixture 1.3 software manual. *Los Angeles: UCLA Human Genetics Software Distribution*. https://dalexander.github.io/admixture/admixture-manual.pdf

Alò, D., & Turner, T. F. (2005). Effects of habitat fragmentation on effective population size in the endangered Rio Grande silvery minnow. *Conservation Biology: The Journal of the Society for Conservation Biology*, *19*(4), 1138–1148.

Anderson, N. W., Hjelmen, C. E., & Blackmon, H. (2020). The probability of fusions joining sex chromosomes and autosomes. *Biology Letters*, *16*(11), 20200648.

Bahre, C. J. (1985). Wildfire in southeastern Arizona between 1859 and 1890. *Desert Plants*. https://repository.arizona.edu/handle/10150/609084

Baker, R. J., & Bickham, J. W. (1986). Speciation by monobrachial centric fusions. *Proceedings of the National Academy of Sciences of the United States of America*, *83*(21), 8245–8248.

Beçak, M. L., Denaro, L., & Beçak, W. (1970). Polyploidy and mechanisms of karyotypic diversification in Amphibia. *Cytogenetics*, *9*(4), 225–238.

Bech, N., Boissier, J., Drovetski, S., & Novoa, C. (2009). Population genetic structure of rock ptarmigan in the ‘sky islands’ of French Pyrenees: implications for conservation. *Animal Conservation*, *12*(2), 138–146.

Bennetzen, J. L., Ma, J., & Devos, K. M. (2005). Mechanisms of recent genome size variation in flowering plants. *Annals of Botany*, *95*(1), 127–132.

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*, *27*(2), 573–580.

Bergamaschi, S., Dawes-Gromadzki, T. Z., Scali, V., Marini, M., & Mantovani, B. (2007). Karyology, mitochondrial DNA and the phylogeny of Australian termites. *Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology*, *15*(6), 735–753.

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., Pütz, J., Middendorf, M., & Stadler, P. F. (2013). MITOS: improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*, *69*(2), 313–319.

Blackman, Leather, & Hardie. (1995). Insect reproduction. *Sex Determination of Insects. CRS Press, Boca*.

Blackmon, H., & Brandvain, Y. (2017). Long-Term Fragility of Y Chromosomes Is Dominated by Short-Term Resolution of Sexual Antagonism. *Genetics*, *207*(4), 1621–1629.

Blackmon, H., & Demuth, J. P. (2014). Estimating tempo and mode of Y chromosome turnover: explaining Y chromosome loss with the fragile Y hypothesis. *Genetics*, *197*(2), 561–572.

Blackmon, H., & Demuth, J. P. (2015a). Coleoptera Karyotype Database. *The Coleopterists' Bulletin*, *69*(1), 174–175.

Blackmon, H., & Demuth, J. P. (2015b). The fragile Y hypothesis: Y chromosome aneuploidy as a selective pressure in sex chromosome and meiotic mechanism evolution. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, *37*(9), 942–950.

Blackmon, H., Justison, J., Mayrose, I., & Goldberg, E. E. (2019). Meiotic drive shapes rates of karyotype evolution in mammals. *Evolution; International Journal of Organic Evolution*, *73*(3), 511–523.

Blackmon, H., Ross, L., & Bachtrog, D. (2017). Sex Determination, Sex Chromosomes, and Karyotype Evolution in Insects. *The Journal of Heredity*, *108*(1), 78–93.

Bogart, J. P. (1980). Evolutionary implications of polyploidy in amphibians and reptiles. *Basic Life Sciences*, *13*, 341–378.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M. A., Rambaut, A., & Drummond, A. J. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, *10*(4), e1003537.

Boveri, T. (1904). *Ergebnisse über die Konstitution der chromatischen Substanz des Zellkerns*. G. Fischer.

Brady, P., & Cummings, M. (2010). Differential response to circularly polarized light by the jewel scarab beetle Chrysina gloriosa. *The American Naturalist*, *175*(5), 614–620.

Britton-Davidian, J., Sonjaya, H., Catalan, J., & Cattaneo-Berrebi, G. (1990). Robertsonian heterozygosity in wild mice: fertility and transmission rates in Rb(16.17) translocation heterozygotes. *Genetica*, *80*(3), 171–174.

Broeck, A. V., Maes, D., Kelager, A., & Wynhoff, I. (2017). Gene flow and effective population sizes of the butterfly Maculinea alcon in a highly fragmented, anthropogenic landscape. *Biologicals: Journal of the International Association of Biological Standardization*. https://www.sciencedirect.com/science/article/pii/S0006320716307923

Burke, K. D., Williams, J. W., Chandler, M. A., Haywood, A. M., Lunt, D. J., & Otto-Bliesner, B. L. (2018). Pliocene and Eocene provide best analogs for near-future climates.

Proceedings of the National Academy of Sciences of the United States of America, *115*(52), 13288–13293.

Byers, J. A. (2000). Wind-aided dispersal of simulated bark beetles flying through forests. *Ecological Modelling*, *125*(2), 231–243.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, *10*, 421.

Carbone, L., Harris, R. A., Gnerre, S., Veeramah, K. R., Lorente-Galdos, B., Huddleston, J., Meyer, T. J., Herrero, J., Roos, C., Aken, B., Anaclerio, F., Archidiacono, N., Baker, C., Barrell, D., Batzer, M. A., Beal, K., Blancher, A., Bohrson, C. L., Brameier, M., … Gibbs, R. A. (2014). Gibbon genome and the fast karyotype evolution of small apes. *Nature*, *513*(7517), 195–201.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, *17*(4), 540–552.

Cazier, M. A. (1951). *The genera Chrysina and Plusiotis of north central Mexico (Coleoptera, Scarabaeidae). American Museum novitates; no. 1516*. https://digitallibrary.amnh.org/bitstream/handle/2246/2371/N1516.pdf?sequence=1

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, *4*, 7.

Charlesworth, D., & Charlesworth, B. (1980). Sex differences in fitness and selection for centric fusions between sex-chromosomes and autosomes. *Genetical Research*, *35*(2), 205–214.

Chase, K. D., Kelly, D., Liebhold, A. M., Bader, M. K.-F., & Brockerhoff, E. G. (2017). Long-distance dispersal of non-native pine bark beetles from host resources. *Ecological Entomology*, *42*(2), 173–183.

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* , *34*(17), i884–i890.

Chesters, D. (2020). The phylogeny of insects in the data-driven era. *Systematic Entomology*, *45*(3), 540–551.

Chippindale, P. T., Bonett, R. M., Baldwin, A. S., & Wiens, J. J. (2004). Phylogenetic evidence for a major reversal of life-history evolution in plethodontid salamanders. *Evolution; International Journal of Organic Evolution*. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.0014-3820.2004.tb01632.x

Couvet, D. (2002). Deleterious effects of restricted gene flow in fragmented populations. *Conservation Biology: The Journal of the Society for Conservation Biology*, *16*(2), 369–376.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics* , *27*(15), 2156–2158.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2). https://doi.org/10.1093/gigascience/giab008

De Oliveira, & Moraes. (2015). Chromosomal evolution in Pleurothallidinae (Orchidaceae: Epidendroideae) with an emphasis on the genus Acianthera: chromosome numbers and

heterochromatin. *Botanical Journal of Scotland*. https://academic.oup.com/botlinnean/article-abstract/178/1/102/2416536

DeBano, L. F. (1995). *Biodiversity and Management of the Madrean Archipelago: The Sky Islands of Southwestern United States and Northwestern Mexico : September 19-23, 1994, Tucson, Arizona*. The Station.

Degrandi, T. M., Barcellos, S. A., Costa, A. L., Garnero, A. D. V., Hass, I., & Gunski, R. J. (2020). Introducing the Bird Chromosome Database: An overview of cytogenetic studies in birds. *Cytogenetic and Genome Research*, *160*(4), 199–205.

Dubois, A. (2004). Developmental pathway, speciation and supraspecific taxonomy in amphibians 1. Why are there so many frog species in Sri Lanka? *Alytes; Paris*, *22*(1), 19–37.

Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., Shamim, M. S., Machol, I., Lander, E. S., Aiden, A. P., & Aiden, E. L. (2017). De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science*, *356*(6333), 92–95.

Dudchenko, O., Shamim, M. S., Batra, S. S., Durand, N. C., Musial, N. T., Mostofa, R., Pham, M., St Hilaire, B. G., Yao, W., Stamenova, E., Hoeger, M., Nyquist, S. K., Korchina, V., Pletch, K., Flanagan, J. P., Tomaszewicz, A., McAloose, D., Estrada, C. P., Novak, B. J., … Aiden, E. L. (2018). The Juicebox Assembly Tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under $1000. In *bioRxiv* (p. 254797). https://doi.org/10.1101/254797

Duellman, W. E. (1989). Alternative life-history styles in anuran amphibians: evolutionary and ecological implications. In M. N. Bruton (Ed.), *Alternative Life-History Styles of Animals* (pp. 101–126). Springer Netherlands.

Dumont, B. L. (2017). Variation and Evolution of the Meiotic Requirement for Crossing Over in Mammals. *Genetics*, *205*(1), 155–168.

Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S. P., Huntley, M. H., Lander, E. S., & Aiden, E. L. (2016). Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Systems*, *3*(1), 95–98.

Dürrbaum, M., & Storchová, Z. (2016). Effects of aneuploidy on gene expression: implications for cancer. *The FEBS Journal*, *283*(5), 791–802.

Epstein, C. J. (1989). Down Syndrome. In *Abnormal States of Brain and Mind* (pp. 43–44). Birkhäuser Boston.

Evans, A. V. (2010). *REFLECTIONS ON ARIZONA'S JEWEL SCARABS-Part 1*. https://arthurevans.wordpress.com/2010/09/27/reflections-on-arizonas-jewel-scarabs-part-1/

Evans, B. J., Carter, T. F., Greenbaum, E., Gvoždík, V., Kelley, D. B., McLaughlin, P. J., Pauwels, O. S. G., Portik, D. M., Stanley, E. L., Tinsley, R. C., Tobias, M. L., & Blackburn, D. C. (2015). Genetics, morphology, advertisement calls, and historical records distinguish six new polyploid species of African clawed frog (Xenopus, Pipidae) from west and central Africa. *PloS One*, *10*(12), e0142823.

Evenden, M. L., Whitehouse, C. M., & Sykes, J. (2014). Factors influencing flight capacity of the mountain pine beetle (Coleoptera: Curculionidae: Scolytinae). *Environmental Entomology*, *43*(1), 187–196.

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* , *32*(19), 3047–3048.

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315.

FitzJohn, R. G. (2012). Diversitree : comparative phylogenetic analyses of diversification in R. *Methods in Ecology and Evolution*, *3*(6), 1084–1092.

Flemming, W. (1882). *Zellsubstanz, Kern und Zelltheilung*. Vogel.

Flemming, W. (1887). Weitere Beobachtungen über die Entwicklung der Spermatosomen bei Salamandra maculosa. *Archiv für mikroskopische Anatomie (1865)*, *31*(1), 71–97.

Freyman, W. A., & Höhna, S. (2018). Cladogenetic and Anagenetic Models of Chromosome Number Evolution: A Bayesian Model Averaging Approach. *Systematic Biology*, *67*(2), 195–215.

Fujita, M. K., & Moritz, C. (2009). Origin and evolution of parthenogenetic genomes in lizards: current state and future directions. *Cytogenetic and Genome Research*, *127*(2–4), 261–272.

Funk, W. C., Tallmon, D. A., & Allendorf, F. W. (1999). Small effective population size in the long-toed salamander. *Molecular Ecology*, *8*(10), 1633–1640.

Furness, A. I., Venditti, C., & Capellini, I. (2022). Terrestrial reproduction and parental care drive rapid evolution in the trade-off between offspring size and number across amphibians. *PLoS Biology*, *20*(1), e3001495.

Garagna, S., Broccoli, D., Redi, C. A., Searle, J. B., Cooke, H. J., & Capanna, E. (1995). Robertsonian metacentrics of the house mouse lose telomeric sequences but retain some minor satellite DNA in the pericentromeric area. *Chromosoma*, *103*(10), 685–692.

Genoways, H. H., & Baker, R. J. (1979). *Biological Investigations in the Guadalupe Mountains National Park, Texas: Proceedings of a Symposium Held at Texas Tech University, Lubbock, Texas, April 4-5, 1975*. National Park Service.

Gill, D. E. (1978). Effective population size and interdemic migration rates in a metapopulation of the red-spotted newt, Notophthalmus viridescens (rafinesque). *Evolution; International Journal of Organic Evolution*, *32*(4), 839.

Glick, L., & Mayrose, I. (2014). ChromEvol: assessing the pattern of chromosome number evolution and the inference of polyploidy along a phylogeny. *Molecular Biology and Evolution*, *31*(7), 1914–1922.

Gordon, J. L., Byrne, K. P., & Wolfe, K. H. (2011). Mechanisms of chromosome number evolution in yeast. *PLoS Genetics*, *7*(7), e1002190.

Gould, J., Beranek, C., Valdez, J., & Mahony, M. (2022). Quantity *versus* quality: A balance between egg and clutch size among Australian amphibians in relation to other life-history variables. *Austral Ecology*, *47*(3), 685–697.

Graf, N. (2019). *Unexpected Visitors*. Yavapai Gardens. https://extension.arizona.edu/sites/extension.arizona.edu/files/attachment/junjul2019.pdf

Green, D. M., Daugherty, C. H., & Bogart, J. P. (1980). Karyology and Systematic Relationships of the Tailed Frog Ascaphus truei. *Herpetologica*, *36*(4), 346–352.

Green, D. M., King, M., & John, B. (1991). Amphibia. Vol. 4: Chordata 2. Animal Cytogenetics. *Copeia*, *1991*(4), 1147.

Gregory. (2002). Animal genome size database. *Http://Www. Genomesize. Com*. https://cir.nii.ac.jp/crid/1570854175130267648

Greiner, S., Lehwark, P., & Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research*, *47*(W1), W59–W64.

Hanken, J. (1999). Chapter 3 - Larvae in Amphibian Development and Evolution. In B. K. Hall & M. H. Wake (Eds.), *The Origin and Evolution of Larval Forms* (pp. 61–IV). Academic Press.

Hanrahan, S. J., & Johnston, J. S. (2011). New genome size estimates of 134 species of arthropods. *Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology*, *19*(6), 809–823.

Harton, G. L., & Tempest, H. G. (2012). Chromosomal disorders and male infertility. *Asian Journal of Andrology*, *14*(1), 32–39.

Hausfather. (2018). How "shared socioeconomic pathways" explore future climate change. *Carbon Brief: Clear on Climate*.

Hawks, D. C. (2002). *Jewel Scarabs*. https://digitalcommons.unl.edu/museumprogram/15/

Hernández-Pacheco, R., Sutherland, C., Thompson, L. M., & Grayson, K. L. (2019). Unexpected spatial population ecology of a widespread terrestrial salamander near its southern range edge. *Royal Society Open Science*, *6*(6), 182192.

Herndon, N., Shelton, J., Gerischer, L., Ioannidis, P., Ninova, M., Dönitz, J., Waterhouse, R. M., Liang, C., Damm, C., Siemanowski, J., Kitzmann, P., Ulrich, J., Dippel, S., Oberhofer, G., Hu, Y., Schwirz, J., Schacht, M., Lehmann, S., Montino, A., … Bucher, G. (2020).

Enhanced genome assembly and a new official gene set for Tribolium castaneum. *BMC Genomics*, *21*(1), 47.

Ho, L. si T., & Ané, C. (2014). A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Systematic Biology*, *63*(3), 397–408.

Holmgren, C. A., Penalba, M. C., Rylander, K. A., & Betancourt, J. L. (2003). A 16,000 14C yr BP packrat midden series from the USA--Mexico Borderlands. *Quaternary Research*, *60*(3), 319–329.

Hornsey, K. G. (1973). The Occurrence of Hexaploid Plants Among Autotetraploid Populations of Sugar Beet, (Beta Vulgaris L.) and the Production of Tetraploid Progeny using a Diploid Pollinator. *Caryologia*, *26*(2), 225–228.

Howe, K., Chow, W., Collins, J., Pelan, S., Pointon, D.-L., Sims, Y., Torrance, J., Tracey, A., & Wood, J. (2021). Significantly improving the quality of genome assemblies through curation. *GigaScience*, *10*(1). https://doi.org/10.1093/gigascience/giaa153

Hu, J., Fan, J., Sun, Z., & Liu, S. (2020). NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* , *36*(7), 2253–2255.

Hughes-Schrader, S. (1950). The chromosomes of mantids (Orthoptera; Manteidae) in relation to taxonomy. *Chromosoma*, *4*(1), 1–55.

Hughes-Schrader, Sally. (1959). On the cytotaxonomy of phasmids (Phasmatodea). *Chromosoma*, *10*(1–6), 268–277.

Imai, H. T., Takahata, N., Maruyama, T., Daniel, A., Honda, T., Matsuda, Y., & Moriwaki, K. (1988). Theoretical bases for karyotype evolution. II. The fusion burst in man and mouse. *The Japanese Journal of Genetics*, *63*(4), 313–342.

Janssens, S. B., Couvreur, T. L. P., Mertens, A., Dauby, G., Dagallier, L.-P. M. J., Vanden Abeele, S., Vandelook, F., Mascarello, M., Beeckman, H., Sosef, M., Droissart, V., van der Bank, M., Maurin, O., Hawthorne, W., Marshall, C., Réjou-Méchain, M., Beina, D., Baya, F., Merckx, V., … Hardy, O. (2020). A large-scale species level dated angiosperm phylogeny for evolutionary and ecological analyses. *Biodiversity Data Journal*, *8*, e39677.

Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K., & Mooers, A. O. (2012). The global diversity of birds in space and time. *Nature*, *491*(7424), 444–448.

Jetz, Walter, & Pyron, R. A. (2018). The interplay of past diversification and evolutionary isolation with present imperilment across the amphibian tree of life. *Nature Ecology & Evolution*, *2*(5), 850–858.

Johnston, J. S., Bernardini, A., & Hjelmen, C. E. (2019). Genome size estimation and quantitative cytogenetics in insects. *Methods in Molecular Biology (Clifton, N.J.)*, *1858*, 15–26.

Katoh, K., Rozewicki, J., & Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, *20*(4), 1160–1166.

Kazitsa, Wu, Wei, Pu, Wu, & Song. (2018). Population Size, Genetic Diversity and Molecular Evidence of a Recent Population Bottleneck in Hynobius chinensis, an Endangered Salamander Species. *Asian Herpetological Research*. https://core.ac.uk/download/pdf/199410526.pdf

Kelley, M., Schmidt, G. A., Nazarenko, L. S., Bauer, S. E., Ruedy, R., Russell, G. L., Ackerman, A. S., Aleinov, I., Bauer, M., Bleck, R., Canuto, V., Cesana, G., Cheng, Y., Clune, T. L.,

Cook, B. I., Cruz, C. A., Del Genio, A. D., Elsaesser, G. S., Faluvegi, G., … Yao, M.-S. (2020). GISS-E2.1: Configurations and climatology. *Journal of Advances in Modeling Earth Systems*, *12*(8). https://doi.org/10.1029/2019ms002025

Kidwell, M. G. (2002). Transposable elements and the evolution of genome size in eukaryotes. *Genetica*, *115*(1), 49–63.

King, J. E., & Van Devender, T. R. (1977). Pollen Analysis of Fossil Packrat Middens from the Sonoran Desert. *Quaternary Research*, *8*(2), 191–204.

Kitano, J., & Peichel, C. L. (2012). Turnover of sex chromosomes and speciation in fishes. *Environmental Biology of Fishes*, *94*(3), 549–558.

Knowles, Gillespie, & Clague. (2009). Sky islands. *Encyclopedia of Earth Sciences*. http://www.johnemccormack.com/uploads/1/3/9/2/139271240/mccormack_2009_sky_islands_review.pdf

Kobel, H. R., & Du Pasquier, L. (1986). Genetics of polyploid Xenopus. *Trends in Genetics: TIG*, *2*, 310–315.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S. J., & Marra, M. A. (2009). Circos: an information aesthetic for comparative genomics. *Genome Research*, *19*(9), 1639–1645.

Kumar, S., Stecher, G., Suleski, M., & Hedges, S. B. (2017). TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Molecular Biology and Evolution*, *34*(7), 1812–1819.

Kuro-o, M., Ikebe, C., & Kohno, S. (1987). Cytogenetic studies of Hynobiidae (Urodela). *Cytogenetic and Genome Research*, *44*(2–3), 69–75.

Lachowska, & Holecova. (1998). Karyotypic data on weevils (Coleoptera, Curculionidae). *Folia Biologica*. https://books.google.com/books?hl=en&lr=&id=5D4g-GNWFt0C&oi=fnd&pg=PA129&ots=M9lcTADW6j&sig=X6tpdDWS_Wg1Gx2YVx65QbtohHs

Laetsch, D. R., & Blaxter, M. L. (2017). BlobTools: Interrogation of genome assemblies. *F1000Research*, *6*(1287), 1287.

Lande, R. (1995). Mutation and conservation. *Conservation Biology: The Journal of the Society for Conservation Biology*, *9*(4), 782–791.

LeConte. (1854). Descriptions of new Coleoptera collected by Thos. H. Webb, MD, in the years 1850–51 and 52, while Secretary to the US and Mexican Boundary …. *Proceedings. Academy of Natural Sciences of Philadelphia*.

Leroy, B., Meynard, C. N., Bellard, C., & Courchamp, F. (2016). virtualspecies, an R package to generate virtual species distributions. *Ecography*, *39*(6), 599–607.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. In *arXiv [q-bio.GN]*. arXiv. http://arxiv.org/abs/1303.3997

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* , *34*(18), 3094–3100.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* , *25*(14), 1754–1760.

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* , *26*(5), 589–595.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* , *25*(16), 2078–2079.

Li, Z., Tiley, G. P., Galuska, S. R., Reardon, C. R., Kidder, T. I., Rundell, R. J., & Barker, M. S. (2018). Multiple large-scale gene and genome duplications during the evolution of hexapods. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(18), 4713–4718.

Liedtke, H. C., Gower, D. J., Wilkinson, M., & Gomez-Mestre, I. (2018). Macroevolutionary shift in the size of amphibian genomes and the role of life history and climate. *Nature Ecology & Evolution*, *2*(11), 1792–1799.

Liehr, T., Buleu, O., Karamysheva, T., Bugrov, A., & Rubtsov, N. (2017). New Insights into Phasmatodea Chromosomes. *Genes*, *8*(11). https://doi.org/10.3390/genes8110327

Lippe, B. (1991). Turner syndrome. *Endocrinology and Metabolism Clinics of North America*, *20*(1), 121–152.

Liu, Y., Chen, C., Xu, Z., Scuoppo, C., Rillahan, C. D., Gao, J., Spitzer, B., Bosbach, B., Kastenhuber, E. R., Baslan, T., Ackermann, S., Cheng, L., Wang, Q., Niu, T., Schultz, N., Levine, R. L., Mills, A. A., & Lowe, S. W. (2016). Deletions linked to TP53 loss drive cancer through p53-independent mechanisms. *Nature*, *531*(7595), 471–475.

Lo, J., Jonika, M. M., & Blackmon, H. (2019). micRocounter: Microsatellite Characterization in Genome Assemblies. *G3* , *9*(10), 3101–3104.

Lockstone, H. E., Harris, L. W., Swatton, J. E., Wayland, M. T., Holland, A. J., & Bahn, S. (2007). Gene expression profiling in the adult Down syndrome brain. *Genomics*, *90*(6), 647–660.

Lokki, J., & Saura, A. (1979). Polyploidy in insect evolution. *Basic Life Sciences*, *13*, 277–312.

Luykx, P. (1990). A cytogenetic survey of 25 species of lower termites from Australia. *Genome / National Research Council Canada = Genome / Conseil National de Recherches Canada*, *33*(1), 80–88.

Maddison, W. P., & Maddison, D. R. (2019). *Mesquite: a modular system for evolutionary analysis. Version 3.51. 2018*.

Marcus, J., Ha, W., Barber, R. F., & Novembre, J. (2021). Fast and flexible estimation of effective migration surfaces. *ELife*, *10*. https://doi.org/10.7554/eLife.61927

Mathien, F. J. (1995). The Chimney Rock Archaeological Symposium: 1990 October 20–21; Durango, Colorado. J. McKim Malville and Gary Matlock, editors. General Technical Report RM-227. U.S. Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experiment Station, Fort Collins, Colorado, 1993. 106 pp., figures, tables, references. Free. *American Antiquity*, *60*(1), 180–181.

Mayrose, I., Barker, M. S., & Otto, S. P. (2010). Probabilistic models of chromosome number evolution and the inference of polyploidy. *Systematic Biology*, *59*(2), 132–144.

McCann, J., Schneeweiss, G. M., Stuessy, T. F., Villaseñor, J. L., & Weiss-Schneeweiss, H. (2016). The impact of reconstruction methods, phylogenetic uncertainty and branch lengths on inference of chromosome number evolution in American daisies (Melampodium, Asteraceae). *PloS One*, *11*(9), e0162299.

Meisel, R. P., Delclos, P. J., & Wexler, J. R. (2019). The X chromosome of the German cockroach, Blattella germanica, is homologous to a fly X chromosome despite 400 million years divergence. *BMC Biology*, *17*(1), 100.

Metcalfe, O'Hara, & Caballero. (2000). Records of Late Pleistocene–Holocene climatic change in Mexico—a review. *Quaternary Science Reviews*. https://www.sciencedirect.com/science/article/pii/S0277379199000220

Metcalfe, S., Say, A., Black, S., McCulloch, R., & O'Hara, S. (2002). Wet conditions during the last glaciation in the Chihuahuan Desert, Alta Babicora basin, Mexico. *Quaternary Research*, *57*(1), 91–101.

Miga, K. H. (2017). Chromosome-Specific Centromere Sequences Provide an Estimate of the Ancestral Chromosome 2 Fusion Event in Hominin Genomes. *The Journal of Heredity*, *108*(1), 45–52.

Miller, Pfeiffer, & Schwartz. (2010). Gateway computing environments workshop (GCE), 2010. *Creating the CIPRES Science Gateway For*.

Morelli, Blackmon, & Hjelmen. (2022). Diptera and Drosophila karyotype databases: a useful dataset to guide evolutionary and genomic studies. *Frontiers in Ecology and Evolution*. https://cehjelmen.github.io/assets/publications/morelli_diptera_frontiers_2022.pdf

Morescalchi, A. (1975). *Chromosome evolution in the caudate Amphibia*. https://pascal-francis.inist.fr/vibad/index.php?action=getRecordDetail&idt=PASCAL7650240204

Morescalchi, A., Odierna, G., & Olmo, E. (1979). Karyology of the primitive slamanders, family hynobiidae. *Experientia*, *35*(11), 1434–1436.

Moretti, A., & Sabato, S. (1984). Karyotype evolution by centromeric fission inZamia (Cycadales). *Plant Systematics and Evolution = Entwicklungsgeschichte Und Systematik Der Pflanzen*, *146*(3), 215–223.

Morón. (1990). The beetles of the world. Vol. 10: Rutelini 1. *Sciences Nat*.

Nagpure, N. S., Pathak, A. K., Pati, R., Rashid, I., Sharma, J., Singh, S. P., Singh, M., Sarkar, U. K., Kushwaha, B., Kumar, R., & Murali, S. (2016). Fish Karyome version 2.1: a chromosome database of fishes and other aquatic organisms. *Database: The Journal of Biological Databases and Curation*, *2016*, baw012.

Oliveira, B. F., São-Pedro, V. A., Santos-Barrera, G., Penone, C., & Costa, G. C. (2017). AmphiBIO, a global database for amphibian ecological traits. *Scientific Data*, *4*(1), 170123.

Oppold, A.-M., & Pfenninger, M. (2017). Direct estimation of the spontaneous mutation rate by short-term mutation accumulation lines in Chironomus riparius. *Evolution Letters*, *1*(2), 86–92.

Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J. R. A., Hellinga, A. J., Lugo, C. S. B., Elliott, T. A., Ware, D., Peterson, T., Jiang, N., Hirsch, C. N., & Hufford, M. B. (2019). Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biology*, *20*(1), 275.

Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* , *35*(3), 526–528.

Pease, J. B., & Hahn, M. W. (2012). Sex chromosomes evolved from independent ancestral linkage groups in winged insects. *Molecular Biology and Evolution*, *29*(6), 1645–1653.

Pélissié, B., Chen, Y. H., Cohen, Z. P., Crossley, M. S., Hawthorne, D. J., Izzo, V., & Schoville, S. D. (2022). Genome Resequencing Reveals Rapid, Repeated Evolution in the Colorado Potato Beetle. *Molecular Biology and Evolution*, *39*(2). https://doi.org/10.1093/molbev/msac016

Perkins, R. D., Gamboa, J. R., Jonika, M. M., Lo, J., Shum, A., Adams, R. H., & Blackmon, H. (2019). A database of amphibian karyotypes. *Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology*, *27*(4), 313–319.

Perry, J., Slater, H. R., & Choo, K. H. A. (2004). Centric fission--simple and complex mechanisms. *Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology*, *12*(6), 627–640.

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*(3), 231–259.

Phillips, S. J., Dudík, M., & Schapire, R. E. (2004). A maximum entropy approach to species distribution modeling. *Proceedings of the Twenty-First International Conference on Machine Learning*, 83.

Porter, C. A., & Sites, J. W., Jr. (1985). Normal disjunction in Robertsonian heterozygotes from a highly polymorphic lizard population. *Cytogenetic and Genome Research*, *39*(4), 250–257.

R Core Team. (2016). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria*. URL https://www.R-project.org/.

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). *Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Systematic biology67 (5): 901-904*.

Ratomponirina, C., Brun, B., Rumpler, Y., & Brandham, P. E. (1988). *Kew Chromosome Conference III*.

Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution / British Ecological Society*, *2*, 217–223.

Rice, A., Glick, L., Abadi, S., Einhorn, M., Kopelman, N. M., Salman-Minkov, A., Mayzel, J., Chay, O., & Mayrose, I. (2015). The Chromosome Counts Database (CCDB) - a community resource of plant chromosome numbers. *The New Phytologist*, *206*(1), 19–26.

Rieseberg, L. H. (2001). Chromosomal rearrangements and speciation. *Trends in Ecology & Evolution*, *16*(7), 351–358.

Ritcher, P. O. (1966). *White grubs and their allies: a study of North American scarabaeoid larvae*. ir.library.oregonstate.edu. https://ir.library.oregonstate.edu/downloads/nk322k13b

Rockman, M. V., & Rowell, D. M. (2002). Episodic chromosomal evolution in Planipapillus (Onychophora: Peripatopsidae): a phylogenetic approach to evolutionary dynamics and speciation. *Evolution; International Journal of Organic Evolution*, *56*(1), 58–69.

Rodriguez-Soana, & Miller. (2007). Behaviors of Adult Agrilus Planipennis (Coleoptera: Buprestidae). *The Great Lakes*. https://scholar.valpo.edu/tgle/vol40/iss1/1/

Ross, L., Blackmon, H., Lorite, P., Gokhman, V. E., & Hardy, N. B. (2015). Recombination, chromosome number and eusociality in the Hymenoptera. *Journal of Evolutionary Biology*, *28*(1), 105–116.

Ruckman, S. N., Jonika, M. M., Casola, C., & Blackmon, H. (2020). Chromosome number evolves at equal rates in holocentric and monocentric clades. *PLoS Genetics*, *16*(10), e1009076.

Saccheri, I., Kuussaari, M., Kankare, M., Vikman, P., Fortelius, W., & Hanski, I. (1998). Inbreeding and extinction in a butterfly metapopulation. *Nature*, *392*(6675), 491–494.

Sanderson, M. J., Boss, D., Chen, D., Cranston, K. A., & Wehe, A. (2008). The PhyLoTA Browser: processing GenBank for molecular phylogenetics research. *Systematic Biology*, *57*(3), 335–346.

Sekar, S., & Karanth, P. (2013). Flying between sky islands: The effect of naturally fragmented habitat on butterfly population structure. *PloS One*, *8*(8), e71573.

Sessions, S. K. (2008). Evolutionary cytogenetics in salamanders. *Chromosome Research: An International Journal on the Molecular, Supramolecular and Evolutionary Aspects of Chromosome Biology*, *16*(1), 183–201.

Shaffer, L., & Theisen. (2010). Disorders caused by chromosome abnormalities. *The Application of Clinical Genetics*, 159.

Sharma, V., Crne, M., Park, J. O., & Srinivasarao, M. (2009). Structural origin of circularly polarized iridescence in jeweled beetles. *Science*, *325*(5939), 449–451.

Sherman, P. W. (1979). Insect Chromosome Numbers and Eusociality. *The American Naturalist*, *113*(6), 925–935.

Shumate, A., & Salzberg, S. L. (2020). Liftoff: accurate mapping of gene annotations. *Bioinformatics* . https://doi.org/10.1093/bioinformatics/btaa1016

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* , *31*(19), 3210–3212.

Simons, A. (2010). A quality control tool for high throughput sequence data. *A Quality Control Tool for High Throughput Sequence Data*.

Smith, S. A., & Brown, J. W. (2018). Constructing a broadly inclusive seed plant phylogeny. *American Journal of Botany*, *105*(3), 302–314.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* , *30*(9), 1312–1313.

Stebbins, G. L. (1958). Longevity, habitat. and release of genetic variability in the higher plants. *Cold Spring Harbor Symposia on Quantitative Biology*, *23*, 365–378.

Stebbins, G. L. (1971). Chromosomal evolution in higher plants. *Chromosomal Evolution in Higher Plants.* https://www.cabdirect.org/cabdirect/abstract/19711606614

Steinemann, S., & Steinemann, M. (2005). Y chromosomes: born to be destroyed. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, *27*(10), 1076–1083.

Sun, L., Johnson, A. F., Donohue, R. C., Li, J., Cheng, J., & Birchler, J. A. (2013). Dosage compensation and inverse effects in triple X metafemales of Drosophila. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(18), 7383–7388.

Sutton, W. S. (1902). ON THE MORPHOLOGY OF THE CHROMOSO GROUP IN BRACHYSTOLA MAGNA. *The Biological Bulletin*, *4*(1), 24–39.

Sved, J. A., Chen, Y., Shearman, D., Frommer, M., Gilchrist, A. S., & Sherwin, W. B. (2016). Extraordinary conservation of entire chromosomes in insects over long evolutionary periods. *Evolution; International Journal of Organic Evolution*, *70*(1), 229–234.

Sylvester, T., Hjelmen, C. E., Hanrahan, S. J., Lenhart, P. A., Johnston, J. S., & Blackmon, H. (2020). Lineage-specific patterns of chromosome evolution are the rule not the exception in Polyneoptera insects. *Proceedings. Biological Sciences*, *287*(1935), 20201388.

Terhorst, J., Kamm, J. A., & Song, Y. S. (2017). Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nature Genetics*, *49*(2), 303–309.

Thomas, G. W. C., Dohmen, E., Hughes, D. S. T., Murali, S. C., Poelchau, M., Glastad, K., Anstead, C. A., Ayoub, N. A., Batterham, P., Bellair, M., Binford, G. J., Chao, H., Chen, Y. H., Childers, C., Dinh, H., Doddapaneni, H. V., Duan, J. J., Dugan, S., Esposito, L. A., … Richards, S. (2020). Gene content evolution in the arthropods. *Genome Biology*, *21*(1), 15.

Tini, M., Bardiani, M., Campanaro, A., Chiari, S., Mason, F., Maurizi, E., Toni, I., Audisio, P., & Carpaneto, G. M. (2017). A stag beetle's life: sex-related differences in daily activity and behaviour of Lucanus cervus (Coleoptera: Lucanidae). *Journal of Insect Conservation*, *21*(5–6), 897–906.

Tree of Sex Consortium. (2014). Tree of Sex: a database of sexual systems. *Scientific Data*, *1*, 140015.

Tymowska. (1991). Polyploidy and cytogenetic variation in frogs of the genus Xenopus. *Amphibian Cytogenetics and Evolution*. https://books.google.com/books?hl=en&lr=&id=A_sngwUBLxcC&oi=fnd&pg=PA259&dq=polyploidy+and+cutogenetic+variance+in+xenopus&ots=x3zC1ZAPlL&sig=d8VZpsPRJ7wkCeUb-Nb2RXJEjMo

Upham, N. S., Esselstyn, J. A., & Jetz, W. (2019). Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation. *PLoS Biology*, *17*(12), e3000494.

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research: JMLR*, *9*(11). https://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf?fbcl

Van Devender, T. R., Freeman, C. E., & Worthington, R. D. (1978). Full-Glacial and Recent Vegetation of Livingston Hills, Presidio County, Texas. *The Southwestern Naturalist*, *23*(2), 289–301.

Van Devender, T. R., Martin, P. S., Thompson, R. S., Cole, K. L., Jull, A. J. T., Long, A., Toolin, L. J., & Donahue, D. J. (1985). Fossil packrat middens and the tandem accelerator mass spectrometer. *Nature*, *317*(6038), 610–613.

Vicoso, B., & Bachtrog, D. (2015). Numerous transitions of sex chromosomes in Diptera. *PLoS Biology*, *13*(4), e1002078.

Voss, S. R., Kump, D. K., Putta, S., Pauly, N., Reynolds, A., Henry, R. J., Basa, S., Walker, J. A., & Smith, J. J. (2011). Origin of amphibian and avian chromosomes by fission, fusion, and retention of ancestral chromosomes. *Genome Research*, *21*(8), 1306–1312.

Wake, & Hanken. (2004). Direct development in the lungless salamanders: what are the consequences for developmental biology, evolution and phylogenesis? *The International Journal of Developmental Biology*. http://www.ijdb.ehu.es/web/paper.php?doi=8877460

Wang, Q., Liu, L., Zhang, S., Wu, H., & Huang, J. (2022). A chromosome-level genome assembly and intestinal transcriptome of Trypoxylus dichotomus (Coleoptera: Scarabaeidae) to understand its lignocellulose digestion ability. *GigaScience*, *11*. https://doi.org/10.1093/gigascience/giac059

White, M. J. D. (1954). *Animal cytology & evolution*. Cambridge university press.

White, M. J. D. (1978a). *Modes of speciation*. WH Freeman.

White, M. J. D. (1978b). Chain Processes in Chromosomal Speciation. *Systematic Biology*, *27*(3), 285–298.

Wick, R. R., Judd, L. M., Gorrie, C. L., & Holt, K. E. (2017). Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Computational Biology*, *13*(6), e1005595.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686.

Williams, B. R., Prabhu, V. R., Hunter, K. E., Glazier, C. M., Whittaker, C. A., Housman, D. E., & Amon, A. (2008). Aneuploidy affects proliferation and spontaneous immortalization in mammalian cells. *Science*, *322*(5902), 703–709.

Wodsedalek, J. E. (1916). CAUSES OF STERILITY IN THE MULE. *The Biological Bulletin*, *30*(1), 1-[56]-1.

Young, F. N. (1957). Notes on the Habits of Plusiotis gloriosa Le conte (Scarabaeidae). *The Coleopterists' Bulletin*, *11*(3/4), 67–70.

Zenil-Ferguson, R., Burleigh, J. G., & Ponciano, J. M. (2018). chromploid: An R package for chromosome number evolution across the plant tree of life. *Applications in Plant Sciences*, *6*(3), e1037.

Zenil-Ferguson, R., Ponciano, J. M., & Burleigh, J. G. (2017). Testing the association of phenotypes with polyploidy: An example using herbaceous and woody eudicots. *Evolution; International Journal of Organic Evolution*, *71*(5), 1138–1148.

Zhao, W.-W., Wu, M., Chen, F., Jiang, S., Su, H., Liang, J., Deng, C., Hu, C., & Yu, S. (2015). Robertsonian translocations: an overview of 872 Robertsonian translocations identified in a diagnostic laboratory in China. *PloS One*, *10*(5), e0122647.

APPENDIX A

ADDITIONAL SUPPLEMENTARY MATERIAL OF CHAPTER 2

Table 0.1 1C genome sizes for Polyneoptera. All new records were run on a Partec Cyflow SL_3 cytometer with Periplaneta americana (1C = 3338 Mbp) used as a standard. AGSD; Animal genome size database.

Order	Family	Genus	Species	Sex	GS (Mbp)	Rep	StDev	Source
Blattodea	Blattellidae	*Blattella*	*germanica*	F	2161.66	3	45.2	This manuscript
Blattodea	Blattellidae	*Blattella*	*germanica*	M	2090.7	3	32.5	This manuscript
Blattodea	Blattidae	*Periplaneta*	*fuliginosa*	-	4498.9	3	91.4	This manuscript
Mantodea	Mantidae	*Stegomantis*	*sp*	M	3461.51	1	-	This manuscript
Mantodea	Mantidae	*Thesprotia*	*graminis*	M	2071.5	1	-	This manuscript
Orthoptera	Acrididae	*Acantherus*	*piperatus*	F	14287.9	1	-	This manuscript
Orthoptera	Acrididae	*Acrolophitus*	*hirtipes*	F	14813.4	1	-	This manuscript
Orthoptera	Acrididae	*Arethea*	*sp*	M	8539.8	1	-	This manuscript
Orthoptera	Acrididae	*Arphia*	*simplex*	F	11823.3	2	744.25	This manuscript
Orthoptera	Acrididae	*Arphia*	*simplex*	M	11134.8	6	436.25	This manuscript
Orthoptera	Acrididae	*Bootettix*	*argentatus*	F	11998.2	2	25.79	This manuscript
Orthoptera	Acrididae	*Bootettix*	*argentatus*	M	11401.1	2	159.19	This manuscript
Orthoptera	Acrididae	*Brachystola*	*magna*	F	16807	1	-	This manuscript
Orthoptera	Acrididae	*Chortophaga*	*viridifasciata*	F	10435.6	4	409.95	This manuscript
Orthoptera	Acrididae	*Chortophaga*	*viridifasciata*	M	10279.7	3	168.65	This manuscript
Orthoptera	Acrididae	*Dactylotum*	*bicolor*	M	10211.2	1	-	This manuscript
Orthoptera	Acrididae	*Encoptolophus*	*costalis*	F	8449	1	-	This manuscript
Orthoptera	Acrididae	*Encoptolophus*	*costalis*	M	7971.5	3	96.42	This manuscript
Orthoptera	Acrididae	*Hadrotettix*	*trifasciatus*	F	18051.1	1	-	This manuscript
Orthoptera	Acrididae	*Hadrotettix*	*trifasciatus*	M	17495.7	2	1010.54	This manuscript
Orthoptera	Acrididae	*Hippuscus*	*ocelote*	F	13810.4	2	739.17	This manuscript
Orthoptera	Acrididae	*Hippuscus*	*ocelote*	M	11940.9	3	245.45	This manuscript
Orthoptera	Acrididae	*Lactista*	*azteca*	F	10007.5	2	12.84	This manuscript
Orthoptera	Acrididae	*Lactista*	*azteca*	M	9555.5	1	-	This manuscript
Orthoptera	Acrididae	*Leprus*	*wheeleri*	F	13968.1	2	58.8	This manuscript
Orthoptera	Acrididae	*Leprus*	*wheeleri*	M	13466.6	7	449.09	This manuscript
Orthoptera	Acrididae	*Melanoplus*	*bispinosus*	M	5813	1	-	This manuscript
Orthoptera	Acrididae	*Melanoplus*	*differentialis*	F	7096.2	3	56.75	This manuscript
Orthoptera	Acrididae	*Melanoplus*	*differentialis*	M	6668.9	3	61.97	This manuscript
Orthoptera	Acrididae	*Melanoplus*	*femurrubrum*	F	6011.2	6	403.34	This manuscript
Orthoptera	Acrididae	*Melanoplus*	*femurrubrum*	M	5277.1	3	88.29	This manuscript
Orthoptera	Acrididae	*Mermiria*	*bivittata*	F	15430.5	9	671.76	This manuscript
Orthoptera	Acrididae	*Mermiria*	*bivittata*	M	14496.9	5	213.22	This manuscript
Orthoptera	Acrididae	*Psinidia*	*amplicornis*	F	9325.4	4	209.71	This manuscript
Orthoptera	Acrididae	*Psinidia*	*amplicornis*	M	8824.1	2	334.41	This manuscript
Orthoptera	Acrididae	*Schistocerca*	*americana*	M	8259.15	1	-	This manuscript
Orthoptera	Acrididae	*Schistocerca*	*nitens*	F	9251.34	1	-	This manuscript
Orthoptera	Acrididae	*Schistocerca*	*obscura*	M	9318.86	1	-	This manuscript
Orthoptera	Acrididae	*Spharagemon*	*cristatum*	F	10129.4	2	24.02	This manuscript
Orthoptera	Acrididae	*Spharagemon*	*cristatum*	M	9774.37	1	-	This manuscript
Orthoptera	Acrididae	*Spharagemon*	*equale*	M	12260.4	1	-	This manuscript
Orthoptera	Acrididae	*Syrbula*	*admirabilis*	F	9589.9	1	-	This manuscript
Orthoptera	Acrididae	*Syrbula*	*admirabilis*	M	9201	1	-	This manuscript
Orthoptera	Acrididae	*Syrbula*	*montezuma*	F	11765.8	1	-	This manuscript
Orthoptera	Acrididae	*Syrbula*	*montezuma*	M	11151.9	2	494.83	This manuscript
Orthoptera	Acrididae	*Trachyrachys*	*kiowa*	F	9357.4	1	-	This manuscript
Orthoptera	Acrididae	*Trachyrachys*	*kiowa*	M	8842.2	3	68.59	This manuscript
Orthoptera	Acrididae	*Trimerotropis*	*pallidipennis*	F	9219	1	-	This manuscript
Orthoptera	Acrididae	*Trimerotropis*	*pallidipennis*	M	9112.3	2	197.87	This manuscript
Orthoptera	Acrididae	*Trimerotropis*	*sp*	F	12134.2	1	-	This manuscript
Orthoptera	Acrididae	*Xanthippus*	*corallipes*	F	13759.3	1	-	This manuscript
Orthoptera	Acrididae	*Xanthippus*	*corallipes*	M	12809.9	1	-	This manuscript
Orthoptera	Acrididae	*Xanthippus*	*corallipes pantherinus*	F	12765.2	1	-	This manuscript
Orthoptera	Tettigoniidae	*Neobarrettia*	*spinosa*	F	5328.1	2	49.96	This manuscript
Orthoptera	Tettigoniidae	*Neobarrettia*	*spinosa*	M	4755.3	1	-	This manuscript

Table 6.1 Continued

Order	Family	Genus	Species	Sex	GS (Mbp)	Rep	StDev	Source
Orthoptera	Tettigoniidae	*Pediodectes*	*sp*	-	9080.7	1	-	This manuscript
Orthoptera	Tettigoniidae	*Pediodectes*	*tinkhami?*	M	7795.83	1	-	This manuscript
Orthoptera	Tettigoniidae	*Scudderia*	*sp*	-	13079.3	2	79.25	This manuscript
Phamatodea	Diapheromeridae	*Megaphasma*	*dentricus*	M	2774.1	1	-	This manuscript
Phamatodea	Pseudophasmatidae	*Anisomorpha*	*buprestoides*	M	2944.1	2	40.43	This manuscript
Blattodea	Blaberidae	*Blaberus*	*fuscus*	-	3286.08	-	-	AGSD
Blattodea	Blaberidae	*Blaptica*	*dubia*	-	4440.12	-	-	AGSD
Blattodea	Blaberidae	*Nauphoeta*	*cinerea*	-	5036.7	-	-	AGSD
Blattodea	Blatellidae	*Blatella*	*nipponica*	-	2083.14	-	-	AGSD
Blattodea	Blattellidae	*Blattella*	*germanica*	-	1956	-	-	AGSD
Blattodea	Blattidae	*Blatta*	*orientalis*	-	2963.34	-	-	AGSD
Blattodea	Blattidae	*Periplaneta*	*americana*	-	2660.16	-	-	AGSD
Blattodea	Cryptocercidae	*Cryptocercus*	*kyebangensis*	-	1134.48	-	-	AGSD
Blattodea	Cryptocercidae	*Cryptocercus*	*punctulatus*	-	1290.96	-	-	AGSD
Blattodea	Blaberidae	*Panchlora*	*nivea*	-	1486.56	-	-	AGSD
Blattodea	Blattellidae	*Parcoblatta*	*pensylvanica*	-	1026.9	-	-	AGSD
Blattodea	Blattidae	*Periplaneta*	*americana*	-	3334.98	-	-	AGSD
Dermaptera	Labiduridae	*Labidura*	*riparia*	-	518.34	-	-	AGSD
Embiidina	Oligotomidae	*Oligotoma*	*saundersii*	-	2601.48	-	-	AGSD
Isoptera	Hodotermitidae	*Hodotermes*	*mossambicus*	-	978	-	-	AGSD
Isoptera	Kalotermitidae	*Glyptotermes*	*fuscus*	-	1525.68	-	-	AGSD
Isoptera	Kalotermitidae	*Glyptotermes*	*nakajimai*	-	870.42	-	-	AGSD
Isoptera	Kalotermitidae	*Neotermes*	*koshunensis*	-	1828.86	-	-	AGSD
Isoptera	Mastotermitidae	*Mastotermes*	*darwiniensis*	-	1271.4	-	-	AGSD
Isoptera	Rhinotermitidae	*Coptotermes*	*formosanus*	-	909.54	-	-	AGSD
Isoptera	Rhinotermitidae	*Coptotermes*	*formosanus*	-	841.08	-	-	AGSD
Isoptera	Rhinotermitidae	*Reticulitermes*	*flavipes*	-	1046.46	-	-	AGSD
Isoptera	Rhinotermitidae	*Reticulitermes*	*speratus kyushuensis*	-	1046.46	-	-	AGSD
Isoptera	Rhinotermitidae	*Reticulitermes*	*speratus speratus*	-	987.78	-	-	AGSD
Isoptera	Rhinotermitidae	*Reticulitermes*	*yaeyamanus*	-	978	-	-	AGSD
Isoptera	Termitidae	*Nasutitermes*	*takasagoensis*	-	1643.04	-	-	AGSD
Isoptera	Termitidae	*Odontotermes*	*formosanus*	-	1447.44	-	-	AGSD
Isoptera	Termitidae	*Pericapritermes*	*nitobei*	-	1858.2	-	-	AGSD
Isoptera	Termopsidae	*Hodotermopsis*	*sjostedti*	-	1242.06	-	-	AGSD
Isoptera	Termopsidae	*Zootermopsis*	*nevadensis*	-	567.24	-	-	AGSD
Mantodea	Hymenopodidae	*Acromantis*	*japonica*	-	4430.34	-	-	AGSD
Mantodea	Mantidae	*Litaneutria*	*sp.*	-	3246.96	-	-	AGSD
Mantodea	Mantidae	*Stagmomantis*	*carolina*	-	3863.1	-	-	AGSD
Mantodea	Mantidae	*Statilia*	*maculata*	-	2982.9	-	-	AGSD
Mantodea	Mantidae	*Tenodera*	*aridifolia*	-	2855.76	-	-	AGSD
Orthoptera	Acrididae	*Acrida*	*conica*	-	10581.96	-	-	AGSD
Orthoptera	Acrididae	*Acrida*	*conica*	-	12273.9	-	-	AGSD
Orthoptera	Acrididae	*Ailopus*	*thalassinus*	-	6533.04	-	-	AGSD
Orthoptera	Acrididae	*Austroicetes*	*pusilla*	-	6151.62	-	-	AGSD
Orthoptera	Acrididae	*Caledia*	*captiva*	-	10660.2	-	-	AGSD
Orthoptera	Acrididae	*Campylacantha*	*olivacea*	-	6425.46	-	-	AGSD
Orthoptera	Acrididae	*Chorthippus*	*apicalis*	-	12332.58	-	-	AGSD
Orthoptera	Acrididae	*Chorthippus*	*binotatus binotatus*	-	10669.98	-	-	AGSD
Orthoptera	Acrididae	*Chorthippus*	*brunneus*	-	8361.9	-	-	AGSD
Orthoptera	Acrididae	*Chorthippus*	*brunneus*	-	9251.88	-	-	AGSD
Orthoptera	Acrididae	*Chorthippus*	*brunneus*	-	9926.7	-	-	AGSD
Orthoptera	Acrididae	*Chorthippus*	*cf. binotatus*	-	10122.3	-	-	AGSD
Orthoptera	Acrididae	*Chorthippus*	*dorsatus*	-	8156.52	-	-	AGSD
Orthoptera	Acrididae	*Chorthippus*	*jacobsi*	-	10601.52	-	-	AGSD
Orthoptera	Acrididae	*Chorthippus*	*jucundus*	-	11618.64	-	-	AGSD
Orthoptera	Acrididae	*Chorthippus*	*longicornis*	-	8391.24	-	-	AGSD
Orthoptera	Acrididae	*Chorthippus*	*nevadensis*	-	11276.34	-	-	AGSD
Orthoptera	Acrididae	*Chorthippus*	*parallelus*	-	12039.18	-	-	AGSD
Orthoptera	Acrididae	*Chorthippus*	*parallelus*	-	13066.08	-	-	AGSD

Table 6.1 Continued

Order	Family	Genus	Species	Sex	GS (Mbp)	Rep	StDev	Source
Orthoptera	Acrididae	*Chorthippus*	*parallelus*	-	13525.74	-	-	AGSD
Orthoptera	Acrididae	*Chorthippus*	*parallelus*	-	14396.16	-	-	AGSD
Orthoptera	Acrididae	*Chorthippus*	*scalaris*	-	14396.16	-	-	AGSD
Orthoptera	Acrididae	*Chorthippus*	*vagans*	-	8449.92	-	-	AGSD
Orthoptera	Acrididae	*Chorthippus*	*vagans*	-	8489.04	-	-	AGSD
Orthoptera	Acrididae	*Chortoicetes*	*terminifera*	-	5858.22	-	-	AGSD
Orthoptera	Acrididae	*Chortoicetes*	*terminifera*	-	7061.16	-	-	AGSD
Orthoptera	Acrididae	*Cryptobothrus*	*chrysophorus*	-	9163.86	-	-	AGSD
Orthoptera	Acrididae	*Eyprepocnemis*	*plorans*	-	9486.6	-	-	AGSD
Orthoptera	Acrididae	*Gastrimargus*	*musicus*	-	8811.78	-	-	AGSD
Orthoptera	Acrididae	*Gomphocerus*	*sibiricus*	-	8753.1	-	-	AGSD
Orthoptera	Acrididae	*Heteracris*	*adspersus*	-	6200.52	-	-	AGSD
Orthoptera	Acrididae	*Humbe*	*tenuicornis*	-	8029.38	-	-	AGSD
Orthoptera	Acrididae	*Locusta*	*migratoria*	-	5163.84	-	-	AGSD
Orthoptera	Acrididae	*Locusta*	*migratoria*	-	5349.66	-	-	AGSD
Orthoptera	Acrididae	*Locusta*	*migratoria*	-	5956.02	-	-	AGSD
Orthoptera	Acrididae	*Locusta*	*migratoria*	-	6132.06	-	-	AGSD
Orthoptera	Acrididae	*Locusta*	*migratoria*	-	6210.3	-	-	AGSD
Orthoptera	Acrididae	*Locusta*	*migratoria*	-	6298.32	-	-	AGSD
Orthoptera	Acrididae	*Macrotona*	*australis*	-	8303.22	-	-	AGSD
Orthoptera	Acrididae	*Melanoplus*	*differentialis*	-	3755.52	-	-	AGSD
Orthoptera	Acrididae	*Melanoplus*	*differentialis*	-	6092.94	-	-	AGSD
Orthoptera	Acrididae	*Melanoplus*	*differentialis*	-	6875.34	-	-	AGSD
Orthoptera	Acrididae	*Melanoplus*	*sanguinipes*	-	5701.74	-	-	AGSD
Orthoptera	Acrididae	*Myrmeleotettix*	*maculatus*	-	11872.92	-	-	AGSD
Orthoptera	Acrididae	*Myrmeleotettix*	*maculatus*	-	12381.48	-	-	AGSD
Orthoptera	Acrididae	*Myrmeleotettix*	*maculatus*	-	13085.64	-	-	AGSD
Orthoptera	Acrididae	*Omocestus*	*viridulus*	-	12870.48	-	-	AGSD
Orthoptera	Acrididae	*Peakesia*	*hospita*	-	10239.66	-	-	AGSD
Orthoptera	Acrididae	*Phaulacridium*	*vittatum*	-	10493.94	-	-	AGSD
Orthoptera	Acrididae	*Podisma*	*pedestris*	-	16557.54	-	-	AGSD
Orthoptera	Acrididae	*Schistocerca*	*cancellata*	-	9281.22	-	-	AGSD
Orthoptera	Acrididae	*Schistocerca*	*gregaria*	-	8361.9	-	-	AGSD
Orthoptera	Acrididae	*Schistocerca*	*gregaria*	-	8518.38	-	-	AGSD
Orthoptera	Acrididae	*Schistocerca*	*gregaria*	-	8762.88	-	-	AGSD
Orthoptera	Acrididae	*Schistocerca*	*paranensis*	-	8440.14	-	-	AGSD
Orthoptera	Acrididae	*Schizobothrus*	*flavovittatus*	-	7335	-	-	AGSD
Orthoptera	Acrididae	*Stauroderus*	*scalaris*	-	15980.52	-	-	AGSD
Orthoptera	Acrididae	*Valanga*	*irregularis*	-	9232.32	-	-	AGSD
Orthoptera	Eumasticidae	*Warramaba*	*virgo*	-	3667.5	-	-	AGSD
Orthoptera	Eumasticidae	*Warramaba*	*virgo*	-	3912	-	-	AGSD
Orthoptera	Gryllacrididae	*Ceuthophilus*	*stygius*	-	9339.9	-	-	AGSD
Orthoptera	Gryllidae	*Acheta*	*domesticus*	-	1956	-	-	AGSD
Orthoptera	Gryllidae	*Acheta*	*domesticus*	-	1956	-	-	AGSD
Orthoptera	Gryllidae	*Acheta*	*domesticus*	-	1956	-	-	AGSD
Orthoptera	Gryllidae	*Acheta*	*domesticus*	-	1956	-	-	AGSD
Orthoptera	Gryllidae	*Acheta*	*domesticus*	-	2327.64	-	-	AGSD
Orthoptera	Gryllidae	*Gryllus*	*pennsylvanicus*	-	1956	-	-	AGSD
Orthoptera	Gryllidae	*Gryllus*	*pennsylvanicus*	-	2014.68	-	-	AGSD
Orthoptera	Gryllidae	*Gryllus*	*pennsylvanicus*	-	2621.04	-	-	AGSD
Orthoptera	Gryllidae	*Hadenoecus*	*subterraneus*	-	1515.9	-	-	AGSD
Orthoptera	Gryllidae	*Laupala*	*cerasina*	-	1887.54	-	-	AGSD
Orthoptera	Gryllidae	*Oecanthus*	*niveus*	-	1672.38	-	-	AGSD
Orthoptera	Gryllotalpidae	*Neoscapteriscus*	*borellii*	-	3334.98	-	-	AGSD
Orthoptera	Tettigoniidae	*Conocephalus*	*sp.*	-	2777.52	-	-	AGSD
Orthoptera	Tettigoniidae	*Neoconocephalus*	*triops*	-	7442.58	-	-	AGSD
Orthoptera	Tridactylidae	*Unknown*	*sp.*	-	2572.14	-	-	AGSD
Phasmatodea	Pseudophasmatidae	*Anisomorpha*	*buprestoides*	-	2904.66	-	-	AGSD
Phasmatodea	Timematidae	*Timema*	*cristinae*	-	1330.08	-	-	AGSD

Table 6.1 Continued

Order	Family	Genus	Species	Sex	GS (Mbp)	Rep	StDev	Source
Phasmatodea	Heteronemiidae	*Diapheromera*	*femorata*	-	2493.9	-	-	AGSD
Phasmatodea	Phasmatidae	*Bacillus*	*atticus atticus*	-	2249.4	-	-	AGSD
Phasmatodea	Phasmatidae	*Bacillus*	*atticus caprai*	-	2180.94	-	-	AGSD
Phasmatodea	Phasmatidae	*Bacillus*	*atticus carius*	-	2875.32	-	-	AGSD
Phasmatodea	Phasmatidae	*Bacillus*	*atticus cyprius*	-	2396.1	-	-	AGSD
Phasmatodea	Phasmatidae	*Bacillus*	*grandii grandii*	-	2552.58	-	-	AGSD
Phasmatodea	Phasmatidae	*Bacillus*	*grandii grandii*	-	2112.48	-	-	AGSD
Phasmatodea	Phasmatidae	*Bacillus*	*rossius redtenbacheri*	-	2122.26	-	-	AGSD
Phasmatodea	Phasmatidae	*Bacillus*	*rossius redtenbacheri*	-	1907.1	-	-	AGSD
Phasmatodea	Phasmatidae	*Bacillus*	*whitei*	-	2220.06	-	-	AGSD
Phasmatodea	Phasmatidae	*Extatosoma*	*tiaratum*	-	7824	-	-	AGSD
Zoraptera	Zorotypidae	*Zorotypus*	*hubbardi*	-	1848.42	-	-	AGSD

Table 0.2 accession numbers of the sequences used for the inference of the phylogeny of Polyneoptera

Taxa name	COI	COX	ND4	X18S	X28S.A	X28S.B
Acanthops falcata	KU507627	KU528759	KU507927	KU320376	-	KU320480
Acanthoxyla geisovii	EU492959	EU492986	-	-	-	-
Acanthoxyla inermis	-	-	-	-	-	EU543518
Acrophylla titan	FJ474258	FJ474335	-	-	-	-
Amitermes darwini	-	EF442698	-	-	-	-
Amitermes germanus	-	EF442703	-	-	-	-
Amitermes parvus	-	EF442704	-	-	-	-
Anechura sp	-	-	-	KC413722	-	-
Anisolabis maritima	MF468289	AB005470	-	-	-	-
Antemna rapax	EF383875	EF384003	FJ802557	EF383553	-	EF383716
Antistia sp	-	-	-	FJ806471	-	-
Archimandrita tessellata	JN615372	DQ874263	-	DQ874110	DQ874197	-
Archimantis sobrina	EF383802	-	FJ802427	EF383602	-	AY491222
Arixenia esau	-	-	-	KX069014	-	-
Atrachelacris sp	AY014360	-	-	-	-	-
Bacillus atticus	-	AY185558	-	KF256446	-	KF256329
Bacillus grandii benazzii	-	AF148314	-	-	-	-
Bacillus grandii grandii	-	AF148301	-	-	-	-
Bacillus lynceorum	-	AF038214	-	-	-	-
Bacillus rossius	-	-	-	-	-	AY125320
Baculum sp	-	-	-	AY121172	-	-
Bifiditermes improbus	-	AF189079	-	-	-	-
Blaberus discoidalis	KF372514	AB014063	-	-	-	-
Blaberus giganteus	-	AB014064	-	-	-	-
Blatta orientalis	KP986400	DQ874267	FJ802410	AY521830	-	AY521741
Blattella bisignata	KT271843	-	-	-	-	-
Blattella germanica	AY176057	DQ874268	-	FJ806322	DQ874201	FJ806519
Blattella sauteri	KY349679	-	-	-	-	-
Bolbe pallida	FJ802759	-	FJ802436	FJ806343	-	FJ806541
Byrsotria fumigata	JN615375	DQ874269	-	DQ874117	DQ874202	-
Calineuria sp	-	-	-	AY521881	-	-
Callimantis antillarum	MF414717	-	-	MF414712	-	MF414715
Carausius morosus	FJ474268	FJ474344	-	AY121170	-	AY125310
Chaetospania sp	-	-	-	KX069010	-	-
Cheddikulama straminea	KT316265	KT316267	-	-	-	-
Chlorus bolivianus	FJ829333	-	-	-	-	-
Choeradodis rhombicollis	EF383805	-	FJ802431	EF383481	-	AY491226
Cliomantis cornuta	FJ802777	-	FJ802470	FJ806360	-	AY491264
Clitarchus hookeri	KF383507	EU492999	-	-	-	EU543521
Clonopsis felicitatis	-	GQ370542	-	-	-	-
Clonopsis gallica	-	AF096287	-	-	-	-
Clonopsis maroccana	-	GQ370575	-	-	-	-

Table 6.2 Continued

Taxa name	COI	COX	ND4	X18S	X28S.A	X28S.B
Clonopsis soumiae	-	GQ370529	-	-	-	-
Coptotermes acinaciformis	HQ878425	FJ384631	-	DQ882634	DQ441893	-
Crenetermes albotarsalis	-	DQ442108	-	-	DQ441898	-
Creobroter laevicollis	FJ802780	FJ806890	FJ802488	FJ806363	-	FJ806562
Cryptocercus primarius	-	-	-	-	KU312232	-
Cryptocercus punctulatus	KY241441	AB005462	-	AY521829	DQ441902	AY521739
Cryptocercus relictus	-	KF855906	-	JX091733	KU312247	-
Cryptotermes austrinus	-	AF189081	-	-	-	-
Cryptotermes brevis	EU253840	EU253879	FJ802415	EU253783	-	FJ806528
Cryptotermes cynocephalus	-	AF189083	-	-	-	-
Cryptotermes domesticus	-	JQ678601	-	-	-	-
Cryptotermes dudleyi	-	FN377808	-	-	-	-
Cryptotermes primus	-	AF189090	-	-	-	-
Cryptotermes queenslandis	-	AF189092	-	-	-	-
Cryptotermes secundus	-	DQ442111	-	DQ882635	DQ441901	-
Ctenomorpha sp	-	FJ474350	-	-	-	-
Cubitermes sp	-	-	-	-	DQ441903	-
Deiphobe sp	-	-	-	-	-	KP340217
Deropeltis erythrocephala	-	DQ874271	-	DQ874121	-	-
Dichromatos lilloanus	FJ829336	-	-	-	-	-
Dichromorpha sp	-	-	-	KM853205	-	-
Dichroplus conspersus	DQ083454	-	-	-	-	-
Dichroplus democraticus	DQ083455	-	-	-	-	-
Dichroplus elongatus	KY595084	-	-	-	-	-
Dichroplus exilis	KY595085	-	-	-	-	-
Dichroplus fuscus	KY595086	-	-	-	-	-
Dichroplus maculipennis	KY595088	-	-	-	-	-
Dichroplus paraelongatus	KY595089	-	-	-	-	-
Dichroplus patruelis	DQ083458	-	-	-	-	-
Dichroplus pratensis	DQ083459	-	-	-	-	-
Dichroplus schulzi	DQ083460	-	-	-	-	-
Dichroplus silveiraguidoi	DQ083461	-	-	-	-	-
Dichroplus vittatus	KY595090	-	-	-	-	-
Didymocorypha lanceolata	EF383847	-	FJ802502	EF383524	-	EF383687
Diploptera punctata	JN615379	DQ874273	-	DQ874123	DQ874207	-
Diponthus sp	-	-	-	KM853222	-	-
Drepanotermes septentrionalis	-	EF442705	-	-	-	-
Dystacta alticeps	FJ802826	FJ806939	FJ802592	EF383571	-	EF383737
Ectobius pallidus	-	DQ874276	-	DQ874126	DQ874210	-
Embia nuragica	JQ907059	-	-	JQ907235	-	JQ906996
Embia ramburi	-	-	-	JQ907247	-	-
Embia tyrrhenica	JQ907057	-	-	JQ907233	-	JQ906994
Empusa sp	-	-	FJ802545	-	-	-
Ephelotermes melachoma	-	EF442709	-	-	-	-
Ephelotermes taylori	-	EF442710	-	-	-	-
Epilampra sp	-	-	-	KY497641	-	-
Eublaberus distanti	KF372518	KF372541	-	EU367508	-	-
Eublaberus posticus	JN615376	DQ874281	-	DQ874131	DQ874215	-
Euborellia moesta	-	AF140545	-	-	-	-
Eurycotis floridana	JN615395	DQ874283	-	DQ874133	-	-
Extatosoma tiaratum	KJ201987	KJ024574	-	AY121155	-	AY125295
Forficula auricularia	MF462143	-	-	Z97594	-	-
Galiblatta sp	-	-	-	KY497640	-	-
Galloisiana nipponensis	-	KC142671	-	-	-	-

Table 6.2 Continued

Taxa name	COI	COX	ND4	X18S	X28S.A	X28S.B
Glyptotermes brevicornis	-	AF189096	-	-	-	-
Glyptotermes iridipennis	-	AF189097	-	-	-	-
Gongylus sp	-	-	FJ802421	-	-	-
Gonolabis sp	-	-	-	KC413693	-	-
Gromphadorhina portentosa	KF372519	DQ181523	-	Z97592	-	EF383626
Grylloblatta sp	-	KP972434	-	-	-	-
Haploembia solieri	JQ907039	-	-	JQ907206	-	-
Harpagomantis tricolor	KR360618	-	-	KR360562	-	KR360595
Hemimerus sp	-	-	-	JQ714388	-	-
Hemithyrsocera sp	KY349565	-	-	-	-	-
Hestiasula sp	-	-	FJ802734	-	-	-
Hierodula patellifera	JF700167	-	-	-	-	-
Humbertiella similis	EF383834	EF383961	FJ802484	EF383511	-	EF383672
Hyalopteryx rufipennis	-	-	-	KM853210	-	KM853480
Incisitermes sp	-	-	-	GQ337715	-	-
Iris sp	-	-	FJ802599	-	-	-
Isoperla grammatica	KU955920	-	-	-	-	-
Kalotermes flavicollis	EU253842	GU931799	-	EU253785	-	FJ806529
Karoophasma biedouwense	-	KF855912	-	-	-	-
Kongobatha sp	-	-	-	FJ806473	-	-
Labidura riparia	JN241998	AF140544	-	AY707356	-	-
Lamproblatta albipalpus	JN615394	KF855913	-	KF855836	-	-
Lanxoblatta emarginata	KF372522	KF372544	-	EU367509	-	-
Leiotettix flavipes	FJ829337	-	-	-	-	-
Leiotettix pulcher	DQ083464	-	-	-	-	-
Leiotettix viridis	AY014353	-	-	-	-	-
Leptynia attenuata	-	AF241441	-	-	-	-
Leptynia caprai	-	AF241431	-	-	-	-
Leptynia montana	-	AF241416	-	-	-	-
Leptysma sp	-	-	-	KM853199	-	-
Ligaria sp	-	-	FJ802551	-	-	-
Liturgusa cursor	KU507661	KU528782	KU507962	KU320411	-	KU320516
Liturgusa maya	KU507664	KU528783	FJ802555	KU320413	-	KU320518
Loboptera decipiens	KF372524	DQ874297	-	KF372502	-	-
Lobopterella dimidiatipes	KX053864	AB005905	-	-	-	-
Lophotermes septentrionalis	-	EF442708	-	-	-	-
Macrognathotermes sunteri	-	DQ442158	-	-	DQ441947	-
Macropanesthia rhinoceros	-	DQ874299	-	DQ874152	DQ874230	-
Macrotermes bellicosus	AY127702	JF923231	-	-	-	-
Mantis religiosa	FJ802846	KP639994	FJ802419	AY859586	-	EF383633
Marellia remipes	-	-	-	KM853256	-	KM853434
Margattea sp	KY349621	-	-	-	-	-
Mastotermes darwiniensis	JN615367	EU253885	-	DQ882638	-	EF383632
Microcerotermes boreus	-	EF442696	-	-	-	-
Microcerotermes nervosus	-	EF079031	-	-	-	-
Microtermes sp	-	-	-	-	JQ429105	-
Miomantis sp	-	-	FJ802627	-	-	-
Nala lividipes	-	-	-	AY707362	-	-
Nasutitermes graveolus	-	DQ442185	-	-	DQ441974	-
Nasutitermes longipennis	-	DQ442190	-	-	DQ441978	-
Nasutitermes triodiae	-	DQ442197	-	-	DQ441986	-
Nauphoeta cinerea	JN615381	DQ874301	-	KF372504	DQ874233	-
Neotermes insularis	-	AF189105	-	-	-	-
Nesogaster sp	-	-	-	AY707358	-	-

Table 6.2 Continued

Taxa name	COI	COX	ND4	X18S	X28S.A	X28S.B
Odontotermes sp	-	-	-	EU253801	-	-
Oligonyx sp	-	-	KU508002	-	-	-
Oligotoma sp	-	-	-	JQ907216	-	-
Ommatolampis sp	-	-	-	KM853267	-	-
Ommexecha sp	-	-	-	KM853269	-	-
Orthodera sp	-	-	FJ802422	-	-	-
Orxines sp	-	-	-	AY121153	-	-
Panchlora nivea	JN615382	-	-	-	-	-
Panchlora viridis	-	DQ181534	-	-	-	-
Parathespis humbertiana	-	-	FJ802660	EF383616	-	EF383782
Paulinia acuminata	-	-	-	KM853257	-	KM853433
Pelmatosilpha sp	-	-	-	EU253773	-	-
Pericapritermes sp	-	-	-	-	DQ442004	-
Periplaneta americana	KR144773	M83971	-	AF370792	-	-
Periplaneta australasiae	JN615391	DQ874310	-	DQ874169	DQ874242	-
Periplaneta brunnea	MG572238	DQ874311	-	DQ874170	-	-
Periplaneta fuliginosa	MG458948	DQ874312	-	DQ874171	-	-
Periplaneta japonica	KC407710	-	-	KC413768	-	-
Perla abdominalis	MF458789	-	-	-	-	-
Perla marginata	KF492799	-	-	-	-	-
Perlodes sp	-	-	-	EF622774	-	-
Phalces sp	-	KT426619	-	-	-	-
Phobaeticus sp	-	-	-	AY121184	-	-
Phoetalia sp	-	-	-	KF372509	-	-
Phyllium bioculatum	-	KJ024570	-	Z97575	-	AY125301
Podacanthus wilkinsoni	GQ927386	GQ927422	-	-	-	-
Polyspilota aeruginosa	FJ802847	GU064726	FJ802575	EF383478	-	FJ806633
Porotermes adamsoni	-	LC193947	-	-	-	-
Prionolopha serrata	MF682223	-	-	-	-	-
Procryptotermes australiensis	-	AF189107	-	-	-	-
Procubitermes sp	-	-	-	-	DQ441918	-
Protermes sp	-	-	-	-	DQ442017	-
Pseudacanthops sp	-	-	KU508010	-	-	-
Pseudacanthotermes militaris	AY127731	DQ442233	-	DQ882644	DQ442020	-
Pseudomiopteryx infuscata	KU507710	-	KU508012	KU320463	-	KU320566
Pseudophasma sp	-	KJ024573	-	-	-	-
Pseudoscopas nigrigena	FJ829342	-	-	-	-	-
Pycnoscelus surinamensis	JN615383	DQ874325	-	DQ874185	DQ874252	-
Reticulitermes flavipes	AY027469	EU689011	-	-	-	-
Reticulitermes lucifugus	KM245729	JQ231192	-	-	-	-
Reticulitermes urbis	KM245722	JQ231191	-	-	-	-
Rhabdoblatta sp	-	AB007543	-	-	-	-
Ronderosia sp	DQ083468	-	-	-	-	-
Schedorhinotermes lamanianus	-	DQ442239	-	DQ882645	DQ442025	-
Schistocerca cancellata	KY980925	KY980962	KY981176	-	-	-
Schistocerca flavofasciata	KY980916	KY980953	-	-	-	-
Schistocerca pallens	KY980891	KY980928	KY981153	KM853186	-	KM853504
Schizocephala bicornis	EF383831	EF383958	FJ802480	EF383508	-	EF383669
Scotussa daguerrei	DQ083469	-	-	-	-	-
Scotussa impudica	KY595091	-	-	-	-	-
Scotussa lemniscata	FJ829338	-	-	-	-	-
Sipyloidea sipylus	FJ474324	FJ474393	-	AY121181	-	AY125321
Sphodromantis sp	-	-	FJ802430	-	-	-
Sphodropoda sp	-	-	FJ802670	-	-	-
Statilia maculata	FJ802905	AB006435	FJ802477	FJ806503	-	FJ806717
Stolotermes sp	-	-	-	EU253798	-	-

Table 6.2 Continued

Taxa name	COI	COX	ND4	X18S	X28S.A	X28S.B
Supella longipalpa	FJ802748	FJ806878	FJ802414	EF383467	-	FJ806526
Symploce sp	-	-	-	DQ874188	-	-
Symplocodes sp	KY349568	-	-	-	-	-
Tenodera aridifolia	-	GU064747	-	AF423805	-	AY125282
Tenodera australasiae	-	GU064735	-	-	-	-
Tenodera superstitiosa	-	GU064753	-	-	-	-
Thesprotia graminis	EF383826	EF383952	FJ802473	EF383502	-	EF383662
Thoracotermes macrothorax	-	DQ442254	-	-	DQ442040	-
Timema poppensis	HQ184611	-	-	-	-	-
Timema shepardi	HQ184658	-	-	-	-	-
Trimerotropis ochraceipennis	JQ513038	-	-	-	-	-
Trimerotropis pallidipennis	JQ513036	GU476998	-	-	-	-
Tropidostethus sp	-	-	-	KM853196	-	-
Tuberculitermes bycanistes	-	DQ442258	-	-	DQ442044	-
Tumulitermes pastinator	-	DQ442261	-	-	DQ442047	-
Unguitermes sp	-	-	-	-	DQ442050	-
Vates sp	-	-	FJ802429	-	-	-
Xestoblatta sp	-	-	-	KF372512	-	-
Xyleus insignis	MF682248	-	-	-	-	-
Xyleus laevipes	MF682234	-	-	-	-	-
Xyleus modestus	KJ889693	-	-	KM853221	-	KM853469
Zoniopoda hempeli	MF682228	-	-	-	-	-
Zoniopoda omnicolor	MF682227	-	-	-	-	-
Zoniopoda tarsata	MF682225	-	-	-	-	-
Zootermopsis angusticollis	MF477188	DQ442267	-	AY859615	-	-
Zootermopsis nevadensis	JN615368	EU253894	AB936819	EU253799	-	-

APPENDIX B

ADDITIONAL SUPPLEMENTARY MATERIAL OF CHAPTER 4

Table 0.3 Summary statistics of Y chromosome assembly at different kmer sizes and two minimum kmer coverage multiplicity cutoff values. Assembly of choice is represented in orange color shading. All stats are for contigs greater than 500bp

Total contigs	Contigs > 500bp	L50	min	N75	N50	N25	E-size	max	Size	kmer-size	multiplicity cutoff
266107	165	70	500	539	617	719	652	1196	103352	18	2
121248	718	284		572	676	852	752	2212	497139	26	
99621	705	275		574	690	878	791	3287	498907	34	
83909	697	271		570	679	847	786	3300	489516	42	
69658	716	276		569	674	876	822	5504	504904	50	
57331	741	282		566	668	869	860	5580	521935	58	
46451	767	287		571	681	907	877	6651	549939	66	
36295	873	324		576	712	948	912	6449	643463	74	
29161	975	356		580	711	965	932	6652	727192	82	
22731	1032	362		589	731	1034	982	6652	792019	90	
17118	1056	367		595	737	1067	1019	6490	823430	98	
12307	1031	334		596	769	1278	1096	6490	833964	106	
8827	918	286		607	794	1401	1199	6602	772221	114	
5998	716	218		608	793	1418	1280	7461	607177	122	
156916	276	115	500	551	621	760	683	1540	178319	18	3
69869	846	315		578	692	926	879	6358	614307	26	
58669	828	309		584	696	920	877	6385	603555	34	
49372	857	314		588	706	954	918	6616	637091	42	
40941	871	319		590	721	977	924	6618	653915	50	
34158	905	321		586	727	1006	988	6585	688917	58	
28330	952	342		582	722	982	962	6616	718485	66	
22860	1028	355		590	734	1093	1011	6644	800110	74	
18746	1039	351		606	767	1200	1063	6644	832898	82	
14893	1019	332		609	775	1301	1114	6644	835822	90	
11370	970	309		617	784	1360	1200	8344	812176	98	
8464	896	278		616	798	1433	1245	6583	765901	106	
6210	762	233		613	792	1422	1328	8729	651025	114	
4343	529	177		592	766	1130	1049	5320	421435	122	