Landscapes of (Co)Evolution: A Tale of Two Signals

by

Victoria R. Caudill

A dissertation accepted and approved in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in Biology

Dissertation Committee:
Dr. Brendan Bohannan, Chair
Peter Ralph, Advisor
Dr. Matt Barber, Core Member
Dr. Lauren Ponisio, Core Member
Dr. Mike Harms, Institutional Representative

University of Oregon
Spring 2024

© 2024 Victoria R. Caudill

DISSERTATION ABSTRACT

Victoria R. Caudill

Doctor of Philosophy in Biology

Title: Landscapes of (Co)Evolution: A Tale of Two Signals

Evolution between species and within specific environments has resulted in a diverse array of traits and genetic variation. These are the result of interactions that have occurred across space and throughout time. As evolution progresses, it generates signals and patterns that can be used to unravel mysteries of the past and provide insight into future possibilities. By examining the signals left behind by the process of evolution, I have gained valuable insights onto what has occurred in the past. In chapter II, I use spatial simulations to explore the (co)evolutionary trajectories of levels of toxin resistance and toxin production in the predator-prey *Thamnophis* garter snake – *Taricha* newt system. Specifically, I examine how possible genetic architectures of the toxin and resistance traits affect the coevolutionary dynamics by manipulating both mutation rate and effect size of mutations across many simulations. I find that coevolutionary dynamics alone were not sufficient in our simulations to produce the striking mosaic of levels of toxicity and resistance observed in nature. Instead, simulations with ecological heterogeneity (in trait costliness or interaction rate) did produce such patterns. In chapter III, I examined landscapes of genetic variation in cichlids, a species complex that has recently radiated. I used the phylogenetic relationship and population genetic measurements (mean nucleotide diversity and divergence) to describe large-scale variation across the genome. These patterns are likely caused by complex effects of inversions, introgression, and linked selection. Together, these findings contribute to building a strong foundation for understanding the evolutionary signals of natural selection and how its' impacts vary along the genome. This dissertation includes previously published and unpublished co-authored material.

TABLE OF CONTENTS

Chapter Page

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

CHAPTER 1

INTRODUCTION

The study of evolution is complex and messy, leaving multiple threads for researchers to grasp. At its core, evolution is the unifying principle of biology that explains how life on Earth has diversified and adapted over billions of years. Evolution is mediated by several different mechanisms including natural selection, adaptation, mutation, gene flow, genetic drift, non-random mating, and genetic variation. These mechanisms can interact in various ways, leading to diverse evolutionary outcomes. Furthermore, evolution can act over multiple timescales—from short-term changes within populations (microevolution) to long-term patterns shaping the diversity of life over geological epochs (macroevolution). The process of evolution is also intricately linked with environmental factors, such as climate change, habitat alteration, and species interactions, which shape the selective pressures acting on organisms.

The foundation of the study of evolution began with simple observations about the relationships between organisms. Over time, these observations have grown into theories and experiments that dissect the intricate relationships between these different mechanisms and timescales of evolution. Gradually, our understanding of evolution has grown, as the field of genetics has been integrated with evolutionary theory. From the many observations conducted, experiments ran, and theory produced, evolutionary biology has become an umbrella with various areas of focus that are encompassed by this term (i.e., speciation, developmental biology, phylogenetics, coevolution, systematics, population genetics, etc..). Each of these subfields have several ways of describing and quantifying signals of evolutionary change. Through there are various fields in evolutionary biology, in this dissertation I will focus on coevolution and population genetics. Researchers have explored the reciprocal evolutionary changes between species interacting within their environment and with eachother. These interactions, also referred to as coevolution, can take place in wide networks consisting of multiple species interacting in various environments across space and time. Understanding these complex interactions is fundamental in providing insights into the shaping of ecological communities and their impacts on biodiversity. Studies on coevolution have mainly focused on host-parasite relationships, plant-pollinator associations, and predator-prey interactions. In chapter II, I study evolutionary trajectories in a coevolutionary arms race of a

predator-prey (snake-newt) system. Using simulation, I explore the geographic variation of newts' toxicity and snakes' resistance as these species interact across space and time. This study focuses on the effects of genetic basis on trait evolution, in the context of realistic reciprocal selection, genetic drift, gene flow, and spatial environmental variation. This work is currently in review and is co-authored with Dr. Peter Ralph.

Evolutionary biology also encompasses the study of common ancestry, where all existing organisms are related through decent from a common ancestor over the course of evolutionary history. As species evolve, different evolutionary mechanisms leave lasting signals within their genomes. These signals are often difficult to discern due to evolutionary mechanisms occurring at different times, in opposition, or jointly. Oftentimes, evolution impacts different areas of a genome differently. In chapter III, I examine the signals left by evolution along the genome of 14 species of cichlid. We found that there are similarities in the peaks and valleys of the landscapes of genetic diversity and divergence of these species. These genetic landscapes are highly correlated for sister species but are less correlated for more phylogenetically distant species. There are also major differences between the linkage groups, suggesting multiple evolutionary mechanisms acting differently along the genome. This work is unpublished and is co-authored with Dr. Anastasia Teterina, Aidan Short, and Dr. Peter Ralph.

Extensive studies of evolution have greatly advanced our understanding of adaptation, genetics, and diversity. My work in this dissertation continues this expansion of our knowledge by investigating some of the many signals that are produced by evolutionary processes.

CHAPTER 2

CHAPTER 2 GENETIC ARCHITECTURE, SPATIAL HETEROGENEITY, AND THE COEVOLUTIONARY ARMS RACE BETWEEN NEWTS AND SNAKES

This work is currently in review. The experimental question and design for this project was developed by myself and Dr. Peter L. Ralph. I ran and analyzed all simulation results and wrote up the manuscript for this project jointly with Dr. Peter L. Ralph.

2.1 Introduction

Coevolution is the reciprocal adaptation of heritable traits between interacting species (Janzen et al. 1980). These dynamic interactions between species shape patterns of adaptation, genetic diversity, and ecological dynamics. Coevolution shapes our world through mutualistic partnerships (such as pollinators and plants) (Janzen 1966), symbiotic dependencies (Thrall et al. 2007) and arms races (e.g., between host and pathogen or predator and prey) (Daugherty and Malik 2012). These multifarious interactions not only contribute to biodiversity, but also define species relationships, and accentuate the development of exaggerated traits. A better understanding of coevolution will help us to better craft conservation strategies, efficiently manage pest populations, and understand the dynamics that underpin stable ecosystems.

A coevolutionary interaction between species in which the individuals of one species benefit from the interaction and the individuals of the other is being harmed or killed and where phenotypes are continuously escalating is called an antagonistic coevolutionary arms race. For instance, a prey might evolve a defense mechanism and in response, its predator might respond by evolving a counter-mechanism, resulting in an ongoing cycle of adaptations and counter-adaptations (e.g., between herbivore resistance and plant defenses, Ehrlich and Raven 1964). Arms races between species have been observed in many species pairs such as *Taricha* newts and *Thamnophis* garter snakes (Brodie 2003), bacteria and phage (Bohannan and Lenski 2000), flax and flax rust (Dodds et al. 2006), and parsnip and parsnip webworm (Berenbaum, Zangerl, and Nitao 1986).

Antagonistic coevolution can be complex, especially when multiple interactions occur across space and time (Forde, Thompson, and Bohannan 2004). Cur-

rent theory suggests that spatial structure facilitates coevolution by constraining phenotypes in local populations that differ across larger geographical areas (Gibert et al. 2013). Other theory has described how the spatial scales of patterns in host-parasite coevolution are determined by spatial movement and the nature of the coevolutionary interaction (Week and Bradburd 2023). An ambitious attempt to describe the spatial dynamics of coevolution, called the geographic mosaic theory, postulates that spatial structure and environmental heterogeneity creates coevolutionary "hot" and "cold" spots in the arms race, leading to a mosaic of selection pressures and a hypothesized dynamic process known as "trait remixing" (Thompson 2005). Recent studies have tried to infer the contribution of coevolution to speciation and diversification (Hembry, Yoder, and Goodman 2014), but there remains uncertainty about the degree to which coevolution is a primary mechanism for diversification of traits (Eaton 2008; Butler et al. 2009; Thompson 2009; Hembry, Yoder, and Goodman 2014; Parchman et al. 2016).

The well-studied predator-prey *Taricha* newt/*Thamnophis* garter snake system is relatively well-understood and features an intriguing spatial mosaic of trait variation over a wide geographical range (Brodie, Feldman, et al. 2005; Hanifin, Brodie, and Brodie 2008; Tseng 2011; Hague, Stokes, et al. 2020; Reimche et al. 2020). On the Pacific Northwest coast of North America, various species of rough-skinned newt in the genus *Taricha* often contain tetrodotoxin that poisons predators. One of their predators, garter snakes in the genus *Thamnophis*, has developed resistance to this toxin. Levels of toxicity and resistance are highly correlated across the region: areas where newt toxicity is high, snake resistance is usually also high; and in areas where newt toxicity is low, snake resistance is usually also low. Furthermore, the snakes appear to be "winning" the coevolutionary arms race: snakes in each area seem to be able to eat the local newts with relatively little ill effect, no matter their toxicity (Hanifin, Brodie, and Brodie 2008; Feldman, Brodie Jr, et al. 2010).

These two striking observations – that the coevolutionary outcome varies strongly across geography, and that the snakes nonetheless seem to be ahead of the newts – provide additional clues about the underlying biological basis of the interaction. Despite substantial work, the underlying cause of the observed geographic mosaic of hotspots and coldspots is still unknown. Such mosaics of coevolving traits are generally thought to be the result of heterogeneous ecological factors (e.g., resource availability and/or differences in community composition), nonadaptive forces (e.g.,

local genetic drift and population structure), or both (Brodie, Ridenhour, and Brodie 2002; Hague, Stokes, et al. 2020; Thompson 2005). It would be of substantial interest to know the balance of these forces in this particular case. To explain the observation that snakes tend to be "winning", Feldman, Brodie Jr, et al. (2010) suggested that the availability of large-effect resistance alleles allowed the snakes a "potential 'escape' from the arms race". This leaves us with two questions: *Is* heterogeneity in ecological factors required to explain the strikingly correlated maps presented in Hanifin, Brodie, and Brodie (2008) and Hague, Stokes, et al. (2020), or can non-adaptive forces lead to a mosaic? And, *does* a less-polygenic architecture provide an advantage in this antagonistic coevolutionary arms race?

To answer these questions, as well as to better understand the dynamic interaction between spatial structure, genetic architecture, and coevolution, we conducted a simulation study, exploring a range of situations plausible for the *Taricha* newt/*Thamnophis* garter snake system. In addition to answering these specific questions, it is intriguing to consider other possible evolutionary outcomes, and what conditions made this outcome possible. For instance, it is easy to imagine alternate worlds in which snakes cannot eat newts (and survive), or in which snakes only eat newts in locations where newts are less toxic. The many determinants of the coevolutionary outcome include the strengths of various aspects of selection, and the genetic opportunity for adaptation. The genetic basis of these traits influences how the traits vary within populations and how they respond to selective pressures and environmental factors (Hoeksema and Forde 2008; Feldman, Brodie Jr, et al. 2010), which can lead to different evolutionary outcomes. In snakes, there are three known gene mutations that lead to high levels of tetrodotoxin resistance (Feldman, Brodie, et al. 2012; McGlothlin et al. 2014). Each of these mutations alters ion channel functioning, and so decreases the ability of a snake carrying the mutation to crawl. The frequencies of each of these mutations vary over geographical space, and do not strongly correlate with local levels of newt toxicity (Geffeney et al. 2005). On the other hand, the genetic basis of newt toxicity is still unknown. A newt's level of toxicity might be inducible and might be a result of environmental or bacterial factors. It is still unknown if it is heritable (Bucciarelli, Shaffer, et al. 2017; Bucciarelli, Alsalek, et al. 2022; Vaelli et al. 2020). However, Bucciarelli, Shaffer, et al. (2017) showed that newts retain a base level of toxicity even when kept in lab conditions, and that young newts who are more toxic take longer to develop.

To better understand the process of adaptive evolution across geographical space, this paper asks three main questions: (1) How does the genetic architecture of the traits in newts and snake affect how they coevolve? (2) Under what situations do we get spatial patterns of correlated traits in newts and snakes as we see in the real world? (3) How fast does coevolution increase resistance and toxicity in these organisms with different combinations of genetic architectures? In particular, we compare different levels of genetic variance and polygenicity using individual-based simulations of continuous geographic space. The results complement field observations by describing situations that are consistent with empirical observations, and exploring other possible outcomes.

2.2 Methods

To explore these questions, we ran spatial individual-based, forward in time simulations with SLiM (version 3, Haller and Messer 2019). The simulation had two species that we call: "newts" and "snakes", each with a quantitative trait for toxicity and resistance, respectively. The simulated range was a large rectangular map that spans 35 units in width and 140 units in length (long rectangular shape similar to the west coast, larger simulations hampered the speed of the simulation). The demographic details of this model were motivated by our current understanding of the coevolutionary interactions of the rough-skinned newt (*Taricha granulosa*) and the garter snake (*Thamnophis sirtalis*). These simulations are a simplistic representation of the system's ecological and demographic complexities, but aim to explore important aspects of the possible interaction dynamics.

The demographic model

Each simulated individual newt or snake was hermaphroditic and had a genome of 10^8 base pairs, on which phenotype-affecting mutations occurred at a rate μ base pair per generation. (Although real newts and snakes of these species are not hermaphroditic, this should not significantly affect the dynamics). Each new mutation had an "effect size" drawn from a Normal distribution with a mean of zero and a standard deviation of σ (the values of μ and σ differed between simulations runs, as described below). Demographic parameters were chosen that both newts and snakes live for about 4 generations, on average. Newts and snakes had phenotypes

we call "toxicity" and "resistance", respectively, that are determined by genotypes in a way specified below. Each species also had fixed values for recombination rate (10^{-8} crossovers per bp per generation) and for parameters controlling offspring dispersal and mate selection. The simulations used SLiM's "non-Wright-Fisher" population model (Haller and Messer 2019) with overlapping generations and fluctuating population sizes, with population dynamics described below. (We used SLiM version 3.7.1 with different "populations" for each species; since then in version 4.0 direct support for multiple species was added (Haller and Messer 2023).)

Every generation, each newt finds a nearby newt with whom to mate and produce offspring. A newt at spatial location x with k neighbors at locations $x_1, \ldots, x_k$ would choose neighbor $i$as a mate with probability

$$\frac{P(x - x_i)}{\sum_{j=1}^{k} P(x - x_j)}, \tag{2.1}$$

where $P()$ is the density of a 2-D Normal distribution with standard deviation of 1 unit. The number of offspring they produce is Poisson distributed with a mean of 1/4, and each offspring thus produced would disperse to a random location whose displacement relative to the parent was drawn from a Normal distribution with a mean of zero and a standard deviation of 1 unit. Local newt population density was computed by smoothing using a Normal kernel: local density at location x is

$$r(x) = \sum_{j=1}^{k} P(x - x_j), \tag{2.2}$$

where $x_1, \ldots, x_j$ are the locations of nearby newts (up to a maximum distance of 3 units).

Higher phenotypes were more costly than lower phenotypes. The probability that a newt at location x survives to the next time step if they have phenotype z is

$$e^{-(z/c)^2} \times \min(0.95, \frac{1}{1.2r(x)}), \tag{2.3}$$

where c is a parameter controlling the costliness of the phenotype (so that fitness decreases as the phenotype increases). The probability of survival decreases with density, due to competition between individuals of the same species. Consequently, this leads to higher mortality in areas of higher densities. The parameters are chosen so that the rough equilibrium density is one individual per unit area. Unless otherwise stated, the value of c was set to 100.

Snake reproduction, offspring dispersal, and survival used the same dynamics as the newts. Furthermore, at each time step, each nearby newt-snake pair could "encounter" each other. To do this, we iterated over all snakes in random order; choosing for each snake a set of nearby newts and resolving these encounters before moving to the next snake. The probability that a snake "encounters" a given newt at distance D is

$$I_R \times e^{-\frac{D^2}{2}}, \tag{2.4}$$

where I_R is the baseline interaction rate for nearby individuals (set to 0.05 unless otherwise noted), and that individuals that were closer were more likely to interact than individuals further apart. When a snake and a newt encounter each other, the outcome of the interaction depends on the difference between their phenotypes: if L is the newts toxicity minus the snake's resistance, then the probability that the snake survives and the newt dies is

$$\frac{1}{1 + e^{-L/10}}. \tag{2.5}$$

The form is chosen so that phenotype differences must change by around 10 units to substantially affect the interaction. If the snake does not eat the newt, the snake dies and the newt survives. For a newt to have a good chance of surviving an interaction with a snake, their level of toxicity needs to be greater than the snake's level of resistance. For each newt a snake consumes, the snake receives a "fitness bonus" of 0.1, which is added to the probability of surviving to the next time step (however, increases past 1.0 have no effect).

We used coalescent simulations produced by msprime (Kelleher, Etheridge, and McVean 2016) to generate initial genetic variation. These simulations had an effective population size of 10,000 and a recombination rate of 10^{-8}. Mutations' effect sizes were drawn from the same distribution as within SLiM, and added to the resulting tree sequence using pyslim (Rodrigues, Ralph, et al. 2023) (these had no effect on the coalescent simulation). To initialize each SLiM simulation, 300 individuals for each species were uniformly distributed across space.

Genetic Architecture

A major goal of our study is to describe how genetic architecture affects this coevolving system. By "genetic architecture" we mean the combination of mutation rate (μ) and the standard deviation of mutation effect size (σ, which we refer to

as "effect size"). We explored a range of values for mutation rate and effect size to span from a highly polygenic to an oligogenic model (Table 1). Many evolutionary dynamics depend primarily on mutational variance, which is

$$V_M = \mu \times \sigma^2. \tag{2.6}$$

We model the phenotypes for each species as exponentials of additive genetic traits. Concretely, if the sum of the effects of all mutations carried by an individual newt is G, then that newt's phenotype (i.e., toxicity) is $e^{G/10}$. Snake resistance is determined in exactly the same way. The exponential transform is used here because toxicity and resistance are non-negative, and because then mutations have multiplicative effects (i.e., increase or decrease the phenotypes by percentages).

Simulation Experiments

We ran four replicate simulations at each distinct set of parameter values, across all experiments. Table 2.1 shows the parameters used in three experiments that test how mutational variance impact the evolution of newt and snake phenotypes through different combinations of mutation rate and mutation effect size. Our first experiment set (Experiment 1) varies both mutation rates and mutation effect sizes. In the second experiment, mutation rate was for both species in each simulation, but allows the species to have different mutation effect sizes (and hence mutational variance). In the third experiment, mutational variance (V_M) was the same for newts and snakes in each simulation, although polygenicity could be different (varying mutation rate and mutation effect size).

Experiment 1 ran simulations with all possible combinations of the four genetic architectures labeled "1a" to "4a" in Table 2.1. Since there are sixteen possible combinations (e.g., snakes had 1a and newts had 3a) and we ran four replicates per combination, there was a total of 64 simulations. Across these genetic architectures, mutation rate that ranged from 10^{-8} to 10^{-11}, and the standard deviation of mutation effect sizes (σ) ranged from 0.005 to 0.05. The genetic architectures are arranged so that mutational variance increases with label (see rightmost column in Table 2.1.

In Experiment 2, we study the effects of mutational variance. To do this, we spanned the same range of parameters as in Experiment 1, in each simulation both newts and snakes have the same mutation rate (μ), but have (possibly) different

mutation effect sizes (σ), and hence different mutational variances ($V_M = \mu\sigma^2$). Note that in Table 2.1, genetic architectures with the same mutation rate are grouped together in shaded groups of four rows, and that architectures in the same group have the same numbered portion of their label. So, in each simulation as a part of this experiment, newts and snakes were each assigned a genetic architecture from the same shaded group: this led to sixteen combinations within each of the four mutation rate groups, and hence 256 simulations across 64 distinct combinations of genetic architectures.

In Experiment 3 we matched mutational variance between the species, and looked for the effects of polygenicity. This was structured similarly to Experiment 2: shaded groups in Table 2.1 now have the same V_M, and we ran simulations with each of the sixty-four possible pairs of genetic architectures for which the two are drawn from the same group (with replicates, 256 total simulations). For instance, in one simulation, newts had genetic architecture with $\mu = 10^{-8}$ and $\sigma = 0.0158$ "1g" and snakes had genetic architecture with $\mu = 10^{-11}$ and $\sigma = 0.5$ "4g"; these two have the same mutational variance ($V_M = 2.5 \times 10^{-12}$), but different polygenicity (newts have a high rate of small mutations; snakes have a low rate of large mutations). (Also note that the set of genetic architectures in Experiment 3 is the same as in Experiment 2; what differs between the experiments is which pairs are assigned to newts and snakes.)

The contribution of coevolution to phenotype change

To assess the strength of coevolution on phenotype change within the simulations (as opposed to genetic drift), we ran additional simulations after modifying the original model. We separately made two important changes, making either (1) the trait (toxicity or resistance) not heritable, and (2) the outcome of the interaction random (instead of dependent on phenotype). For (1), every new individual had a phenotype not determined by their genetics but instead as $e^{G/10}$ with G chosen independently from a Normal distribution with mean set to 3 and a standard deviation set at 2 (creating a phenotype near 1). For (2), we made the outcome of each snake-newt interaction depend on the result of a fair coin flip instead of the difference in phenotype: with 50% probability the snake eats a newt, otherwise the snake dies.

To measure the speed of coevolution, we identified the "final mean phenotype"

Experiment	Genetic Arch. (Newt/Snake)	Mutation Rate (μ)	Mutation Effect Size (σ)	Mutational Variance
1: Low to High Mutational Variance with changing mutation rate	1a	10^{-8}	0.005	2.5×10^{-13}
	2a	10^{-9}	0.05	2.5×10^{-12}
	3a	10^{-10}	0.5	2.5×10^{-11}
	4a	10^{-11}	5.0	2.5×10^{-10}
2: Low to High Mutational Variance (same mutation rate) How does increasing mutational variance	1b	10^{-8}	0.005	2.5×10^{-13}
	2b	10^{-8}	0.0158	2.5×10^{-12}
	3b	10^{-8}	0.05	2.5×10^{-11}
	4b	10^{-8}	0.158	2.5×10^{-10}
	1c	10^{-9}	0.0158	2.5×10^{-13}
	2c	10^{-9}	0.05	2.5×10^{-12}
	3c	10^{-9}	0.158	2.5×10^{-11}
	4c	10^{-9}	0.5	2.5×10^{-10}
	1d	10^{-10}	0.05	2.5×10^{-13}
	2d	10^{-10}	0.158	2.5×10^{-12}
	3d	10^{-10}	0.5	2.5×10^{-11}
	4d	10^{-10}	1.58	2.5×10^{-10}
	1e	10^{-11}	0.158	2.5×10^{-13}
	2e	10^{-11}	0.5	2.5×10^{-12}
	3e	10^{-11}	1.58	2.5×10^{-11}
	4e	10^{-11}	5.0	2.5×10^{-10}
3: Same Mutation Variance How does effects of polygenicity effect newt and snake phenotype?	1f	10^{-8}	0.005	2.5×10^{-13}
	2f	10^{-9}	0.0158	2.5×10^{-13}
	3f	10^{-10}	0.05	2.5×10^{-13}
	4f	10^{-11}	0.158	2.5×10^{-13}
	1g	10^{-8}	0.0158	2.5×10^{-12}
	2g	10^{-9}	0.05	2.5×10^{-12}
	3g	10^{-10}	0.158	2.5×10^{-12}
	4g	10^{-11}	0.5	2.5×10^{-12}
	1h	10^{-8}	0.05	2.5×10^{-11}
	2h	10^{-9}	0.158	2.5×10^{-11}
	3h	10^{-10}	0.5	2.5×10^{-11}
	4h	10^{-11}	1.58	2.5×10^{-11}
	1i	10^{-8}	0.158	2.5×10^{-10}
	2i	10^{-9}	0.5	2.5×10^{-10}
	3i	10^{-10}	1.58	2.5×10^{-10}
	4i	10^{-11}	5.0	2.5×10^{-10}

Table 2.1. A summary of all parameter sets used in simulations. Within each of the three experiments, newts and snakes were assigned all combinations of genetic architectures for which both rows are in the same group (colored blocks). Parameter sets are labeled (second column) by an identifier parameter sets in the same group have the same letter in their label.

as the average mean phenotype over the last 1000 time steps, and the "equilibrium time" as the first time mean phenotype reached 98% of the final mean phenotype. We then reported the speed as the difference in mean phenotype between the equilibrium time and time step 100, divided by the number of elapsed time steps.

Heterogeneous landscapes

We also conducted simulations where some parameters changed across geographical space. In these simulations, we used the genetic architectures from Experiment 1 (see Table 2.1) to see how newt and snake phenotypes changed when the costliness of the phenotypes or the interaction rate varied across space. In each, the relevant parameter varied across the map in a linear gradient. Each individual's location was then used to determine the correct value for that individual.

To simulate geographical variation in phenotype cost, the parameter c of equation (3.3) decreased linearly across the map, from $c = 50$ in the south to $c = 250$ in the north. So, individuals that lived in the top portion of the map had a larger fitness penalty for having a high phenotype than did individuals living in the bottom portion of the map. We ran simulations in which there was geographical variation in cost for both newts and snakes, as well as simulations in which cost varied for only one species. The fitness cost ranged from 50 to 250 and impacted the probability of snake and newt survival (equation 3).

To simulate geographical variation in interaction rate, each snake's location was used to determine the interaction rate, i.e., the parameter I_R from equation (2.4), which is the base probability with which a snake would encounter a nearby newt. This map had newts and snakes interacting more at the bottom of the map than at the top of the map: I_R ranged from 0.01 to 0.1 linearly with north-south position.

Data collection

We used code in SLiM to collect data on newt and snake phenotypes across the entire geographical area. In particular, we collected local newt and snake mean phenotypes by using SLiM's `summarizeIndividuals()` function to divide the map up into a 5×20 grid and calculating the local mean phenotypes within each grid

cell. The local mean phenotypes were then used to calculate spatial correlations between newt and snake phenotypes.

Data availability

The simulation and analysis code is accessible through github at https://github.com/Vcaudill/Coevolution.

2.3 Results

Newt and snake Evolution

We first evaluated under which situations where phenotypes were actually coevolving. If coevolution was occurring, we anticipated an increase in newt and snake average phenotypes over time. Figure 2.1 shows three common outcomes of the simulation, with mean phenotypes for newts (red) and snakes (blue): fast evolution, slow evolution, and no evolution. Simulation with relatively fast evolution had the average newt and snake phenotype rising rapidly and reaching a steady-state point before 10,000 generations, while relatively slow simulations took longer than 20,000 generations to equilibrate as "slow". In simulations with no evolution, the average phenotype of one or both species' changed very little throughout the simulation.

Figure 2.2 shows the average change phenotypes (averaged across both species) over the course of the simulation for all genetic architecture combinations from Experiment 1. When the genetic architecture of either species had low mutational variance (this was genetic architecture 1a, with lowest V_M), there was no coevolution: it's mean phenotype did not increase (or decrease), regardless of the other species' genetic architecture. However, even when one species could not evolve, the other species' phenotype still might, due to genetic drift (e.g., snake phenotype increased when newt phenotype could not increase in Figure 2.1C). The strongest effect of coevolution – i.e., the largest change in both species' phenotypes occurred when both had intermediate mutational variance and polygenicity (genetic architectures 2a and 3a). A species with the final genetic architecture (4a), which had the largest amount of mutational variance (primarily from large-effect mutations) still was able to coevolve, but showed a lower mean phenotype change.

It is interesting to note that the change in newt and snake mean phenotype is not symmetrical with reciprocal genetic architecture combinations. For exam-

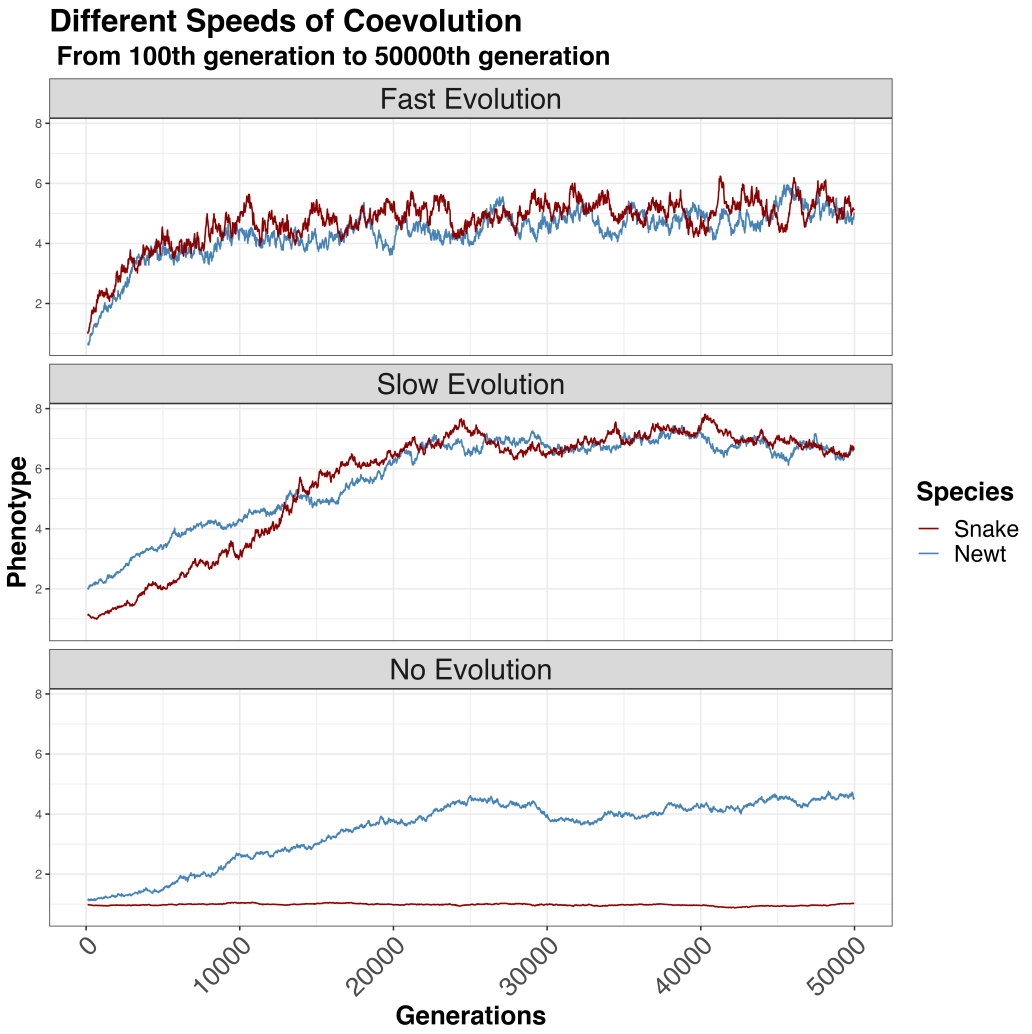

Figure 2.1. Mean phenotype dynamics over time in simulations showing three speeds of evolution; fast, slow, and no. In the top panel (fast evolution), newt (red) and snake (blue) average phenotypes go up quickly, reaching an equilibrium at around 10,000 generations. In the middle panel (slow evolution), newt and snake phenotypes rise more slowly, reaching equilibrium at around 20,000 generations. In cases of "no evolution", at least one species' mean phenotype remains flat, in this example (bottom panel), the mean newt phenotype remains flat.

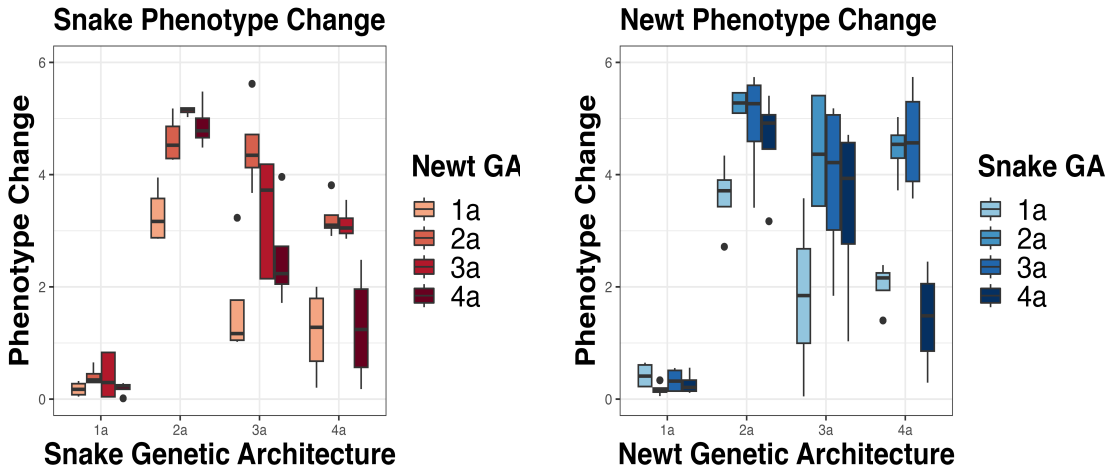

Figure 2.2. Mean phenotypes for each combination of genetic architectures in Experiment 1. Each box plot shows the range of mean phenotypes observed across time (i.e., the entire simulation) and simulation replicate for one combination of genetic architectures. The x-axis represented the genetic architecture of snake (left) or newt (right), and color represents the genetic architecture of the opposite species. Note that when either species has genetic architecture 1a (with lowest genetic variance), that species has consistently low phenotype; otherwise, their mean phenotype depends on the genetic architecture of the other species.

ple, when both newts and snakes had genetic architecture 2a the change in newt phenotype was larger than the change in snake phenotype. There was also a relationship between phenotype and population size (see Supplementary Figures A.1 and A.2): species with higher average phenotype tended to have larger population sizes, as one might expect if that species were "winning" the coevolutionary arms race. (Even though we saw newt and snake population sizes fluctuate we did not observe extinction.) In simulations without the lowest mutation variance (genetic architecture 1a), newt and snake mean phenotypes seemed to be coevolving. However, this pattern could in principle be caused by drift instead of coevolution. To exclude this possibility, we next modified key components of the simulation to examine what was driving the change in phenotype.

How much is coevolution driving phenotype change?

To verify that phenotypic changes were the result of coevolution, we altered both heritability and (separately) the snake-newt interaction while keeping every-

thing else as similar as possible, so we could see what phenotypic changes were expected in the absence of coevolutionary forces (see Methods for details). If genetic drift were responsible for the observed increase in phenotype, we would see similar increases in phenotype in these simulations in which coevolution was not possible.

Figure 2.3 shows the average change in phenotype in these simulations, averaged across both species, for different genetic architecture combinations across generations 100 to 50,000, compared to the previously described coevolutionary simulations shown in Figure 2.1. For clarity, we do not show simulations in which either species had low mutational variance (i.e., genetic architecture 1a). As expected, we saw an increase in newt and snake phenotype only in the standard simulations (orange boxplots), and not when either heritability or dependence of the interaction on phenotype was removed (other boxes). In simulations where the interaction does not depend on phenotype (green boxplots), phenotypes decreased. In these, the phenotype provided no benefit, so the mean phenotype dropped to a level determined by mutation-selection balance (where "selection" is due to the costliness of the phenotype). Simulations in which the trait was not heritable (barely visible, grey boxplots) had no change in newt and snake phenotype for all genetic architectures, as expected. (However, these simulations still provide a meaningful control from levels of spatial correlations expected in the absence of coevolution.)

Spatially Heterogeneous Landscapes

In the real world, newt and snake resistance are correlated across a broad geographical region (Hanifin, Brodie, and Brodie 2008; Reimche et al. 2020) in locations where newts are very toxic, snakes tend to be very resistant to their toxin, and vice-versa. The empirical (product-moment) correlation between the toxicity and resistance values reported in Hanifin, Brodie, and Brodie (2008) is $r = 0.77$. Do our simulations recapitulate this spatial correlation?

The answer, so far, is "no": Figure 2.4 shows that no genetic architecture combination had a comparable degree of spatial correlation. (To obtain a roughly equivalent measure from our simulations, we split the entire area into 100 smaller squares (in a 5×20 grid), calculated local mean phenotypes of each species in the squares 2.5, and computed the correlation between these two vectors of local mean phenotypes; see Figure 2.5 for an example.) Although local fluctuations driven by coevolution could in principle have created spatial correlations (Thompson 2005), it

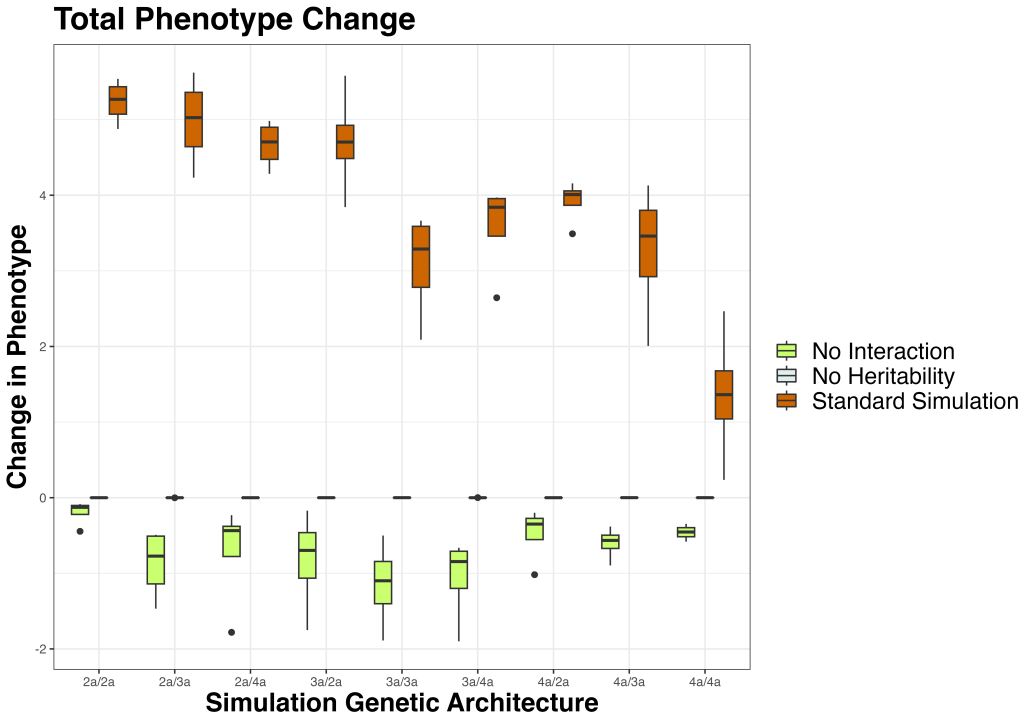

Figure 2.3. Newt and snake mean phenotypes (averaged together), comparing standard simulations to simulations without heritability or random interaction outcomes (see Methods). In the standard simulations (orange) the mean phenotype for newts and snakes increases. When there is no heritability (grey), phenotypes remain close to zero. When the newt-snake interaction outcome is random (not based on phenotype) the average newt and snake phenotypes decreased (green).

appears that such fluctuations do not appear strongly, at least across this range of parameter values in a homogeneous landscape.

However, the landscape inhabited by the real newts and snakes is not constant. It seems very plausible that the real environment is heterogeneous – for instance, (Reimche et al. 2020) found that an elevation gradient in the Sierra Nevadas was correlated to levels of toxicity and resistance in a sister species of newts and snakes. Furthermore, variation in population density, predator pressure, and/or habitat could easily lead to varying rates of interaction between newts and snakes across the landscape. So, we also ran simulations in heterogeneous landscapes. We changed, in separate simulations, the costliness of the phenotype and the rate of interaction between newts and snakes across the map. When the costliness of the phenotype varied across space, phenotypes were even more strongly correlated than in the

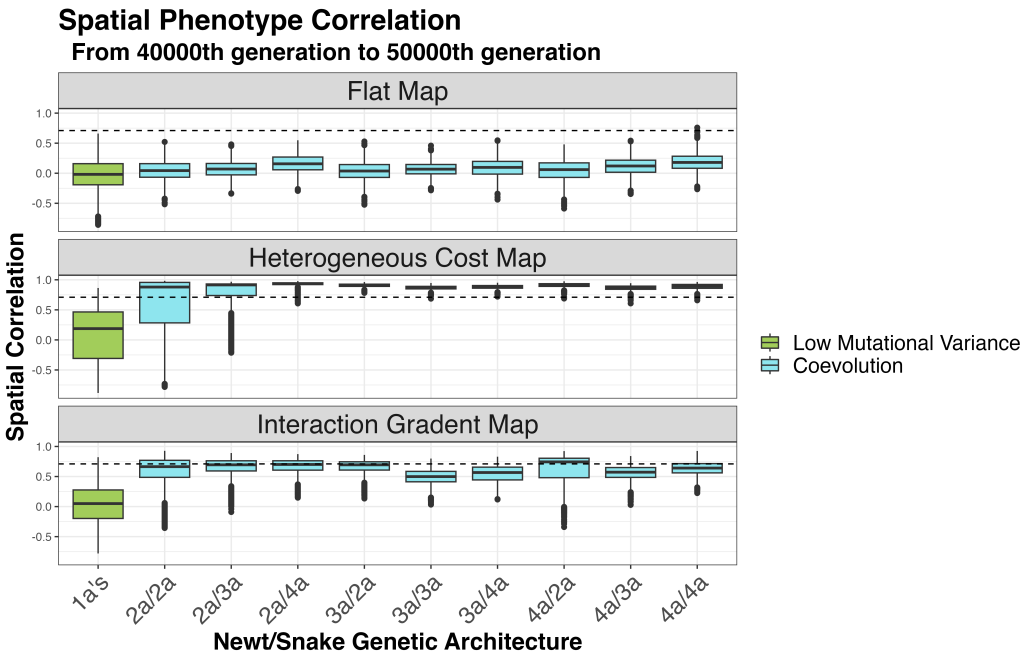

Figure 2.4. Spatial newt and snake phenotype correlations for all genetic architecture combinations, for three types of map. In each plot, the leftmost (green) boxplot displays correlations across all combinations of genetic architectures containing the lowest genetic variance (1a); other boxes (blue) show the range of spatial correlations across replicates and time steps. The dashed line shows the empirical newt and snake spatial correlation (reported by Hanifin, Brodie, and Brodie 2008). When there is no spatial heterogeneity in the simulation (top), there is little spatial correlation. Higher spatial correlations occur when newts and snake coevolve on a map with heterogeneity in phenotype costliness (middle) or interaction rate (bottom). Correlations are similar across all genetic architecture combinations with high enough mutational variance.

empirical data (mean $r = 0.84$, median $r = 0.90$, empirical $r = 0.77$). To make sure that this correlation caused by coevolution and not changing cost alone, we re-ran this simulation without a phenotype-based interaction (as described above) and saw no correlations (see Supplementary Figure A.3). Simulations in which only one species' costliness varied across the map had lower but still strong correlations (snakes vary: mean $r = 0.47$; median r = 0.49; newts vary: mean $r = 0.57$, median r = 0.70; see Supplementary Figure A.3). In addition to having high correlations between phenotypes, all simulations with varying parameters across the map had a wide range of phenotypic values, as seen in the empirical data (as in Figure 2.4).

We also saw high correlations in simulations where the interaction rate varied

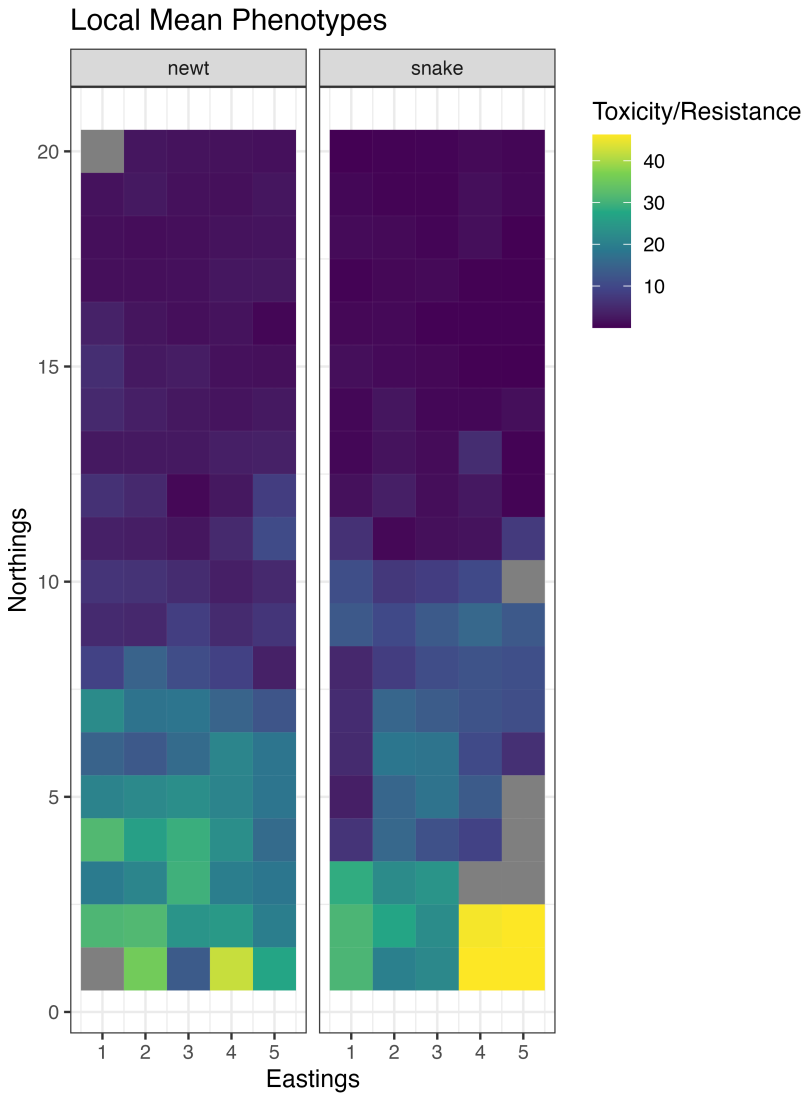

Figure 2.5. Local newt and snake mean phenotypes across the geographical area at generation 44501 in a simulation in which costliness of the phenotype was a spatial gradient (higher costliness at the top of the map). The genetic architectures used were snake: 4a and newt: 3a; other combinations are roughly similar. Each box shows the mean phenotype of individuals living within that grid cell. Brighter colors represent higher levels of toxicity or resistance.

across the map (mean $r = 0.60$). Varying interaction rate led to instances where local extinction and subsequent recolonizations occurred in small sections of the map. Adjusting the range of interaction rates or perhaps other parameters would likely increase correlations, but we did not explore parameter space further.

Effects of Genetic Architecture on Coevolution

In Figure 2.2, we saw that the coevolutionary equilibrium values of newt and snake mean phenotypes differed substantially depending on the genetic architecture (and, interestingly, in an asymmetric way). However, we saw in Figure 2.4 that genetic architecture did not affect the spatial correlation of newt and snake phenotype. How else does genetic architecture affect the coevolutionary dynamics? And, does it affect which species is "winning"? Our simulations in Experiment 1 had genetic architectures that simultaneously varied mutational variance and polygenicity, so to disentangle these two effects, we ran two additional sets of simulations in which we (Experiment 2) constrained newts and snakes to have the same mutation rate (μ) but differing effect size (σ) and hence differing mutational variance (V_M from equation (2.6)); and (Experiment 3) constrained newts and snakes to have the same mutational variance but different combinations of mutation rate and effect size.

We saw that the speed of coevolution depends more on mutational variance than it does on polygenicity for both newts (red) and snakes (blue) (Figure 2.6). When mutational variance was high the speed of coevolution was fast, essentially taking less time for the simulation to reach a steady-state. When mutational variance was lower, it took longer for the simulation to reach a steady-state.

Does it make sense to say that one species is "winning"? If one species has a higher mean phenotype than the other, then that species will more often "win" in encounters between the two species. Figure 2.7 shows mean phenotype differences between the species across the genetic architecture combinations of Experiment 1 (excluding those combinations with architecture 1a, as usual). Here, we see that population size differences and mean phenotype differences track each other (also see in Supplementary Figure A.2). We also see that newts more often have the higher mean phenotype – apparently, the asymmetry in their ecological interaction gives newts the coevolutionary edge. However, in at least one situation of Figure 2.7, the snakes (on average) had a larger mean phenotype and population

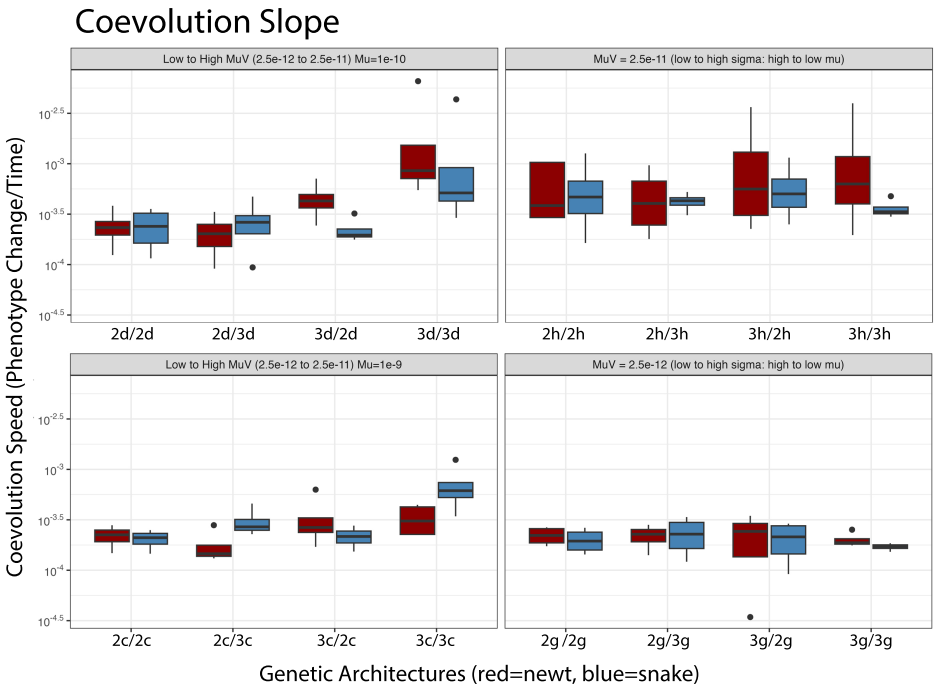

Figure 2.6. The speed of coevolution depends on mutational variance and not just the mutation rate or mutation effect size. Each boxplot shows the range of coevolution speed (see Methods) across simulation replicates at that combination of genetic architectures. **Left** plots show combinations in which mutation rate is matched between species, arranged so that mutational variance increases moving left to right. **Right** plots show combinations in which mutational variance is matched between species, arranged so that effect size (σ) increases and mutation rate (μ) decreases moving left to right. Note that the combination 3d/3d has the same amount of genetic variation as all combinations in the top-right panel (2h/2h, 2h/3h, 3h/2h, and 3h/3h) and that the combination 2c/2c has the same amount of genetic variation as those in the bottom-right panel.

size.

Supplementary Figure A.4 shows similar plots for Experiment 2, where the two species have different mutational variances but the same mutation rate, across four distinct mutation rates. In these simulations, although equilibrium phenotypes depended on the genetic architectures, in general the species with the larger mutational variance usually did *worse*, i.e., had a lower phenotype than the other species. (This is seen in Supplementary Figure A.4 by the purple box plots moving down to the right.) However, when the two species had the same mutational variance, newts usually had higher phenotype than snakes.

On the other hand, Figure A.5 shows that if both species have the same mutational variance, then differences in polygenicity do not have a strong effect on the outcome. (There are perhaps relatively minor differences in the equilibrium, but patterns are unclear.) In summary, we have good evidence for a strong effect of mutational variance, but not polygenicity, on which species is winning the coevolutionary arms race.

2.4 Discussion

In summary, we found that spatial heterogeneity in ecological factors was important for creating a spatial mosaic of correlated phenotypes. On the other hand, our simulations could not produce such coevolutionary mosaics in the absence of spatial heterogeneity, regardless of the genetic architecture. The genetic architecture did affect the dynamics of coevolution: the speed of coevolution depended primarily on mutational variance (in particular, if mutational variance was too low, the species would not coevolve). However, genetic architectures with the same mutational variance but different mutation rates and polygenicities had nearly identical coevolutionary dynamics.

One of our main goals was to identify conditions that would create the striking spatial correlations seen between the two species' phenotypes in the wild. Even though our simulations had enough space for local adaptation to occur, we saw no such geographical correlations in simulations on uniform landscapes, regardless of genetic architecture. However, environmental gradients in various aspects of the underlying biology were sufficient to create geographical correlations of similar magnitude to that seen in the wild. This result, again, did not depend strongly on genetic architecture.

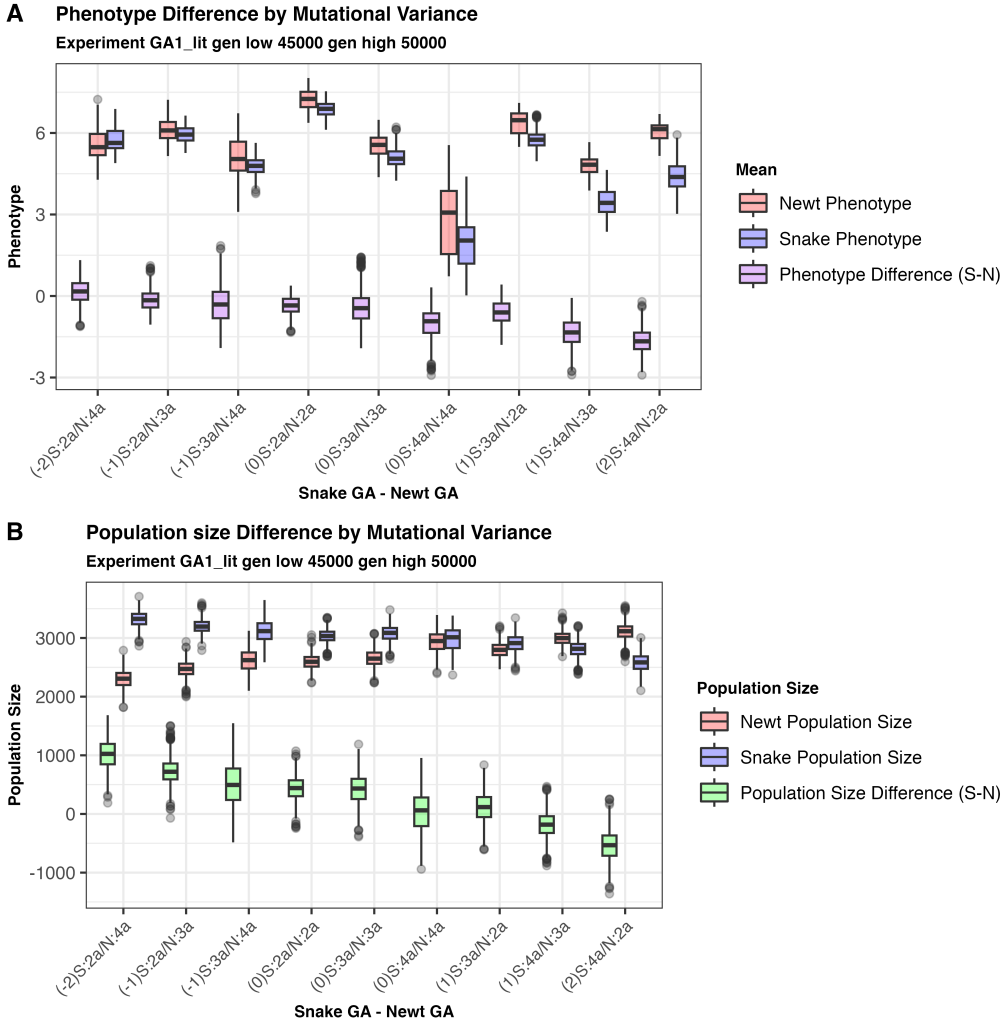

Figure 2.7. Distribution of **(A)** mean newt and snake phenotypes, and **(B)** newt and snake population sizes, and their differences (snake minus newt), from a regularly spaced set of generations between 45,000 and 50,000. All combinations of genetic architecture from Experiment 1 are shown, except those containing the lowest mutational variance (1a). The combination of genetic architectures is shown on the x-axis labels, prepended with the difference of snake and newt $\log_{10}(V_M)$ values: for instance, the leftmost set of boxplots, labeled "(-2)S:2a/N:4a", refers to simulations in which snakes have genetic architecture 2a, newts have genetic architecture 4a; and the snake's genetic architecture has 100 times less mutational variance than does the newt's.

Genetic architecture affected the dynamics of coevolution: species with larger mutational variance could increase their phenotype faster (and so "win", at least transiently). Furthermore, effects of genetic architecture on the final steady state seemed mostly attributable to mutational variance – genetic architectures with the same mutational variance but different polygenicities did not show large differences, at least over the fairly coarse scale we examined. Our results (e.g., Figure 2.7) suggest that there are indeed differences in the steady state achieved by different combinations of genetic architectures, but these differences are usually smaller than the generation-to-generation noise observed over the course of our simulations.

Relationship to Other Models of Coevolution

Many previous theoretical papers also study coevolutionary dynamics. For instance, Nuismer, Thompson, and Gomulkiewicz (2000) study how coevolution might lead to clines in allele frequency using deterministic models in either discrete or continuous space in which each species carries a single biallelic locus that mediates the coevolutionary dynamics. Nuismer, Thompson, and Gomulkiewicz (2000) found many situations in which coevolution maintained spatial variation in allele frequencies, although their model did not include escalatory (i.e., arms race) antagonism. Nuismer, Ridenhour, and Oswald (2007) develop a quantiative model of trait coevolution in which – much like in our own – quatitative traits are additive subject to stabilizing selection, and the result of antagonistic inter-species interactions is determined by a logistic function of the difference in trait values (so that each species "tries to exceed" the other). Nuismer, Ridenhour, and Oswald (2007) deal with genetic architecture in three ways: by assuming fixed genetic variance and using a quantitative genetics model; by assuming weak selection and quasi-linkage equilibrium; and by doing deterministic numerical simulations in which genetic variance is due to a fixed number of explicitly represented loci evolving under strong selection. Nuismer, Gomulkiewicz, and Ridenhour (2010) also used a quantitative genetics model (i.e., of many small-effect loci but without explicitly representing the loci) in both analytical calculations and individual-based simulations of an island model. Nuismer, Gomulkiewicz, and Ridenhour (2010) found that conditions under which traits were correlated across space differed between across types of coevolutionary interactions. They also found differences between the analytical predictions and individual-based simulation results.

Our study looks at many of the same questions, using some of the same quantitative genetics tools. However, many aspects of models that are fixed in previous work (e.g., number of polymorphic loci or population size) are emergent properties of our simulations. We focus on using individual-based simulations to test how genetic architectures affect coevolutionary trajectories with explicit models of continuous space and genetic inheritance. It would not be possible to address some of our questions using previous models: for instance, we study a broad range of genetic architecture polygenicities, which would not be possible under a purely quantitative genetics model. Furthermore, our model explicitly represents ecological dynamics – so, for instance, local extinction/recolonization is possible, unlike under the fixed-population-density models of many previous papers. Additional points of realism of our simulations that make analysis of mathematical models difficult include recombination, fluctuating population sizes, and stochasticity. These are relatively straightforward to implement thanks to the simulation software, SLiM, especially in its' newest release with improved support for interacting species (Haller and Messer 2023).

What Does This Tell us About the Newts and Snakes?

Our results support the idea that the spatial patterns in toxicity and resistance observed in *Taricha* newts and *Thamnophis* garter snakes is a result of spatial heterogeneity in some ecologically important parameters, rather than simply spatial structure and resulting decoupling of local dynamics. It has been observed, for instance, that toxicity of one *Taricha* species is correlated with elevation in the Sierra Nevada (Reimche et al. 2020). There are many possible aspects that might vary across space; variation in both trait costliness and interaction rate had similar effects. Hanifin, Brodie, and Brodie (2008) presented data across a large area of the Pacific Northwest coast of North America, from Canada to Southern California, which spans a wide range of temperature, rainfall, and biodiversity. It is easy to imagine that, for instance, the costliness to toxin-resistant snakes of being less able to crawl quickly might vary with temperature, or that varying density of other species (e.g., toads or owls that prey on newts) might lead to spatial patterns in interaction rates (Toju and Sota 2006; Craig, Itami, and Horner 2007).

Feldman, Brodie Jr, et al. (2010) hypothesized that snakes were "winning" the coevolutionary arms-race against newts due to the snakes' genes of large effect. We

found that mutational variance, rather than polygenicity, was the important determinant of which species was ahead (Figure 2.7). In fact, we found that in our simulations, newts – not snakes – had higher average phenotypes across most combinations of genetic architectures, although the effect was relatively small and somewhat hard to predict. Other scenarios are possible that would explain this situation: for instance, perhaps high phenotype values are more costly for newts than for snakes (in our simulations, we might set c smaller for newts than for snakes). Or, perhaps there are relatively hard constraints on the upper limit of tetrodotoxin production by newts, thus limiting the evolvability of the newt's phenotype (which is particularly plausible if the toxin is produced by an environmentally-acquired bacteria (Vaelli et al. 2020)).

In summary, what have we learned about the newts and the snakes? Two major takeaways are that (a) the observed spatial correlations in phenotypes is entirely consistent with the proposed evolutionary story developed in the empirical literature, but that (b) existence of these spatial patterns does not strongly constrain the set of possible underlying genetic architectures.

Spatial Patterns and Interactions

When evolution occurs across continuous space, evolutionary change and differentiation might lead to differences in different locations. However, we saw no such spatial patterns in simulations with a uniform environment. This might be due to the mixing action of migration, but even across a very large range, we still might not see substantial spatial patterns if evolution towards the same equilibrium was occurring at the same rate in all locations. (If ecological factors are heterogeneous, as in Figure 2.5, the equilibrium may differ across space.) In the real world, spatial patterns are not ubiquitous: for instance, Hoeksema and Forde (2008) did not find a relationship between spatial scale and degree of local adaptation.

Spatial patterns in biological systems are possible even without variation in the underlying environment. Even on a scale of meters, spatial patterns can spontaneously emerge from ecological dynamics, e.g., of vegetation in grasslands (Thompson and Daniels 2010; Sasaki 1997). Spatial cycling in reaction-diffusion equations is known to produce striking spatial patterns in certain circumstances (e.g., Murray 1982; Britton 1990; Schreiber and Killingback 2013). It could be, for instance, that local extinction-recolonization dynamics might play a role in the observed spa-

tial patterns. This might work, for instance, by cycles of (a) newts evolve toxicity; (b) snakes evolve resistance; (c) snakes eat most of the newts; (d) snakes lose resistance; (e) newts recolonize from a nearby relatively snake-free environment. Alternatively, if snakes have an "all-or-nothing" genetic architecture of resistance and newts have relatively little genetic variance for toxicity at any time, then one might imagine a cycle of: (a) newts evolve toxicity; (b) snakes evolve resistance; (c) snakes are sufficiently resistant that variation in newt toxicity levels provides no advantage; (d) newts lose costly toxicity; (e) snakes lose resistance. Across space, these cycles could be decoupled between regions, and barriers to dispersal would make it more likely they are not synchronized, leading to spatial patterns. However, there is no evidence that such cycles occur in the real world. In this paper, we have not attempted to search for conditions that would produce such cycling; instead, our goal was to see whether, under reasonably plausible conditions, such dynamics might occur. It would, however, be interesting to investigate.

Our simulations highlighted the importance of understanding coevolutionary systems in a broader context: in our simulations, newts and snakes evolved in response not only to the selective pressure of coevolution, but also the heterogeneous environment. This point was also made by Yoder and Nuismer (2010), who suggested that heterogeneity in the environment was also necessary for coevolutionary dynamics to increase phenotypic diversity. Furthermore, similar coevolutionary mosaics could be produced by distinct genetic and environmental conditions, and so care must be taken in ascribing specific underlying mechanisms to patterns observed in the world (Craig and Itami 2021). This work highlights the importance of considering things outside of the focal species interaction (e.g., indirect interactions).

Genetic Architecture

The underlying genetic mechanism of a trait determines how a trait can evolve and has implications on diversification and genetic variation (Hague, Feldman, et al. 2017). In our simulations, a species with sufficiently low mutational variance would never adapt, due to a lack of genetic variation. (For instance, if mutational variance for snakes was sufficiently low and newts were somewhat toxic, then rare resistance mutations in snakes might not increase snake phenotypes sufficiently on their own to meaningfully increase snake fitness.) Lack of genetic vari-

ation is thought to be a limiting factor in some potentially coevolutionary interactions (Hoeksema and Forde 2008).

Quantitative genetics models can predict how traits might change due to selective pressures (Lynch and Walsh 1998). These models typically depend only on genetic variance, without specifying the polygenicity of the underlying genetic architecture. We ran simulations under various architectures that ranged from simple (a few mutation of large effect) to complex (many mutations of minor effect), and although mutational variance indeed was the primary determining factor, polygenicity did affect steady state of the phenotype, although in complex ways. It would be interesting to model how deviations from the highly polygenic limit described by quantitative genetics affects results in practice.

Consistent with predictions of quantitative genetics, the speed of coevolution was mediated by the level of mutational variance created by the genetic architecture, rather than mutation rate or mutation effect size separately. The species that had a higher mutational variance often evolved and reached a steady state quicker. However, there were instances were the highest mutational variance limited a species' ability to coevolve. This was potentially due to large effect mutations interacting with the costliness of the phenotype: a particularly large effect mutation might cause an individual to have a phenotype so high that it perished, thus slowing the speed of coevolution.

Perhaps counterintuitively, we found that higher mutational variance seemed to put a species at a coevolutionary *disadvantage* – see e.g., Figure 2.7. To understand this, consider what is happening in the simulations represented by the furthest-left boxplots of that figure. Here, newts have much higher mutational variance than snakes, and hence a much wider distribution of phenotypes: if this is sufficiently wide, then even if the mean newt phenotype is higher than the mean snake phenotype, there will still be many newts that can be eaten by the typical snake. Furthermore, snakes have a much higher population size than newts, so a given newt is more likely to encounter a snake than vice-versa, and so the selection pressure on newts to have a high phenotype is stronger than on snakes. A similar argument applies to the other end of the parameter range.

Limitations and Continuing Questions

Our simulation was inspired by ecological interactions of real newts and snakes, but has a great many simplifications. For instance, newt and snake demographies (e.g., dispersal mechanism, birth rates, death rate, and age structure) were identical, despite *Taricha* newts living substantially longer (on average) than *Thamnophis* garter snakes in the wild. These aspects of life history impact the evolutionary trajectories of traits. Furthermore, many aspects of newt toxicity are unknown: it is thought that newt's tetrodotoxin may be produced by a bacterium (as do other tetrodotoxin-producing species, Vaelli et al. 2020; Bucciarelli, Alsalek, et al. 2022), but whether any such bacterium would be acquired from the environment or vertically transmitted is unknown. However, it seems likely that there would still be genetic variation related to the production and/or tolerance of the toxin.

In our simulation, if an interaction occurs between a newt and a snake, one perishes while the other survives. In the wild, a snake may attempt to eat a newt, fail, but both species survive (Reimche et al. 2020). Such interactions where both survive may lead to different results if trait plasticity is considered. It has been shown that a newt's toxicity can increase and remain high for some time after the newt is be disturbed (Bucciarelli, Shaffer, et al. 2017). Therefore, a failed attack could leave a newt in a better position if more snakes were nearby. Relatively little is known about how often snakes and newts interact in the wild, and how much that varies across space or time.

Finally, it would be interesting to incorporate more specifics about the known genetic architecture of tetrodotoxin resistance. We did not pursue this direction in this paper because we were interested not only in conclusions about the real system, but also broader evolutionary questions about the role of genetic architecture in these sorts of coevolutionary dynamics. We have aimed more generally to capture important aspects of the dynamics without being distracted by unnecessary details. Modeling, particularly simulation-based exploration, walks a fine line between specificity (to faithfully represent an aspect of reality) and generality (so that results are generalizable). However, future work in this area could reflect new understandings of the system, as well as more realistically model differences between the species such as genome sizes and life cycles.

2.5 Bridge

The results from this chapter highlighted the importance of the various mechanisms that impact evolutionary trajectories. In this chapter, evolution exhibited stochastic behavior, but was influenced by a range of parameters used to model a specific evolutionary system. In the next chapter I explore the real-world outcomes of evolution by examining the signals left in the genomes of 14 cichlid species. I will investigate how different mechanisms of evolution have impacted their genetic landscapes

2.6 References

Berenbaum, M. R., A. R. Zangerl, and J. K. Nitao, *Constraints on chemical coevolution: Wild parsnips and the parsnip webworm* (en), Evolution **40**, no. 6, 1215–1228, DOI: 10.1111/j.1558-5646.1986.tb05746.x.

Bohannan, B.J.M. and R.E. Lenski, *Linking genetic change to community evolution: insights from studies of bacteria and bacteriophage* (en), Ecology Letters **3**, no. 4, 362–377, DOI: 10.1046/j.1461-0248.2000.00161.x.

Britton, N. F., *Spatial Structures and Periodic Travelling Waves in an Integro-Differential Reaction-Diffusion Population Model*, SIAM J. Appl. Math. **50**, no. 6, 1663–1688, DOI: 10.1137/0150099.

Brodie, E. D., *Reciprocal Selection at the Phenotypic Interface of Coevolution* (en), Integrative and Comparative Biology **43**, no. 3, 408–418, DOI: 10.1093/icb/43.3.408.

Brodie, Edmund D Jr, B J Ridenhour, and Edmund D III Brodie, *THE EVOLUTIONARY RESPONSE OF PREDATORS TO DANGEROUS PREY: HOTSPOTS AND COLDSPOTS IN THE GEOGRAPHIC MOSAIC OF COEVOLUTION BETWEEN GARTER SNAKES AND NEWTS* (en), Evolution; international journal of organic evolution **56(10)**, 2067–2082.

Brodie, Edmund D., Chris R. Feldman, et al., *Parallel Arms Races between Garter Snakes and Newts Involving Tetrodotoxin as the Phenotypic Interface of Coevolution* (en), Journal of Chemical Ecology **31**, no. 2, 343–356, DOI: 10.1007/s10886-005-1345-x.

Bucciarelli, G.M., Farid Alsalek, et al., *Toxic Relationships and Arms-Race Coevolution Revisited* (en), Annual Review of Animal Biosciences **10**, no. 1, 63–80, DOI: 10.1146/annurev-animal-013120-024716.

Bucciarelli, Gary M., H. Bradley Shaffer, et al., *An amphibian chemical defense phenotype is inducible across life history stages* (en), Scientific Reports **7**, no. 1, 8185, DOI: 10.1038/s41598-017-08154-z.

Butler, Richard J. et al., *Testing co-evolutionary hypotheses over geological timescales: interactions between Mesozoic non-avian dinosaurs and cycads* (en), Biological Reviews **84**, no. 1, 73–89, DOI: 10.1111/j.1469-185X.2008.00065.x.

Craig, Timothy P. and Joanne K. Itami, *A geographic mosaic of coevolution between Eurosta solidaginis (Fitch) and its host plant tall goldenrod Solidago altissima (L.)* (en), Evolution **75**, no. 12, 3056–3070, DOI: 10.1111/evo.14391.

Craig, Timothy P., Joanne K. Itami, and John D. Horner, *Geographic variation in the evolution and coevolution of a tritrophic interaction: geographic coevolution in Eurosta solidaginis* (en), Evolution **61**, no. 5, 1137–1152, DOI: 10.1111/j.1558-5646.2007.00099.x.

Daugherty, Matthew D. and Harmit S. Malik, *Rules of Engagement: Molecular Insights from Host-Virus Arms Races* (en), Annual Review of Genetics **46**, no. 1, 677–700, DOI: 10.1146/annurev-genet-110711-155522.

Dodds, Peter N. et al., *Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes* (en), Proceedings of the National Academy of Sciences **103**, no. 23, 8888–8893, DOI: 10.1073/pnas.0602577103.

Eaton, C.D., *Coevolutionary Research* (en), Encyclopedia of Ecology, Elsevier, 659–663, DOI: 10.1016/B978-008045405-4.00777-1.

Ehrlich, Paul R and Peter H Raven, *Butterflies and Plants: A Study in Coevolution* (en), Evolution **18**, no. 4, 586–608.

Feldman, Chris R, Edmund D Brodie Jr, et al., *Genetic architecture of a feeding adaptation: garter snake (Thamnophis) resistance to tetrodotoxin bearing prey* (en), Proceedings of the Royal Society B: Biological Sciences **277**, no. 1698, 3317–3325, DOI: 10.1098/rspb.2010.0748.

Feldman, Chris R., Edmund D. Brodie, et al., *Constraint shapes convergence in tetrodotoxin-resistant sodium channels of snakes* (en), Proceedings of the National Academy of Sciences **109**, no. 12, 4556–4561, DOI: 10.1073/pnas.1113468109.

Forde, Samantha E., John N. Thompson, and Brendan J. M. Bohannan, *Adaptation varies through space and time in a coevolving host–parasitoid interaction* (en), Nature **431**, no. 7010, 841–844, DOI: 10.1038/nature02906.

Geffeney, Shana L. et al., *Evolutionary diversification of TTX-resistant sodium channels in a predator–prey interaction* (en), Nature **434**, no. 7034, 759–763, DOI: 10.1038/nature03444.

Gibert, Jean P. et al., *The Spatial Structure of Antagonistic Species Affects Coevolution in Predictable Ways*, The American Naturalist **182**, no. 5, 578–591, DOI: 10.1086/673257.

Hague, Michael T. J., Amber N. Stokes, et al., *The geographic mosaic of arms race coevolution is closely matched to prey population structure* (en), Evolution Letters **4**, no. 4, 317–332, DOI: 10.1002/evl3.184.

Hague, Michael T.J., Chris R. Feldman, et al., *Convergent adaptation to dangerous prey proceeds through the same first-step mutation in the garter snake Thamnophis sirtalis* (en), Evolution **71**, no. 6, 1504–1518, DOI: 10.1111/evo.13244.

Haller, Benjamin C and Philipp W Messer, *SLiM 3: Forward Genetic Simulations Beyond the Wright–Fisher Model* (en), Molecular Biology and Evolution **36**, no. 3, 632–637, DOI: 10.1093/molbev/msy228.

———, *SLiM 4: multispecies eco-evolutionary modeling*, The American Naturalist **201**, no. 5, E127–E139.

Hanifin, Charles T, Edmund D Jr. Brodie, and Edmund D III Brodie, *Phenotypic Mismatches Reveal Escape from Arms-Race Coevolution* (en), PLoS Biology **6**, no. 3, e60, DOI: 10.1371/journal.pbio.0060060.

Hembry, David H., Jeremy B. Yoder, and Kari Roesch Goodman, *Coevolution and the Diversification of Life* (en), The American Naturalist **184**, no. 4, 425–438, DOI: 10.1086/677928.

Hoeksema, Jason D. and Samantha E. Forde, *A Meta-Analysis of Factors Affecting Local Adaptation between Interacting Species* (en), The American Naturalist **171**, no. 3, 275–290, DOI: 10.1086/527496.

Janzen, Daniel H, *Coevolution of Mutualism Between Ants and Acacias in Central America* (en), Evolution **20**, no. 3, 249–275.

Janzen, Daniel H et al., *When is it coevolution*, Evolution **34**, no. 3, 611–612.

Kelleher, Jerome, Alison M Etheridge, and Gilean McVean, *Efficient Coalescent Simulation and Genealogical Analysis for Large Sample Sizes*, PLOS Computational Biology **12**, no. 5, 1–22, DOI: 10.1371/journal.pcbi.1004842.

Lynch, Michael and Bruce Walsh, *Genetics and analysis of quantitative traits*, **1**.

McGlothlin, Joel W. et al., *Parallel Evolution of Tetrodotoxin Resistance in Three Voltage-Gated Sodium Channel Genes in the Garter Snake Thamnophis sirtalis*, Molecular Biology and Evolution **31**, no. 11, 2836–2846, DOI: 10.1093/molbev/msu237, eprint: https://academic.oup.com/mbe/article-pdf/31/11/2836/13169900/msu237.pdf.

Murray, J D, *Parameter space for Turing instability in reaction diffusion mechanisms: a comparison of models*, J Theor Biol **98**, no. 1, 143–163.

Nuismer, Scott L., Richard Gomulkiewicz, and Benjamin J. Ridenhour, *When Is Correlation Coevolution?* (en), The American Naturalist **175**, no. 5, 525–537, DOI: 10.1086/651591.

Nuismer, Scott L., Benjamin J. Ridenhour, and Benjamin P. Oswald, *ANTAGONISTIC COEVOLUTION MEDIATED BY PHENOTYPIC DIFFERENCES BETWEEN QUANTITATIVE TRAITS* (en), Evolution **61**, no. 8, 1823–1834, DOI: 10.1111/j.1558-5646.2007.00158.x.

Nuismer, Scott L., John N. Thompson, and Richard Gomulkiewicz, *COEVOLUTIONARY CLINES ACROSS SELECTION MOSAICS*, Evolution **54**, no. 4, 1102–1115.

Parchman, Thomas L. et al., *Genome divergence and diversification within a geographic mosaic of coevolution* (en), Molecular Ecology **25**, no. 22, 5705–5718, DOI: 10.1111/mec.13825.

Reimche, Jessica S. et al., *The geographic mosaic in parallel: Matching patterns of newt tetrodotoxin levels and snake resistance in multiple predator–prey pairs* (en), Journal of Animal Ecology **89**, no. 7, 1645–1657, DOI: 10.1111/1365-2656.13212.

Rodrigues, Murillo F., Peter L. Ralph, et al., *PySLiM manual*, URL: https://tskit.dev/pyslim/docs/latest.

Sasaki, Akira, *Clumped Distribution by Neighbourhood Competition*, J. Theoret. Biol. **186**, no. 4, 415–430, DOI: 10.1006/jtbi.1996.0370.

Schreiber, Sebastian J. and Timothy P. Killingback, *Spatial heterogeneity promotes coexistence of rock–paper–scissors metacommunities*, Theoretical Population Biology **86**, 1–11, DOI: https://doi.org/10.1016/j.tpb.2013.02.004.

Thompson, J.N., *The Geographic Mosaic of Coevolution*, Interspecific Interactions, University of Chicago Press.

Thompson, John N., *The Coevolving Web of Life (American Society of Naturalists Presidential Address)* (en), The American Naturalist **173**, no. 2, 125–140, DOI: 10.1086/595752.

Thompson, Sally E. and Karen E. Daniels, *A Porous Convection Model for Grass Patterns.* The American Naturalist **175**, no. 1, E10–E15, DOI: 10.1086/648603, eprint: https://doi.org/10.1086/648603, PMID: 19911904.

Thrall, Peter H. et al., *Coevolution of symbiotic mutualists and parasites in a community context*, Trends in Ecology and Evolution **22**, no. 3, 120–126, DOI: 10.1016/j.tree.2006.11.007.

Toju, Hirokazu and Teiji Sota, *Imbalance of Predator and Prey Armament: Geographic Clines in Phenotypic Interface and Natural Selection.* (en), The American naturalist, **167(1)**, DOI: 10.1086/498277.

Tseng, Michelle, *In the light of evolution: essays from the laboratory and field.*

Vaelli, Patric M et al., *The skin microbiome facilitates adaptive tetrodotoxin production in poisonous newts* (en), eLife **9**, DOI: 10.7554/eLife.53898.

Week, Bob and Gideon Bradburd, *Host-Parasite Coevolution in Continuous Space Leads to Variation in Local Adaptation across Spatial Scales*, The American Naturalist **0**, no. 0, 000–000, DOI: 10.1086/727470, eprint: https://doi.org/10.1086/727470.

Yoder, Jeremy B. and Scott L. Nuismer, *When Does Coevolution Promote Diversification?* (en), The American Naturalist **176**, no. 6, 802–817, DOI: 10.1086/657048.

CHAPTER 3

CHAPTER 3 THE GENETIC LANDSCAPES OF LAKE MALAWI CICHLIDS

The experimental question and design for this project was developed by myself, Dr. Anastasia Teterina, Aidan Short, and Dr. Peter L. Ralph. I gathered all data, ran a majority of the analysis, and wrote the manuscript for this chapter along with Dr. Peter L. Ralph. Dr. Anastasia Teterina advised on many steps of data cleaning and analysis, and Aidan Short collaborated on producing some of the introgression analysis. Both aided in the interpretation of results.

3.1 Introduction

Population genetics serves as a fundamental framework in the study of evolution, elucidating the mechanisms driving genetic variation within and between populations (Lewontin 1974; Rodrigues, Kern, and Ralph 2024). Through the examination of genetic drift, gene flow, mutation, recombination, and natural selection, population genetics has provided insights into how populations evolve and adapt over time. These different forces have all played a role in shaping patterns of variation in genomes, and describing the "balance" between these forces is an important emerging issue (Renaut et al. 2013).

Populations can be subject to different evolutionary forces at any point in their lineages. These evolutionary forces can leave signals in the genetic landscapes, but these signals are often combatant and easily distorted or complicated to measure (Van Doren et al. 2017; Wang et al. 2020; Rodrigues, Kern, and Ralph 2024). Advances in sequencing technology have allowed examination of genetic landscapes of closely related species, revealing peaks and valleys in genetic diversity and divergence that were similar across species and species pairs. These peaks and valleys were once thought to be markers of speciation (Cruickshank and Hahn 2014), but can also occur due to the process of shared evolutionary history (Wang et al. 2020).

Patterns of genetic diversity and divergence have long been used to uncover the molecular signs of evolution. For instance, comparisons between synonymous and non-synonymous changes in coding sequence, either within species (the McDonald-Kreitman test, McDonald and Kreitman 1991) or between species (dN/dS, Kimura 1977; Kryazhimskiy and Plotkin 2008), can indicate that evolution is either acting to remove or encourage new mutations in a particular gene or set of genes. The

signal provided by these measures works primarily because natural selection directly affects the chance that each new mutation under selection rises in frequency and goes to fixation: "positive selection" tends to increase frequencies of new alleles, while "negative selection" tends to decrease them. More indirectly, selection often decreases genetic diversity in regions near to the loci at which alleles are under direct selection (Kaplan, Hudson, and Langley 1989; Hudson and Kaplan 1995; Charlesworth, Morgan, and Charlesworth 1993). This predicts a negative correlation between genetic diversity and recombination rate that has been widely observed (Langley et al. 2012; Corbett-Detig, Hartl, and Sackton 2015; Burri 2017). However, these indirect (or, "linked") effects of selection only affect levels of within-species genetic diversity – only the direct effects of selection affect the rate at which between-species fixations accumulate (Birky and Walsh 1988). In summary, positive selection tends to decrease within-species diversity but increase the rate of accumulation of between-species divergence as compared to neutral regions, while negative selection tends to decrease both. Furthermore, linked effects are expected to be correlated with recombination rate, while direct effects are not. These differences provide a possible avenue to understand the relative intensities of positive and negative selection on a genomic scale in a particular group of taxa (and for non-coding regions), which is a long-standing open question in evolutionary biology.

However, the problem of deciphering signals is made more difficult because of a complex array of other processes and patterns that can affect genetic diversity and divergence differently along the genome. For instance, the direct effects of positive selection increase the *rate* of between-species divergence, but since divergence also depends on diversity within the ancestor, regions with a higher intensity of positive selection may not actually have higher divergence. This makes it necessary to compare divergences between increasingly distant taxa, rather than in only one or two species. An increased mutation rate in one region of the genome can resemble the net effects of direct selection, except that higher mutation rates will increase diversity. GC-biased gene conversion also resembles the direct effects of positive selection (Duret and Galtier 2009), but its known dependence on particular nucleotide signatures makes its effects identifiable (Rodrigues, Kern, and Ralph 2024) rather than decreasing it. Balancing selection can strongly increase diversity, possibly over large regions if recombination is suppressed, e.g., as for an inversion (Kirkpatrick and Barton 2006; Kirkpatrick 2010).

Finally, much of this theory and intuition for genetic diversity and divergence

is based on models of a single species tree. However, even some well-separated species interbreed, and although the picture of species as distinct, randomly-mating units (possibly exchanging genetic material with each other) is a good description for many purposes, it is not clear how well that description applies to some groups of taxa. Furthermore, introgressing genetic material is subject to particular selective forces: for instance, gene flow may be more directional in gene-rich regions if one species has a higher level of genetic load (Barton and Gale 1993; Bierne et al. 2002; Harrison and Larson 2016) or if one species has adapted to an environment the other is moving into. Indeed, introgression has been shown in some well-studied cases to be correlated with gene density and/or recombination rate (Harris and Nielsen 2016).

Since there are a large number of confounding factors, to take full advantage of the evolutionary information provided by genetic diversity and divergence, we need to compare across regions of the genome, chromosomes, and species pairs of increasing divergence. In particular, it is important to compare across different groups of taxa that might have different confounders (e.g., gene density and recombination rate might be positively or negatively correlated) and different balance of forces (e.g., introgression might be more or less important). This will help us understand what signals are repeatable, how to control for various biological factors, and how to extract a reliable signal.

Several studies of closely related taxa have found correlations along the genome in genetic diversity (Burri et al. 2015; Irwin et al. 2016; Van Doren et al. 2017; Battey 2020). Studies have found such correlations not only between genetic diversity within species but also between genetic divergences between species in *Ficedula* flycatchers (Burri et al. 2015), stonechats-flycatchers (Van Doren et al. 2017), great apes (Rodrigues, Kern, and Ralph 2024) and *Diplacus* bush monkeyflowers (Stankowski et al. 2019). These correlations of diversity and divergence persist over much longer periods of time than what is expected under neutrality, implying that these correlations are a reflection of shared processes acting in similar ways along the genomes of these separate species.

The great apes studied by Rodrigues, Kern, and Ralph (2024) represents a very well-studied system in which five species (human, chimpanzee, bonobo, gorilla and orangutan) are well-separated in evolutionary time: typical inter-node distances in the phylogeny are upwards of $10N_e$, so there is very little shared ancestral variation retained between the species. However, a number of subspecies are more

closely related, giving resolution at shorter time scales. Correlations in diversity or divergence between closely related taxa were around 0.9, and decayed to around 0.4 over the depth of the phylogeny (about 12 million years). The conclusion (based on comparison to simulations) was that correlations were likely maintained by a combination of positive and negative selection, as well as a possible contribution of mutation rate variation. Introgression is not thought to have a strong effect on the genomes, and the genomes are generally very syntenic, which simplified the analyses but did not provide the opportunity to study the effects of introgression or rearrangements.

The group of *Diplacus* bush monkeyflowers studied by Stankowski et al. (2019) is a relatively recent species radiation around 1Mya in (or near) southern California. Taxa show varying degrees of ability to interbreed, but many are geographically isolated, so the amount of ongoing hybridization is unclear. However, at least some taxa have known hybrid zones (Sobel and Streisfeld 2015). In *Diplacus* monkeyflowers, gene density is negatively correlated with recombination rate, whereas in great apes, this correlation is positive. Since the strength of linked selection increases with gene density and decreases with recombination rate, this suggests that the strength of linked selection varies much more across the genome in monkeyflowers than it does in great apes. However, monkeyflower genomes are smaller than those of great apes. Stankowski et al. (2019) found that the correlation between F_{ST} and gene density strongly *increases* with the time separating the taxa, while the correlation of F_{ST} and recombination rate strongly decreases. Furthermore, the correlation between within-taxon genetic diversity and between taxon divergence (d_{XY}) decreases from nearly 1 for closely related taxa to near zero for the most distant taxa. Direct comparison between Stankowski et al. (2019) and Rodrigues, Kern, and Ralph (2024) is difficult because they use different measures of genetic diversity (e.g., F_{ST} versus d_{XY}) and of phylogenetic distance (time).

Both of the studies followed similar procedures with known phylogenetic relationships in different datasets highlighting unique aspects of genome evolution. To add to this growing field, we examined the genetic landscapes of Lake Malawi haplochromine cichlids. Similar to monkeyflowers, Lake Malawi cichlids are a recent radiation with an estimated 500-860 species that diverged within the last 800,000 years (Brawand et al. 2014; Santos, Lopes, and Kratochwil 2023). The repeated rapid expansion of cichlids is thought to be caused by cycles of expansion, diversification, and sexual selection (Kocher 2004; Wagner, Harmon, and Seehausen 2012;

Brawand et al. 2014). The genetic variation needed for such a rapid burst of speciation might be due by frequent hybridization (Meier et al. 2017; Malinsky, Svardal, et al. 2018; Svardal, Salzburger, and Malinsky 2021). This rapid expansion with a large amounts of gene flow has made inferring genetic relationships difficult (Kocher 2004; Malinsky, Svardal, et al. 2018).

In addition to difficulties of reconstructing cichlid phylogenitic history, cichlids are known to have large inversions (Conte et al. 2019; Svardal, Salzburger, and Malinsky 2021), structural variation (Conte et al. 2019; Svardal, Salzburger, and Malinsky 2021; Quah et al. 2024), and high rates of sex chromosome turnover (Conte et al. 2019). Interestingly, cichlids are known to have low levels of diversity and divergence even though they exhibit many phenotypic and behavioral differences (Svardal, Salzburger, and Malinsky 2021). In this study we aim to investigate if cichlids, a recent radiation, with large inversions and great amount of introgression, show correlations in genetic diversity and divergence along the genome similar to those seen in the great apes and *Diplacus* monkeyflowers. We also investigate how genetic diversity and divergence interact with intrinsic properties of genomes (gene density, recombination rate, site accessibility, gene repeats).

3.2 Methods

Samples and Alignment

We used publicly available data from Malinsky, Svardal, et al. (2018), by downloading sequencing data for each species with multiple individuals using SRA-toolkit (fastq-dump.3.0.0), resulting in 53 individuals in 14 species (sample sizes in parentheses): *Champsochromis caeruelus* (2), *Chilotilapia rhoadesii* (4), *Copadichromis virginalis* (7), *Ctenopharynx intermedius* (3), *Dimidiochromis strigatus* (2), *Protomelas ornatus* (4), *Fossorochromis rostratus* (4), *Hemitilapia oxyrhynchus* (2), *Lethrinops lethrinus* (2), *Mylochromis anaphyrmus* (5), *Otopharynx speciosus* (2), *Placidochromis subocularis* (7), *Placidochromis* cf *longimanus* (5), and *Tremitochranus placodon* (4) (accession numbers in supplementary table B.1). Of these species, 13 are shallow benthic and one, *C. virginalis*, is utaka.

The 53 cichlid genomes were aligned to the *Metriaclima zebra* reference genome (Ensembl Release 110, GCA 000238955.5) through the standard BWA-GATK pipeline (https://gatk.broadinstitute.org). Reads were aligned using BWA-mem (v. 0.7.17-

r1188, Li and Durbin 2009) and unmapped reads were removed using samtools (v. 1.6, Danecek et al. 2021). PCR dupicates were marked using picard v. 2.27.5-4-g5295289-SNAPSHOT (http://broadinstitute.github.io/picard/). Coverage for each individual was estimated using samtools depth. A list of indels were created using GATK v. 3.7 RealignerTargetCreator and then realigned using IndelRealigner. Variants were called with GATK HaplotypeCaller v. 4.3.0.0 (McKenna et al. 2010). Indels were marked by SelectVariants and then retained with VariantFiltration. Repeats were masked based on the available repeat mask for the genome from Ensembl using generate_maked_ranges.py (https://gist.github.com/danielecook/cfaa5c359d99bcad3200). High-quality indels and 10 base pairs around them, and regions with bad coverage, specifically regions where the coverage was less than three and more than three times the mean coverage for that sample in at least 75% of the samples were also masked. Chromosomes were divided into 1Mb non-overlapping windows, and any windows that had less than 300Kb nonmissing data were removed from the analysis using bedtools (Quinlan and Hall 2010).

For each 1Mb window we calculated accessibility, gene repeat density, gene density, and recombination rate. "Accessible sites" are sites that are not masked (i.e., no repeats, indels, or duplicates); we summarized "accessibility" for each window as the number of accessible sites divided by 1 million. Gene repeat density was calculated using bedtools coverage with the file generated from generate_maked_ranges.py and and the accessibility mask. Gene densities were also estimated with bedtools coverage, a gene density file from the *M. zebra* Ensembl release (https://ftp.ensembl.org/pub/release-110/gtf/maylandia_zebra_UMD2a.110.gff3), and again the accessibility mask. To obtain a genetic map, we obtained marker sequences from the *M. mbenjii* × *A. koningsi* map from Conte et al. (2019), and used blast (v. 2.2.29, Camacho et al. 2009) to find the coordinates of the markers on the *M. zebra* reference genome. We then calculated a rough estimation recombination rates for each window (in cM/Mb) using genetic map distances from the genetic map of *M. mbenjii* x *A. koningsi* from Conte et al. (2019). Two outlier recombination rates were excluded from the study. All cM/Mb maps provided in appendix E.

Phylogeny

We first generated a phylip file from our filtered VCF file with vcf2phylip.py (https://github.com/edgardomortiz/vcf2phylip) with default parameters, and then

generated a phylogenetic tree with all individuals using iqtree (v. 2.2.5 Nguyen et al. 2015), using *C. virginalis* as the outgroup and 1000 bootstrap replicates. The best fitting model selected by iqtree was TVM+F+R9, and the resulting phylogenetic tree grouped all individuals of the same species together (see Figure B.1).

To calculate a phylogenetic distance between species, we built a second tree using only one individual per species in the same manner as above, using iqtree's best fitting model (TVM+F+I+R3). The individual with the best coverage was selected for each species. For visualization purposes, we also estimated phylogenetic trees separately by linkage groups.

Population genetics

We calculated genetic diversity within species ("π") and genetic divergence between species ("d_{XY}") in each 1Mb window using the windowed_divergence and windowed_diversity functions from scikit-allel (v. 1.3.5, Miles et al. 2024). This calculates both diversity and divergence as the proportion of accessible sites in the window that differ between a pair of individuals, averaged either over pairs of individuals from the same species (for π) or over pairs of individuals with one from each species (for d_{XY}).

We calculated Spearman's correlations between genetic diversity landscapes, divergence landscapes, and diversity and divergence landscapes in python. Spearman's correlations between the genome functions (gene density, gene repeats, accessibility, and recombination rate) and genetic divergence were also calculated.

We plotted phylogenetic trees using the ape package (v. 5.7-1, Paradis and Schliep 2019) in R (v. 4.3.2, R Core Team 2021) and Rstudio (v. 2023.09.1+494). Distance between tips were calculated using cophenetic and dispRity (v. 1.8, Guillerme 2018).

Figure 3.1 shows how phylogenetic distances between different pairs of diversity and/or divergence are calculated in this study. The phylogenetic distance between species A and B, $D(A, B)$, is used for π vs π correlations, and is shown in Figure 3.1A:

$$D(A, B) = d(A, B), \tag{3.1}$$

where A and B are populations and d is phylogenetic distance (shown in red in Figure3.1). The distance between a pair of species, A and B, with another species C, $D(AB, C)$, is used for d_{XY} vs π correlations, and is shown in Figure 3.1B). $D(AB, C)$

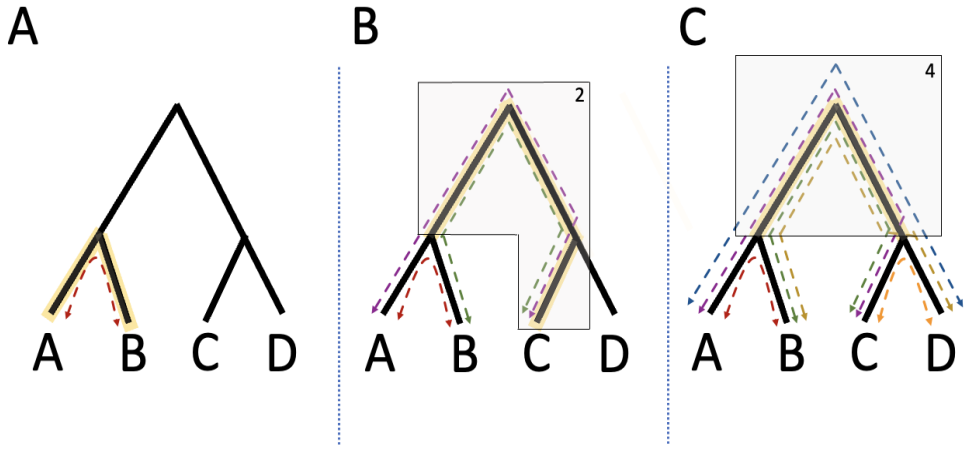

Figure 3.1. An example phylogenetic tree with four populations (A, B, C, D). π is calculated within one population, while d_{XY} is calculated between two populations. Yellow shading highlights the path used to calculate phylogenetic distance in equations (3.1), (3.2), and (3.3). In **(A)** the red path is the distance between population A and B also seen in equation (3.1). In **(B)** the yellow path highlights the distance between the node between A and B, and C. To calculate phylogenetic distance we can use equation (3.2), which takes the path between A and C (purple) and B and C (green) and takes out the distance between A and B (red). Since the two paths overlap, the two is divided out. In **(C)** the yellow path highlights the distance between the node between A and B, the node between C and D. To calculate phylogenetic distance we can use equation (3.3), which takes the path between A and C (purple), B and C (green), A and D (blue), B and D (yellow) and takes out the distance between A and B (red) and C and D (orange). Since the four paths overlap, the four is divided out.

is the phylogenetic distance from species C to the path between A and B. This is computed as:

$$D(AB, C) = \frac{d(A, C) + d(B, C) - d(A, B)}{2}, \tag{3.2}$$

where again A, B, and C are populations. This calculation adds the distance between A and C (purple), the distance between B and C (green), and then subtracts the distance between A and B (red). Two paths overlap so a two is divided out. The distance between two pairs of species, A–B and C–D, denoted $D(AB, CD)$, is used for d_{XY} vs d_{XY} correlations, and is shown in Figure 3.1C). It is the distance in the tree between the path from A to B and the path from C to D, if these do not overlap; if they do, it is minus one times the length of their overlap. It is computed

as follows:

$$D(AB, CD) = \frac{d(A,C) + d(A,D) + d(B,C) + d(B,D) - 2d(A,B) - 2d(C,D)}{4}, \tag{3.3}$$

This calculation adds the distance between A and C (purple), the distance between A and D (blue), the distance between B and C (green), the distance between B and D (yellow), then subtracts the distance between A and B (red) and between C and D (orange) twice. There are four paths overlapping so a four is divided out.

Data visualisation and analyses were completed using ggtree (v. 10.3.0, Xu et al. 2022) and ggplot2 (v. 3.4.4, Wickham 2009), phytools (v. 2.1-1, Revell 2024), reshape2 (v. 1.4.4, Wickham 2007), and tidyr (v. 1.3.0, Wickham, Vaughan, and Girlich 2023).

Introgression Analyses

To estimate the impact of introgression on genome-wide levels of genetic divergence and diversity, we first tested for evidence of introgression by calculating Patterson's D (Green et al. 2010) for all possible trios of in-group taxa using the Dtrios command in Dsuite (Malinsky, Matschiner, and Svardal 2021) based on the relationships inferred from genomic data in our phylogenetic tree with *C. virginalis* as the outgroup. To estimate the history of hybridization among these taxa we then used Dsuite to identify recent and historical signatures of hybridization by calculating the f-branch statistic based on all significant values of Patterson's D (see supplementary index I). To estimate genome-wide variation in introgression, we calculated the admixture proportion (f_d) (Martin, Davey, and Jiggins 2015) in 1Mb non-overlapping windows for all possible trios of ingroup taxa that test for introgression between a pair of taxa that display high levels of recent and historical introgression based on f-branch values. To calculate f_d we used the ABBABABAwindows.py Python script (https://github.com/simonhmartin/genomics_general) with three species (a "trio" P_1, P_2, and P_3) and an outgroup (*C. virginalis*). f_d measures shared ancestry between P_2 and P_3 that is not shared by P_1, and hence quantifies introgression from P_3 into P_2 that has occurred since the divergence of P_1 and P_2. We also calculated f_d by linkage group

3.3 Results

Our study investigated the genetic landscape of cichlids through patterns of similarities and differences in diversity and divergence in 1Mb windows along the genome. Out of a total of 771 windows, 30 were excluded due to low levels of accessibility (see methods). We compared how the relationship between diversity and divergence changed over a range of phylogenetic distances. Additionally, we examined the relationship between diversity and divergence with other components of the genome (gene density, recombination, gene repeats, site accessibility), and how these relate to measures of introgression.

Divergence across the genome

Divergence along all linkage groups in the genome is shown in Figure 3.2A. Each point shows divergence between one pair of species over a 1Mb window (genetic diversity for all linkage groups is shown in supplementary figure B.2). Divergences between species were low but varied substantially along the genome, ranging from about 0.1% to 0.4%, while diversity ranged from 0.05% to 0.3% similar to the levels of diversity seen in Malinsky, Svardal, et al. (2018).

The divergences between different pairs of taxa along each linkage group were highly correlated (plots of each individual linkage groups are shown in appendix C and D). For instance, Figure 3.2B zooms into a typical linkage group, LG5, with lines showing the divergence between each species pair. In this linkage group, as in most others, divergence from all other taxa to *C. virginalis* is higher than divergences between other pairs of taxa (comparisons to *C. virginalis* shown as green lines). We use *C. virginalis* as the outgroup, although it does not show the highest divergence to other taxa on all linkage groups. Divergence differs more between different windows along the genome than it does between different pairs of taxa, unlike what was seen in great apes (Rodrigues, Kern, and Ralph 2024) or monkeyflowers (Stankowski et al. 2019). However, the relative ordering of genetic diversity between different species pairs is highly consistent across windows (a feature also seen in both great apes and monkeyflowers).

However, some linkage groups present different patterns than we saw on linkage group 5. Figure 3.2C zooms into linkage group 11, where divergences between each pair of species fall into one of three distinct groups. Furthermore (although

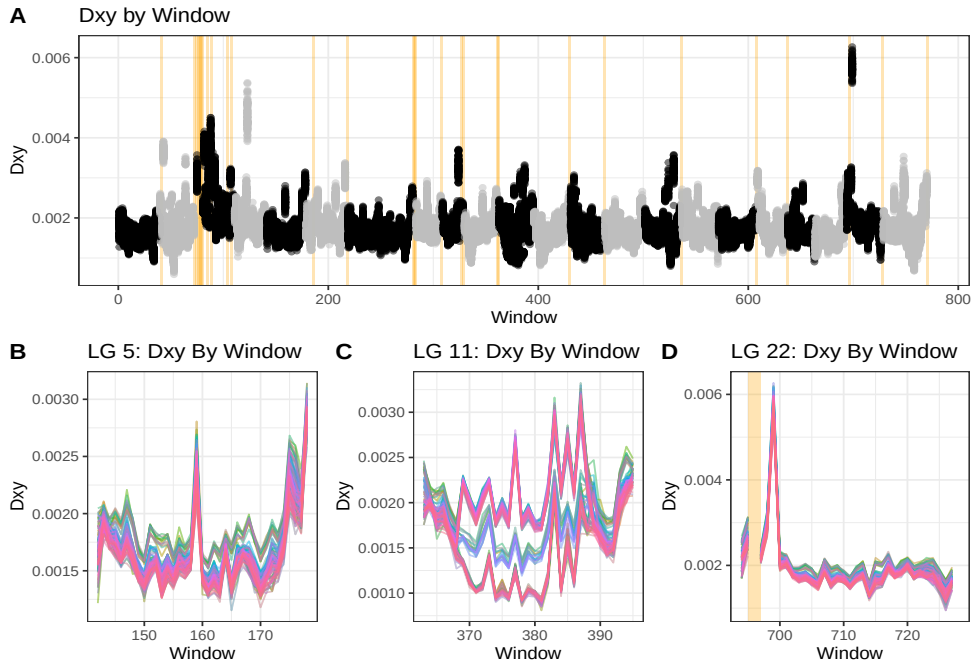

Figure 3.2. **(A)** Divergence in 1Mb windows between all species pairs on all linkage groups, with divergence for each species pair in each window shown as a point. The 30 windows (out of a total of 771) omitted due to low accessibility are shown with a grey background. **(B,C,D)** Divergences in 1Mb windows for linkage groups 5, 11, and 22 respectively, with divergences for each species pair shown as a single line.

this is not evident in the plot), each individual species' comparisons to other species all fall either in the top and bottom, or in the top and middle groups. For instance, comparisons to *C. virginalis* (which are all greenish lines) all fall in either the top or middle group. This is consistent with the presence of an inversion, and indeed, linkage group 11 is known to harbor a large inversion that is polymorphic between species (Conte et al. 2019), and is associated with bower-building behavior (York et al. 2018). Interestingly, divergence between the two orientations is similar to divergence outside of the inversion: an alternative pattern would show that divergence in an inversion is elevated due to long-term maintenance under balancing selection. One interpretation of the decreased divergence within both orientations of the inversion could be that the inversion was polymorphic in the ancestor or occurred early in the radiation (as suggested by Conte et al. (2019)), and has since been prevented from introgressing between taxa by sexual selection, unlike much of the rest of the genome.

Figure 3.2D zooms into linkage group 22, the linkage group with the window having highest divergences (of those that were not removed). A window in this linkage group shows much higher divergences between all pairs of taxa than do any of the other windows; since the difference between this an other windows is so high, it creates particularly high correlations between species comparisons on this linkage group.

We do not see patterns as striking as linkage group 11 on any other linkage group, even though Conte et al. (2019) reported large inversions on additional linkage groups (2, 9 and 20). There is also the possibility that these patterns are difficult to perceive due to sex-determining loci (reported on linkage groups 7, 9 and 11), large region of low recombination (linkage groups 9), chromosomal fusion (linkage group 7) (Conte et al. 2019).

Estimating phylogenetic distance

The whole-genome phylogeny estimated by iqtree is reasonably well-resolved, and shown in Supplementary Figure B.3. Since it is thought that the relationships between these cichlid species are not well-represented by a single phylogeny, we also experimented with building local phylogenies in various ways. In particular, we also estimated phylogenies separately using each linkage group; the resulting trees differed somewhat by linkage group, but were generally similar to Figure B.3 on all linkage groups except LG11, where the inversion is visible (see appendix H). However, using linkage-group-specific phylogenies does not affect downstream results substantially, so we proceeded with using the single, whole-genome phylogeny.

We do not claim that this phylogeny completely represents the history of these cichlid species: at minimum, there are known to be a number of substantial introgression events across the tree (so a phylogenetic network might be more appropriate)(Santos, Lopes, and Kratochwil 2023). However, for downstream analyses it will be useful to have a measure of "time" that separates different comparisons, and this phylogeny should give us a good proxy for that notion.

Correlations between landscapes

Figure 3.3 shows all pairwise correlations between landscapes of diversity and divergence, plotted against the distance between them in the phylogeny (calculated

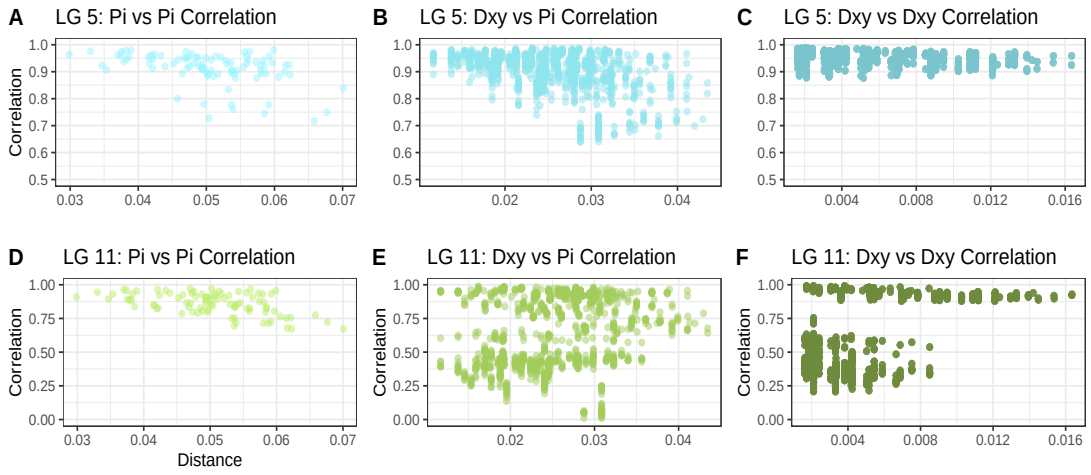

Figure 3.3. Correlations between different landscapes of genetic diversity or divergence plotted against separating phylogenetic time (see text for details), for linkage groups 5 and 11. **(Left: $\pi - \pi$)** Each point shows the correlation between landscapes of genetic diversity (π) for two taxa along linkage group 5, plotted against the time separating the two taxa (i.e., the sum of the times from each of the two taxa back to their MRCA). **(Center: $\pi - d_{XY}$)** Correlations between a landscape of diversity in one taxon and divergence between two other taxa, plotted against time from the first taxon to the path connecting the other taxa in the phylogenetic tree. **(Right: $d_{XY} - d_{XY}$)** Correlations between two landscapes of diversity, plotted against separating phylogenetic distance (see text). Only comparisons with positive separating phylogenetic distance are shown.

using equations (3.1), (3.2), or (3.3)) on linkage groups 5 and 11. For instance, Figure 3.2B shows the 14 landscapes of genetic diversity along linkage group 5, while Figure 3.3A shows the 91 pairwise correlations between these 14 landscapes; the x coordinate for the correlation between $\pi(A)$ and $\pi(B)$ is the distance between species A and species B in the phylogeny. The x coordinate for correlations between diversity and divergence (Figure 3.3B) and between two landscapes of divergence (Figure 3.3C) is found in a similar way as the distance in the phylogeny between the paths connecting the species (see Methods).

Correlations are generally high, and then decrease with time. For instance, correlations between landscapes of diversity on linkage group 5 all have a correlation coefficient that starts around 0.95 and that drops off steadily with time. This makes sense, as the effects of any process that produces or maintains similarities between landscapes are expected to weaken with time: shared ancestral

diversity will become less important as coalescence occurs in each taxon. Furthermore, we might expect selection to be acting with a very similar intensity on the genomes of two recently diverged species, but selection will act more differently across the genome the more the species diverge. We also see that correlations between (within-species) genetic diversity are much noisier than correlations between (between-species) genetic divergence, a feature also noted in Rodrigues, Kern, and Ralph (2024). This makes sense as local inbreeding and drift are expected to have a larger effect on these patterns for diversity.

Although correlations decrease with time for both linkage groups, the correlations with divergence for linkage group 11 form two distinct groups, each decreasing. Furthermore, even some closely related comparisons on LG11 have correlations near 0, something not seen on LG5. This is likely explained by the presence of an inversion: as we saw in Figure 3.2, divergences between each pair of species fall into one of three distinct groups, corresponding to whether the two species have the same orientation of the inversion (and which orientation they have), or different orientations. In other words, the two orientations of the inversion would correspond to two long haplotypes, and we see the three possible comparisons between the two haplotypes. We do not see a similar pattern for within-species diversity, likely due to landscapes of diversity being strongly affected by genetic drift.

The example of linkage group 11 is useful because this gives us a view at a large and well-resolved scale of the underlying process that acts mostly on a much smaller genomic scale. For instance, if the inversion on LG11 was a much smaller portion of the chromosome, the effect on correlations would be correspondingly smaller: the two clouds in Figure 3.3E or F would be much closer together. However, if there are many instances of small regions of the genome (in inversions or not) that move between species, the result could produce a more gradual decrease of correlation with time (as in Figure 3.3B or C).

Although inversions have been reported on other linkage groups, we do not see clearly separated groups like those shown for LG11 in Figure 3.2. However, we do see patterns of banding similar to those of LG11 in Figure 3.3 for other linkage groups, most strongly, LG2 and LG13 (see appendix F). It seems likely that the same mechanism underlies these patterns.

Correlation between divergences of pairs of great ape species is only about 0.75 for the most closely related comparisons and decays to around 0.5 for the correlations between the most distant comparisons (about 10^6 generations separated in

the tree). Correlations between cichlids, on the other hand are, excluding linkage groups with likely inversions, mostly much higher: above 0.8 even for the most distant comparisons (which are separated by a few hundreds of thousands of generations). One reason for this is likely the relevant time scale: the largest phylogenetic distances between cichlids is more comparable to the smaller phylogenetic distances between great apes. However, it is likely that ongoing introgression between cichlids also plays a strong role.

Correlations with genomic features

Genetic diversity and divergence are expected to be correlated with genomic features such as gene density and recombination, because these features regulate the strength of linked selection. Indeed, as shown in Figure 3.4, the genomic features we studied (gene density, recombination rate, accessibility, and repeat content) are often correlated with landscapes of diversity and divergence: on many linkage groups, correlations are in the range 0.2–0.5. There were significant correlations between averaged windowed divergence and gene repeats (p-value 3.05×10^{-7}, correlation 0.188522), averaged windowed divergence and accessibility (p-value $< 2.2 \times 10^{-16}$, correlation -0.4534826), and averaged windowed divergence and recombination rate (p-value 0.0122, correlation 0.09879411). However, the value and even sign of those correlations differs substantially by linkage group.

Appendix G shows these correlations against time: in other words, if $d_{XY}(A, B)_i$ is the divergence between species A and B in window i and f_i is a genomic feature of that window, then $\text{cor}(d_{XY}(A, B), f)$ is plotted against the phylogenetic time separating A and B. Stankowski et al. (2019) and Rodrigues, Kern, and Ralph (2024) found striking trends with time in such plots, but our results show no consistent trends with time.

Although correlations do not change with phylogenetic distance, they do differ substantially between linkage groups. This is quite different to what Stankowski et al. (2019) saw: correlations were similar across all *Diplacus* chromosomes. However, correlations with recombination rate and exon density also varied substantially by chromosome in the great apes (data replotted from Rodrigues, Kern, and Ralph (2024) in Supplementary Figure B.5). This emphasizes the need for more examples of this sort of data: even within cichlids, if we had restricted attention to linkage group 1 (which has a correlation near 0.4 with recombination rate) we might con-

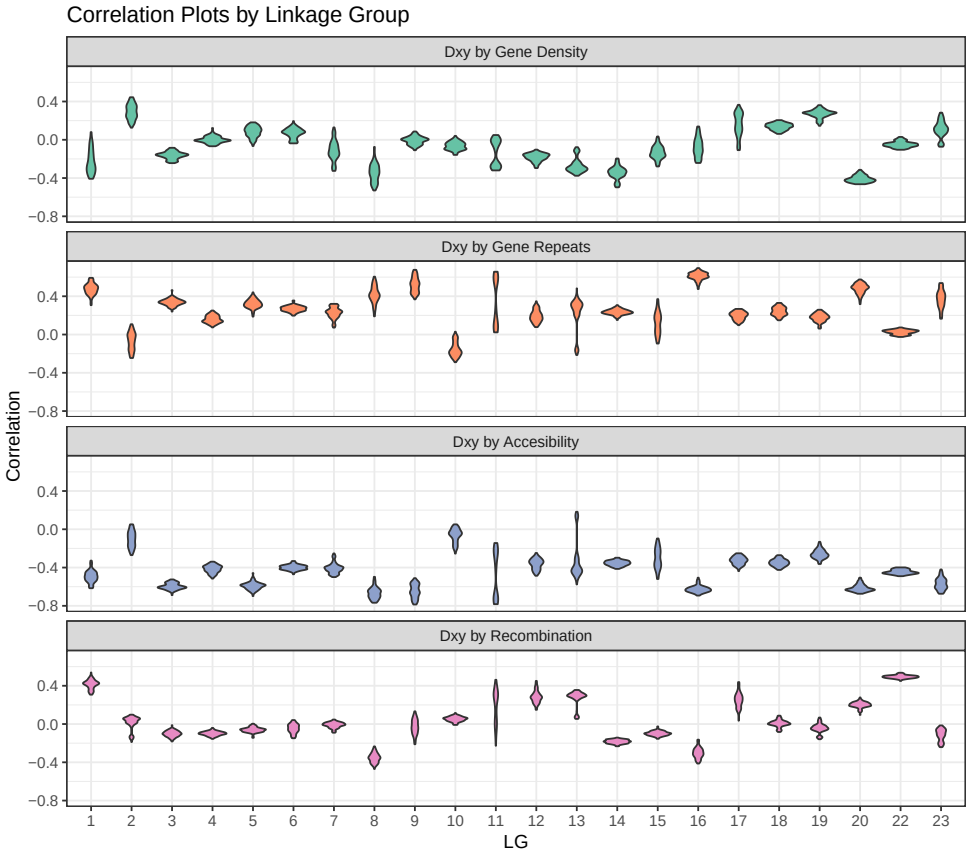

Figure 3.4. Shows the correlation between four genome functions (gene density, gene repeats, accessibility, recombination) and d_{XY} for each linkage group. In each boxplot, correlations between the genome functions and d_{XY} contains information from all species pairs. The correlation between gene density and d_{XY} range from -0.5 to 0.5, with about half of the correlations crossing through 0. The correlation between gene repeats and d_{XY} range from -0.25 to 0.75, with a majority of the correlations being positive. The correlation between accessibility and d_{XY} range from -0.75 to 0.25, with a majority of the correlations being negative. The correlation between recombination and d_{XY} range from -0.5 to 0.5, with a majority of the correlations being close to 0. There are bimodal distributions of correlation values in several linkage groups with the most noticeable occurring in linkage groups 2, 11, and 13.

clude something very different than for linkage group 8 (which has a correlation around -0.4 with recombination rate). For cichlids, it seems that large-scale features are much more stochastic than in other species.

Supplementary Figure B.4 shows correlations of divergence with genomic features, separated not by linkage group but by species comparison.

Introgression

Consistent with previous studies on the history of hybridization and introgression among Lake Malawi cichlids, the f-branch statistic showed evidence of extensive recent and historical introgression among the taxa of this radiation. The strongest signals are of recent hybridization between *speciosus* and *caeruelus*, as well as historical hybridization between *speciosus* and the common ancestor of *caeruelus*, *strigatus*, *rhoadesii*, *ornatus*, *oxyrhynchus*, *subocularis*, and *placodon* (observed in heat maps seen in appendix I).

To investigate how these histories affect landscapes of diversity and divergence, we used f_d to measure the degree of shared ancestry (i.e., the amount of historical introgression) in windows along the genome among various species pairs. Calculations of f_d performed using taxa at increasing levels of divergence should identify introgression that occurred further back in time, because f_d can only identify introgression that has occurred since the divergence of the taxa used as P_1 and P_2 for the calculation of f_d (Martin, Dasmahapatra, et al. 2013; Short and Streisfeld 2023). Roughly speaking, f_d measures amount of ancestry that is shared by two taxa, P_2 and P_3, that is not shared by another taxon, P_1; taking P_1 and P_2 to be sister taxa in the phylogeny, this indicates historical gene flow from P_3 to P_2 more recently than the divergence between P_1 and P_2. Figure 3.5 shows f_d from *speciosus* (P_3) to *caeruelus* (P_2) since the divergence of *strigatus*, *rhoadesii*, *ornatus*, *oxyrhynchus*, *subocularis*, *placodon*, or *rostratus* (P_1).

Figure 3.5A shows evidence of higher levels of historical introgression when the most genetic diverged taxon, *rostratus*, was used as P_1. We ran additional f_d calculations to identify introgression that has occurred since the divergence of *speciosus* and its sister taxon, *intermedius* (which display similar relative divergences to *rostratus* and *caeruelus*, see Supplementary Figure B.3) and also identified similarly high levels of introgression. Our identification of greater f_d values in these two trios containing *caeruelus*, suggests that hybridization has occurred at multi-

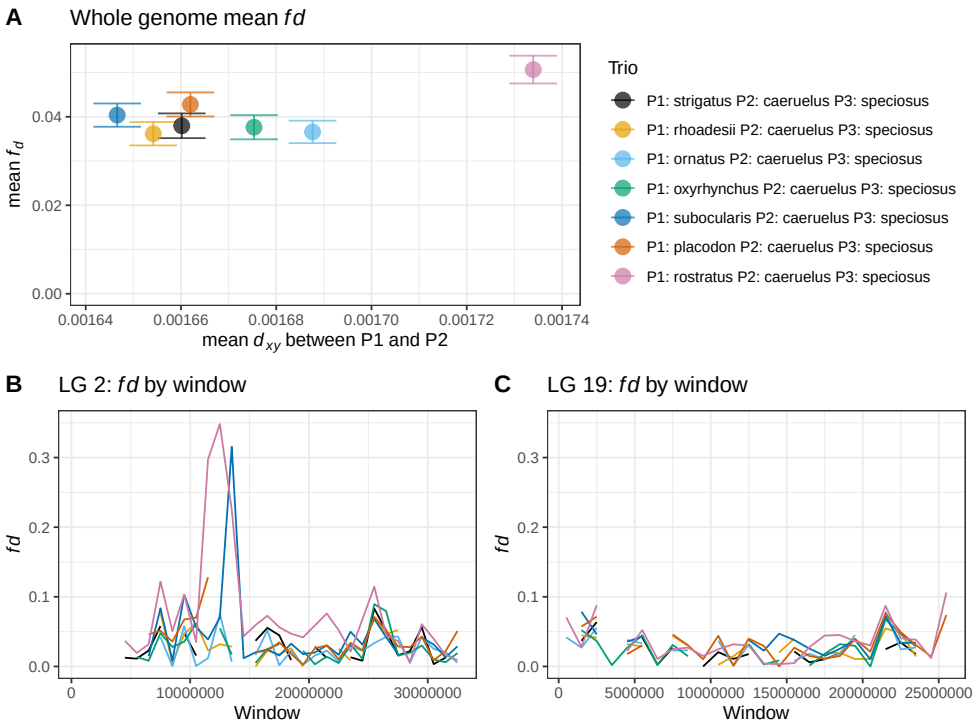

Figure 3.5. **(A)** Average f_d for 1Mb windows across the entire genome, colored by the taxa used for P_1. **(B,C)** f_d for the same sets of taxa in 1Mb windows for linkage groups 2 and 19 respectively, with f_d for each trio shown as a single line.

ple time points throughout their divergence history.

Figure 3.5B and C shows that the signal of introgression varied by linkage group. For example, f_d is larger on linkage group 2 than on linkage group 19. Furthermore, consistent with our genome-wide estimates of f_d, we identified greater values on linkage group 2 using *rostratus* as P1 ($f_d > 0.0625$) as compared to taxa that are less diverged from *caeruelus* (< 0.05). On Chromosome 19 however we identified no such pattern with the average f_d value consistently remaining between 0.02 and 0.04 regardless of which taxon was used as P1.

We identified a negative correlation between f_d and d_{XY} between *rostratus* and *caeruelus* (≈ -0.103), and a strong, negative correlation between f_d and d_{XY} between *caeruelus* and *speciosus* (≈-0.320) (Supplementary figure B.6), which is consistent with introgression occurring from *speciosus* into *caeruelus*. This suggests that introgression is largely acting to decrease divergence between *caeruelus* and *speciosus* as well as between *rostratus* and *caeruelus*. Surprisingly, we identified a

strong negative correlation between f_d and π for both *speciosus* and *caeruelus* ($\approx$-0.2), which raises the possibility that widespread adaptive introgression has contributed to the retention of introgressed alleles across the genome of these species.

3.4 Discussion

In this study, we examined genetic diversity and divergence in 1Mb windows along the genomes of 14 cichlid species. Genetic diversity was low but varied between and within linkage groups, ranging from around 0.05% to 0.3%, and divergence ranged from 0.1% to 0.4%, similar to values seen in Malinsky, Svardal, et al. (2018) and Svardal, Salzburger, and Malinsky (2021). Both within-species diversity and between-species divergence had very similar valleys and peaks across most linkage groups, although a striking pattern of three distinct "pathways" is visible on LG11, likely due to a polymorphic inversion reported in Conte et al. (2019).

Although correlations between landscapes of diversity or divergence were high – mostly above 0.7 – correlations decreased with time, i.e., were lower between more phylogenetically distant comparisons. The patterns and levels of correlation differed by linkage group, as for example on LG11. This decrease of correlation with time has also been observed in other groups, e.g., great apes (Rodrigues, Kern, and Ralph 2024) and bush monkeyflowers (Stankowski et al. 2019).

To understand the balance of forces that maintain these correlations between diversity and divergence it can help to examine the relationship to other genomic features. Across the cichlid genomes we examined, genetic diversity and divergence were correlated with recombination rate, repeat content, and accessibility, although the strength of the correlations differed between linkage groups. Furthermore, the correlation of any of these features with divergence did not strongly increase or decrease with phylogenetic distance between the taxa, unlike in monkeyflowers (Stankowski et al. 2019) or great apes (Rodrigues, Kern, and Ralph 2024).

We observed that levels of introgression differed between trios and varied between linkage groups. The patterns that are present in this dataset are likely caused by complex effects of inversions, introgression, sexual selection, rearrangements and linked selection discussed in further detail below.

Features of genomic landscapes

As reviewed above, one of the most commonly observed features of genomic landscapes is a positive correlation between genetic diversity and recombination rate (Corbett-Detig, Hartl, and Sackton 2015). This is taken as strong evidence for the effects of linked selection, which also predicts a negative correlation of genetic diversity and gene density. Indeed, genetic diversity is positively correlated with recombination rate and negatively correlated with gene density genome-wide both in our study as well as in Burri et al. (2015), Stankowski et al. (2019), Rodrigues, Kern, and Ralph (2024), and Wang et al. (2020). The same is true of genetic divergence as well in most of these studies. However, the correlations we see differ substantially by linkage group: as shown in Figure 3.4, these correlations when looked at on different linkage groups can be of opposite sign, and sometimes quite large. It is not usually reported whether these correlations differ so much by chromosome in other taxa: at least in great apes, the data from Rodrigues, Kern, and Ralph (2024) shows substantial variation between chromosomes, but a consistent sign. This fact, and the fact that these correlations tend to be relatively weak (usually less than 0.4) suggests that the intensity of linked selection is likely changing a lot over time (e.g., as species radiate and genomes rearrange).

Chromosomal rearrangements

A number of large chromosomal rearrangements such as inversions and centromere repositioning events have been reported in cichlid species, as well as smaller-scale gene duplications (Conte et al. 2019; Svardal, Salzburger, and Malinsky 2021). The prevalence of these large rearrangements seems to be higher than in the other taxa we compare to: for instance, there is only a single chromosomal fusion event separating humans from other great apes and few other rearrangements, and chromosomal structure of *Fidecula* flycatchers is very well-conserved (Burri et al. 2015). The degree of chromosomal synteny in *Diplacus* (Stankowski et al. 2019) and *Populus* (Wang et al. 2020) is less clear, but at least fewer instances of inversions or other large rearrangements have been reported. Speculatively, rearrangements might be having a stronger effect on cichlids than on these other taxa.

Perhaps the most obvious signal produced by a rearrangement would be that of a relatively old inversion: if the inversion is older than typical coalescence time,

the two orientations of the inversion generally manifest as two distinct haplotypes, so that genetic variation tends to nest within the two distinct groups defined by the orientations. This agrees with what we see on linkage group 11. However, a more recent inversion would look different: then, the inverted haplotype would be nested within genetic variation on non-inverted haplotypes, and so have less obvious effects. Furthermore, smaller inversions produce smaller signals. It is interesting to see how a large inversion creates discontinuous patterns in the relationship between time and correlation (e.g., Figure 3.3); the net effects of many smaller events would create a gradual decrease. Since the putative inversion on linkage group 11 seems to be relatively old and show a pattern that does not agree with the genome-wide phylogeny, this may suggest that it pre-dated the radiation (as suggested by Svardal, Salzburger, and Malinsky (2021)). Many other rearrangements are known to be present in cichlids, but we may not expect to see their effects if they are not polymorphic in the taxa we study.

Sex determining regions

Cichlids are known to have an array of different sex determining loci and systems, with a high rate of turnover (Santos, Lopes, and Kratochwil 2023). In Lake Malawi cichlids alone, five sex determining systems have been noted and described (Feller et al. 2021). Both ZW and XY sex determining systems have been found within the same species of *Metriaclima* (Ser, Roberts, and Kocher 2010). Such rapid change in sex determination could create very different patterns of gene flow and genetic diversity across chromosomes and across time. Incompatibilities between different sex determination systems can be coupled with a reduction in or complete suppression of recombination (Conte et al. 2019). Changing the sex determination system can also skew the sex ratio (Feller et al. 2021), which impacts population dynamics. Selection pressures acting on sex-linked traits may drive adaptation and divergence, leading to the emergence of novel phenotypes (e.g., cichlid color patterns) and speciation events (Wagner, Harmon, and Seehausen 2012; Conte et al. 2019). Changes in sex determination across different cichlid species might have contributed to the rapid radiation of cichlids (Feller et al. 2021) and could be a major reason that we see such strikingly different patterns in genomic landscapes and their rates of change across linkage groups.

Hybridization and introgression

It is well known that cichlids can hybridize with one another and that this tendency to hybridize is thought to be partially responsible for the success of several cichlid radiations (Meier et al. 2017; Malinsky, Svardal, et al. 2018; Svardal, Salzburger, and Malinsky 2021). Repeated hybridization between both near and distant lineages has led to introgression in between cichlid lineages, leading to complex and difficult-to-infer historical relationships (Santos, Lopes, and Kratochwil 2023). Cichlid species with a history of introgression tend to have higher genetic diversity in some species groups, but most genetic variation present in the ancestral Lake Malawi cichlids has been lost (Svardal, Salzburger, and Malinsky 2021). We found signals of introgression between several species of cichlid, that differed between linkage groups (Figure 3.5). In particular, there is one large several megabase signal on linkage group 2 (visible in two species comparisons), highlighting that introgression can be relatively even or effect large regions. Interestingly, introgression is usually expected to increase genetic diversity, but we observed less genetic diversity in regions of the genome with more evidence of introgression (Supplementary Figure B.6). This observation could be due to positive selection on introgressed alleles, or potentially other factors. Finally, the concept of introgression depends on the existence of a species tree, and so a different framework might better describe the recent dynamics of these species.

Limitations and continuing questions

We found that genomic landscapes of diversity and divergence were more similar between species that were more closely related according to the genome-wide phylogeny of Figure B.3. However, there is not a single phylogenetic relationship that encompasses the entire genome: different portions of the genome have sometimes very different phylogenies. By using the phylogeny estimated from the whole genome, we hope to get a reasonable description of relationships: e.g., which taxa tend to be more closely related on most of the genome. In particular, the choice of outgroup is well-supported across the genome, since it has noticeably higher divergence to all other species on most linkage groups (see appendix C). A somewhat different set of species or a different phylogenetic inference method would lead to a different phylogeny, but we do not expect the main observations to be affected. It

could be better to analyze the data without reference to a phylogeny, but it seems difficult to describe relationships between species otherwise.

There are also certainly some degree of errors generated by aligning to a single reference genome. Some of the potential problems caused through the alignment process can create false regions of diversity or divergence, e.g., a duplication that is not present in the reference genome, or unmasked repetitive regions. Perhaps more concerningly, hyper-divergent portions of the genome might not map well, thus reducing divergence in the region. Furthermore, we simply ignore any non-syntenic regions of the genome: it would be interesting to see how the amount of such regions increases with time. In particular, it would be interesting to survey changes in repetitive element content or location, along the lines of Quah et al. (2024) or Stitzer et al. (2021).

Cichlids have attracted a lot of attention from evolutionary biologists due to the phenotypic diversity generated by their recent adaptive radiation. The cichlid genome reflects this complex evolutionary history, with striking patterns that appear at chromosomal scale along the genome. Patterns that we see are different from what is observed in other species, so it is interesting to speculate that these differences are related to the history and ecology of cichlids. For instance: is their propensity for speciation reflected in patterns of genetic variation over time? Would we see similar patterns in other taxa that have recently radiated (e.g., tomatoes, pupfish, etc)? Further research integrating genomic data with ecological and geological information will help elucidate the links between cichlid evolution, genetic variation, and speciation dynamics. Simulations could also be used to determine how these different evolutionary forces interplay with one another to leave various patterns in genetic data. Cichlids present a challenging case study about the relationship between evolutionary dynamics and genetic variation. This study thus provides the field with an additional set of observations for a species with a complex evolutionary history.

3.5 References

Barton, Nicholas H and Katherine S Gale, *Genetic analysis of hybrid zones*, Hybrid zones and the evolutionary process, Oxford University Press, 13–45.

Battey, Christopher J., *Evidence of linked selection on the Z chromosome of hybridizing hummingbirds** (en), Evolution **74**, no. 4, 725–739, DOI: 10.1111/evo.13888.

Bierne, N et al., *Deleterious mutations in a hybrid zone: can mutational load decrease the barrier to gene flow?*, Genet Res **80**, no. 3, 197–204.

Birky, C W and J B Walsh, *Effects of linkage on rates of molecular evolution.* (en), Proceedings of the National Academy of Sciences **85**, no. 17, 6414–6418, DOI: 10.1073/pnas.85.17.6414.

Brawand, David et al., *The genomic substrate for adaptive radiation in African cichlid fish* (en), Nature **513**, no. 7518, 375–381, DOI: 10.1038/nature13726.

Burri, R., *Dissecting differentiation landscapes: a linked selection's perspective*, Journal of Evolutionary Biology **30**, no. 8, 1501–1505, DOI: 10.1111/jeb.13108, eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jeb.13108.

Burri, Reto et al., *Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers* (en), Genome Research **25**, no. 11, 1656–1665, DOI: 10.1101/gr.196485.115.

Camacho, Christiam et al., *BLAST+: architecture and applications* (en), BMC Bioinformatics **10**, no. 1, 421, DOI: 10.1186/1471-2105-10-421.

Charlesworth, B, M T Morgan, and D Charlesworth, *The effect of deleterious mutations on neutral molecular variation.* (en), Genetics **134**, no. 4, 1289–1303, DOI: 10.1093/genetics/134.4.1289.

Conte, Matthew A et al., *Chromosome-scale assemblies reveal the structural evolution of African cichlid genomes* (en), GigaScience **8**, no. 4, DOI: 10.1093/gigascience/giz030.

Corbett-Detig, Russell B., Daniel L. Hartl, and Timothy B. Sackton, *Natural Selection Constrains Neutral Diversity across A Wide Range of Species* (en), PLOS Biology **13**, no. 4, e1002112, DOI: 10.1371/journal.pbio.1002112.

Cruickshank, Tami E. and Matthew W. Hahn, *Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow* (en), Molecular Ecology **23**, no. 13, 3133–3157, DOI: 10.1111/mec.12796.

Danecek, Petr et al., *Twelve years of SAMtools and BCFtools* (en), GigaScience **10**, no. 2, giab008, DOI: 10.1093/gigascience/giab008.

Duret, Laurent and Nicolas Galtier, *Biased Gene Conversion and the Evolution of Mammalian Genomic Landscapes*, Annual Review of Genomics and Human Genetics **10**, no. 1, 285–311, DOI: 10.1146/annurev-genom-082908-150001, eprint: http://www.annualreviews.org/doi/pdf/10.1146/annurev-genom-082908-150001, PMID: 19630562.

Feller, Anna F. et al., *Identification of a novel sex determining chromosome in cichlid fishes that acts as XY or ZW in different lineages* (en), Hydrobiologia **848**, no. 16, 3727–3745, DOI: 10.1007/s10750-021-04560-7.

Green, Richard E et al., *A draft sequence of the Neandertal genome*, science **328**, no. 5979, 710–722.

Guillerme, Thomas, *dispRity: A modular R package for measuring disparity* (en), Methods in Ecology and Evolution **9**, no. 7, 1755–1763, DOI: 10.1111/2041-210X.13022.

Harris, Kelley and Rasmus Nielsen, *The Genetic Cost of Neanderthal Introgression*, Genetics **203**, no. 2, 881–891.

Harrison, R G and E L Larson, *Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones*, Mol Ecol **25**, no. 11, 2454–2466, DOI: 10.1111/mec.13582.

Hudson, R. R. and N. L. Kaplan, *Deleterious background selection with recombination*, Genetics **141**, no. 4, 1605–1617.

Irwin, Darren E. et al., *Recurrent selection explains parallel evolution of genomic regions of high relative but low absolute differentiation in a ring species* (en), Molecular Ecology **25**, no. 18, 4488–4507, DOI: 10.1111/mec.13792.

Kaplan, N L, R R Hudson, and C H Langley, *The "hitchhiking effect" revisited.* (en), Genetics **123**, no. 4, 887–899, DOI: 10.1093/genetics/123.4.887.

Kimura, M, *Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution*, Nature **267**, no. 5608, 275–276, DOI: 10.1038/267275a0.

Kirkpatrick, M, *How and why chromosome inversions evolve*, PLoS Biol **8**, no. 9, DOI: 10.1371/journal.pbio.1000501.

Kirkpatrick, Mark and Nick Barton, *Chromosome Inversions, Local Adaptation and Speciation*, Genetics **173**, no. 1, 419–434, DOI: 10.1534/genetics.105.047985, eprint: https://academic.oup.com/genetics/article-pdf/173/1/419/42222075/genetics0419.pdf.

Kocher, Thomas D., *Adaptive evolution and explosive speciation: the cichlid fish model* (en), Nature Reviews Genetics **5**, no. 4, 288–298, DOI: 10.1038/nrg1316.

Kryazhimskiy, Sergey and Joshua B Plotkin, *The population genetics of dN/dS*, PLoS genetics **4**, no. 12, e1000304–e1000304, DOI: 10.1371/journal.pgen.1000304.

Langley, C H et al., *Genomic variation in natural populations of Drosophila melanogaster*, Genetics **192**, no. 2, 533–598, DOI: 10.1534/genetics.112.142018.

Lewontin, R.C., *The Genetic Basis of Evolutionary Change*, Biological Series, Columbia University Press.

Li, Heng and Richard Durbin, *Fast and accurate short read alignment with Burrows–Wheeler transform* (en), Bioinformatics **25**, no. 14, 1754–1760, DOI: 10.1093/bioinformatics/btp324.

Malinsky, Milan, Michael Matschiner, and Hannes Svardal, *Dsuite-Fast D-statistics and related admixture evidence from VCF files*, Molecular ecology resources **21**, no. 2, 584–595.

Malinsky, Milan, Hannes Svardal, et al., *Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow* (en), Nature Ecology & Evolution **2**, no. 12, 1940–1955, DOI: 10.1038/s41559-018-0717-x.

Martin, Simon H, Kanchon K Dasmahapatra, et al., *Genome-wide evidence for speciation with gene flow in Heliconius butterflies*, Genome research **23**, no. 11, 1817–1828.

Martin, Simon H, John W Davey, and Chris D Jiggins, *Evaluating the use of ABBA–BABA statistics to locate introgressed loci*, Molecular biology and evolution **32**, no. 1, 244–257.

McDonald, John H. and Martin Kreitman, *Adaptive protein evolution at the Adh locus in Drosophila* (en), Nature **351**, no. 6328, 652–654, DOI: 10.1038/351652a0.

McKenna, Aaron et al., *The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data* (en), Genome Research **20**, no. 9, 1297–1303, DOI: 10.1101/gr.107524.110.

Meier, Joana I. et al., *Ancient hybridization fuels rapid cichlid fish adaptive radiations* (en), Nature Communications **8**, no. 1, 14363, DOI: 10.1038/ncomms14363.

Miles, Alistair et al., *cggh/scikit-allel: v1.3.8*, DOI: https://zenodo.org/records/10876220.

Nguyen, Lam-Tung et al., *IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies* (en), Molecular Biology and Evolution **32**, no. 1, 268–274, DOI: 10.1093/molbev/msu300.

Paradis, Emmanuel and Klaus Schliep, *ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R* (en), Bioinformatics **35**, no. 3, 526–528, DOI: 10.1093/bioinformatics/bty633.

Quah, Fu Xiang et al., *A pangenomic perspective of the Lake Malawi cichlid radiation reveals extensive structural variation driven by transposable elements* (en), DOI: 10.1101/2024.03.28.587230.

Quinlan, Aaron R. and Ira M. Hall, *BEDTools: a flexible suite of utilities for comparing genomic features* (en), Bioinformatics **26**, no. 6, 841–842, DOI: 10.1093/bioinformatics/btq033.

R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Renaut, S. et al., *Genomic islands of divergence are not affected by geography of speciation in sunflowers* (en), Nature Communications **4**, no. 1, 1827, DOI: 10.1038/ncomms2833.

Revell, Liam J., *phytools 2.0: an updated R ecosystem for phylogenetic comparative methods (and other things)* (en), PeerJ **12**, e16505, DOI: 10.7717/peerj.16505.

Rodrigues, Murillo F, Andrew D Kern, and Peter L Ralph, *Shared evolutionary processes shape landscapes of genomic variation in the great apes* (en), Genetics **226(4)**, DOI: 10.1093/genetics/iyae006.

Santos, M. Emília, João F. Lopes, and Claudius F. Kratochwil, *East African cichlid fishes* (en), EvoDevo **14**, no. 1, 1, DOI: 10.1186/s13227-022-00205-5.

Ser, Jennifer R., Reade B. Roberts, and Thomas D. Kocher, *MULTIPLE INTERACTING LOCI CONTROL SEX DETERMINATION IN LAKE MALAWI CICHLID FISH* (en), Evolution **64**, no. 2, 486–501, DOI: 10.1111/j.1558-5646.2009.00871.x.

Short, Aidan W and Matthew A Streisfeld, *Ancient hybridization leads to the repeated evolution of red flowers across a monkeyflower radiation*, Evolution Letters **7**, no. 5, 293–304.

Sobel, James M. and Matthew A. Streisfeld, *Strong premating reproductive isolation drives incipient speciation in Mimulus aurantiacus: REPRODUCTIVE ISOLATION IN M. AURANTIACUS* (en), Evolution **69**, no. 2, 447–461, DOI: 10.1111/evo.12589.

Stankowski, Sean et al., *Widespread selection and gene flow shape the genomic landscape during a radiation of monkeyflowers* (en), PLOS Biology **17**, no. 7, e3000391, DOI: 10.1371/journal.pbio.3000391.

Stitzer, Michelle C. et al., *The genomic ecosystem of transposable elements in maize*, PLOS Genetics **17**, no. 10, 1–30, DOI: 10.1371/journal.pgen.1009768.

Svardal, Hannes, Walter Salzburger, and Milan Malinsky, *Genetic Variation and Hybridization in Evolutionary Radiations of Cichlid Fishes* (en), Annual Review of Animal Biosciences **9**, no. 1, 55–79, DOI: 10.1146/annurev-animal-061220-023129.

Van Doren, Benjamin M. et al., *Correlated patterns of genetic diversity and differentiation across an avian family* (en), Molecular Ecology **26**, no. 15, 3982–3997, DOI: 10.1111/mec.14083.

Wagner, Catherine E., Luke J. Harmon, and Ole Seehausen, *Ecological opportunity and sexual selection together predict adaptive radiation* (en), Nature **487**, no. 7407, 366–369, DOI: 10.1038/nature11144.

Wang, Jing et al., *Evidence for widespread selection in shaping the genomic landscape during speciation of Populus* (en), Molecular Ecology **29**, no. 6, 1120–1136, DOI: 10.1111/mec.15388.

Wickham, Hadley, *ggplot2: Elegant Graphics for Data Analysis* (en), Springer New York, New York, NY, DOI: 10.1007/978-0-387-98141-3.

———, *Reshaping Data with the **reshape** Package* (en), Journal of Statistical Software **21**, no. 12, DOI: 10.18637/jss.v021.i12.

Wickham, Hadley, Davis Vaughan, and Maximilian Girlich, *tidyr: Tidy Messy Data*, R package version 1.3.0.

Xu, Shuangbin et al., *Ggtree : A serialized data object for visualization of a phylogenetic tree and annotation data* (en), iMeta **1**, no. 4, e56, DOI: 10.1002/imt2.56.

York, Ryan A. et al., *Behavior-dependent cis regulation reveals genes and pathways associated with bower building in cichlid fishes* (en), Proceedings of the National Academy of Sciences **115**, no. 47, DOI: 10.1073/pnas.1810140115.

CHAPTER 4

CONCLUSION

This dissertation describes my work investigating the many different signals produced by the process of evolution. The evolutionary mechanisms of natural selection, adaptation, mutation, gene flow, genetic drift, and non-random mating all interact in various ways over multiple timescales. These mechanisms have left both ecological and molecular patterns in all life on Earth. The goal of this dissertation was to observe and examine these pattens to expand our current knowledge on how the processes of evolution impact populations.

In my second chapter, I investigated the evolutionary trajectories of a coevolutionary arms race between a predator (snakes) and its' prey (newts). I used spatial simulations to test the effects of genetic basis and spatial heterogeneity on trait evolution. I found that newts and snakes could coevolve under most genetic architectures, except when the mutational variance of a specific architecture was too small. Mutational variance was an important factor that greatly impacted coevolutionary trajectories. These results agree with previous research confirming how a lack of genetic variation created by low mutation variance can effectively halt evolution is some coevolutionary interactions. As I explored the geographic variation of newts' toxicity and snakes' resistance across space and time, I found that my simulations could only produce coevolutionary mosaics if there was spatial heterogeneity existing within the simulation regardless of the genetic architecture. The result from these simulations highlights the importance of including spatial aspects into the study of evolution, especially when dealing with larger geographical areas. Findings from this chapter impacts the field of coevolution theory by improving our understanding of how the genetic basis of traits can impact coevolutionary trajectories and by showing the importance of including geographical space when studying evolution.

In my third chapter, I switch from exploring evolutionary processes though simulation to observing the signals of past evolution by examining the landscapes of cichlid genomes. The many different mechanisms of evolution have left patterns that can be interpreted. In cichlids I found that there were low levels of genetic diversity and divergence across the genome. Even though the levels of genetic diversity and divergence were low, I observed that they were strongly correlated. I found that as species became more phylogenetically different these correlations weakened.

There were also strong correlations between levels of genetic divergence and genome functions. The most interesting finding from this study was the striking differences between linkage groups. Some of the locations of these striking patterns align with previous research and point to these signals coming from a combination of inversions and sexual selection. Although these signals provide clues about the history of evolutionary forces acting on these genomes, this history is sufficiently complex that we cannot yet see the entire picture from these data. The findings from my second chapter fit well within our current knowledge of genetic landscapes especially when compared to the genomic landscapes of great apes and monkeyflowers. The research that I did for this chapter adds additional knowledge point into this growing field.

My dissertation encompasses a small, but specific portion of the field of evolution. The findings in my dissertation contributed to our understanding of evolutionary signals of natural selection and its' impacts along the genome, by emphasizing the intricate mechanisms underlying genetic variation and adaptation, revealing how natural selection acts on specific genomic regions to shape phenotypic diversity and ecological interactions. I hope the findings in my research inspire further investigation and innovation in the field of evolutionary biology, fostering a deeper understanding of the mechanisms driving biological diversity and adaptation. The threads left behind by evolution are complex and messy, but they are ready for researchers to grasp.

APPENDIX A

COEVOLUTION APPENDIX

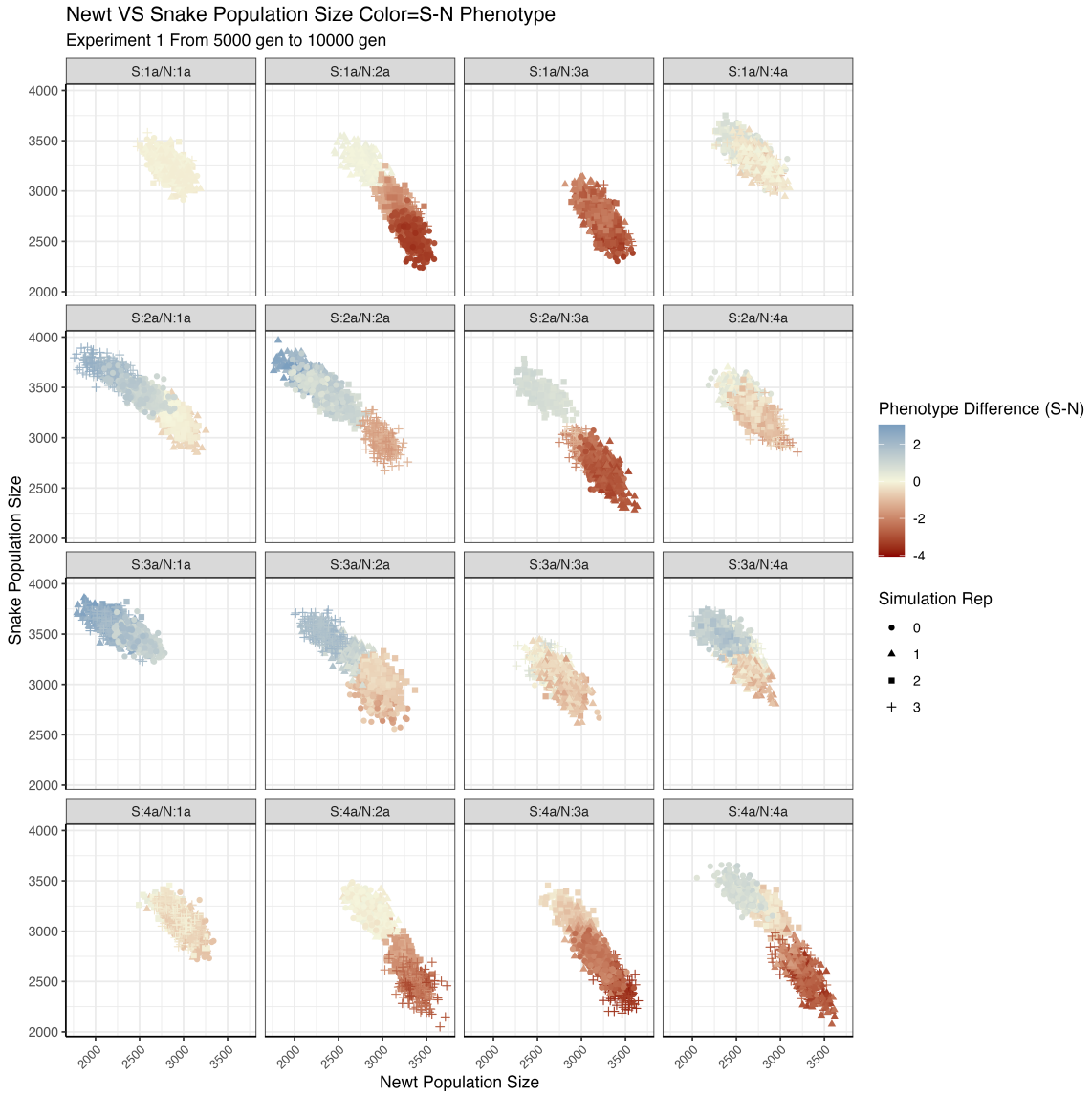

Figure A.1. Newt and snake population sizes for the sixteen genetic architecture combinations of Experiment 1 over generations 5000 to 10000. Points are colored by the difference between average snake and average newt phenotypes. In each plot the population sizes at a evenly spaced set of time points are shown for each of the four replicates (replicate ID shown by point type). Newt and snake population sizes are generally negatively correlated: when newts have higher phenotypes (red points), newt populations tend to be larger and snake populations smaller. Note there is substantial variation between replicates, probably due to differing initial conditions.

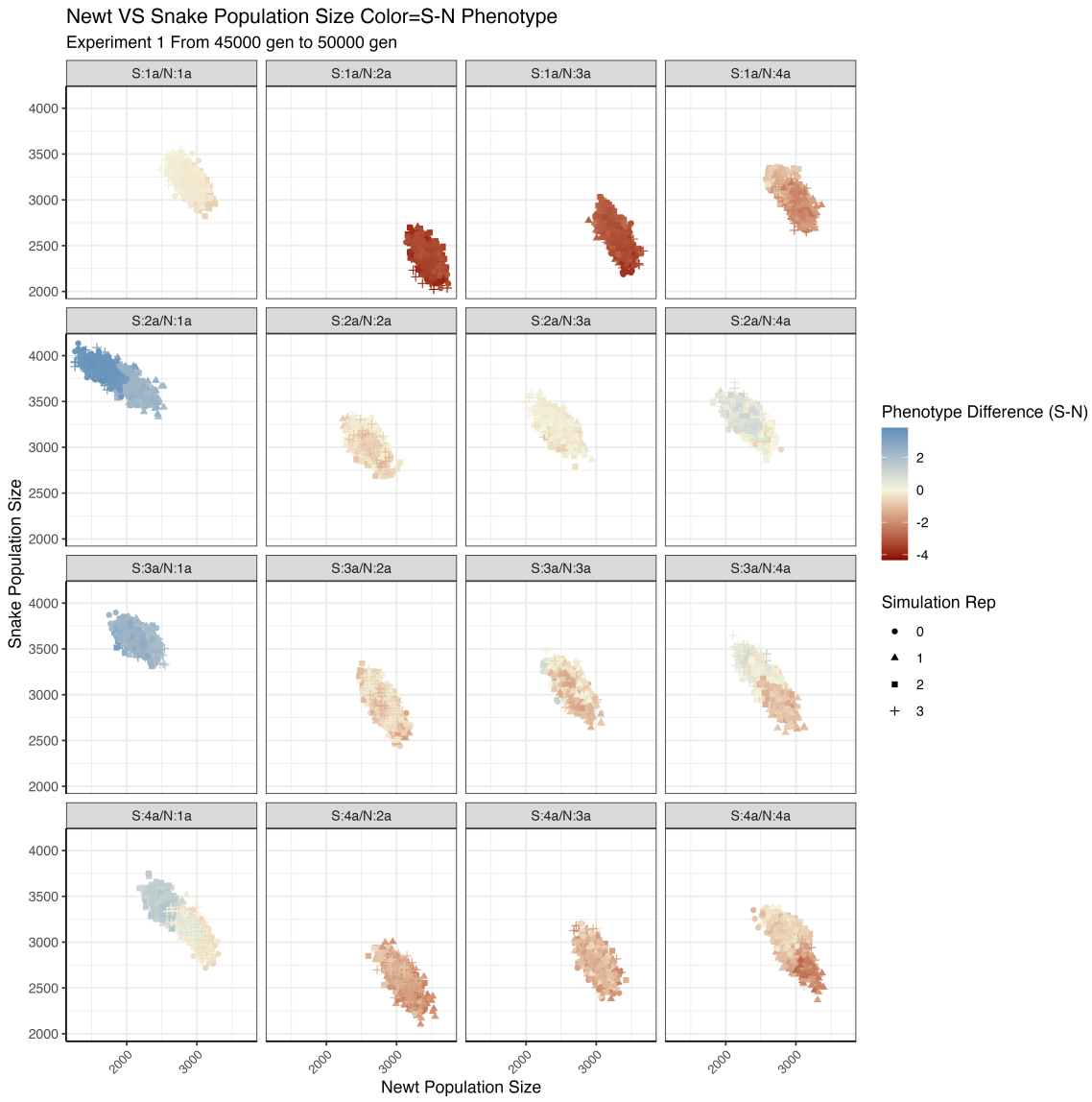

Figure A.2. As in Figure A.1, except that time points are between 45,000 and 50,000 generations (the end of the simulations). Note that there is less variation between replicates than in the earlier time points.

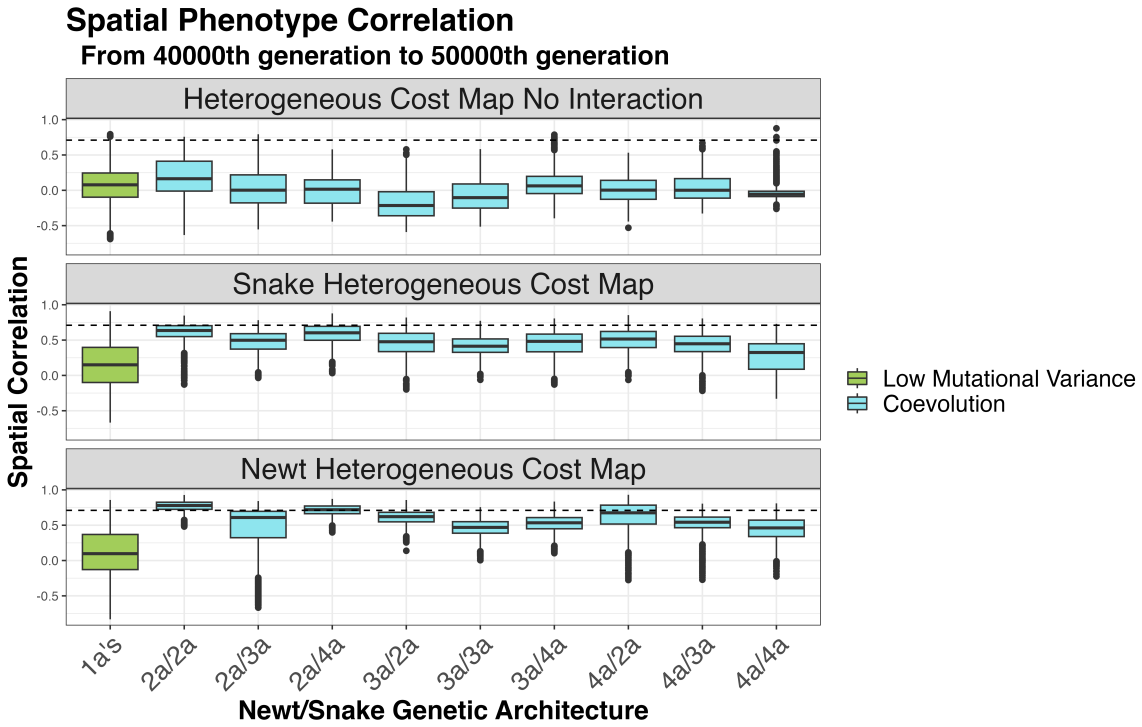

Figure A.3. Distributions of spatial correlations between newt and snake phenotypes, across the combinations of genetic architectures of Experiment 1. Each boxplot shows the range of spatial correlations computed across an evenly spaced set of time points from 40,000 to 50,000 generations. **(top)** "Heterogeneous cost - no interaction" refers to simulations where costliness for both species varies across the landscape, but interaction outcome does not depend on phenotype (i.e., is just the result of a coin toss); **(middle)** "Heterogeneous cost - snake" refers to simulations that are as usual except that the costliness of the snake phenotype varies across the landscape (but not newt phenotypes); **(bottom)** "Heterogeneous cost - newt" refers to the converse situation, where newt phenotype costliness varies, but not snake.

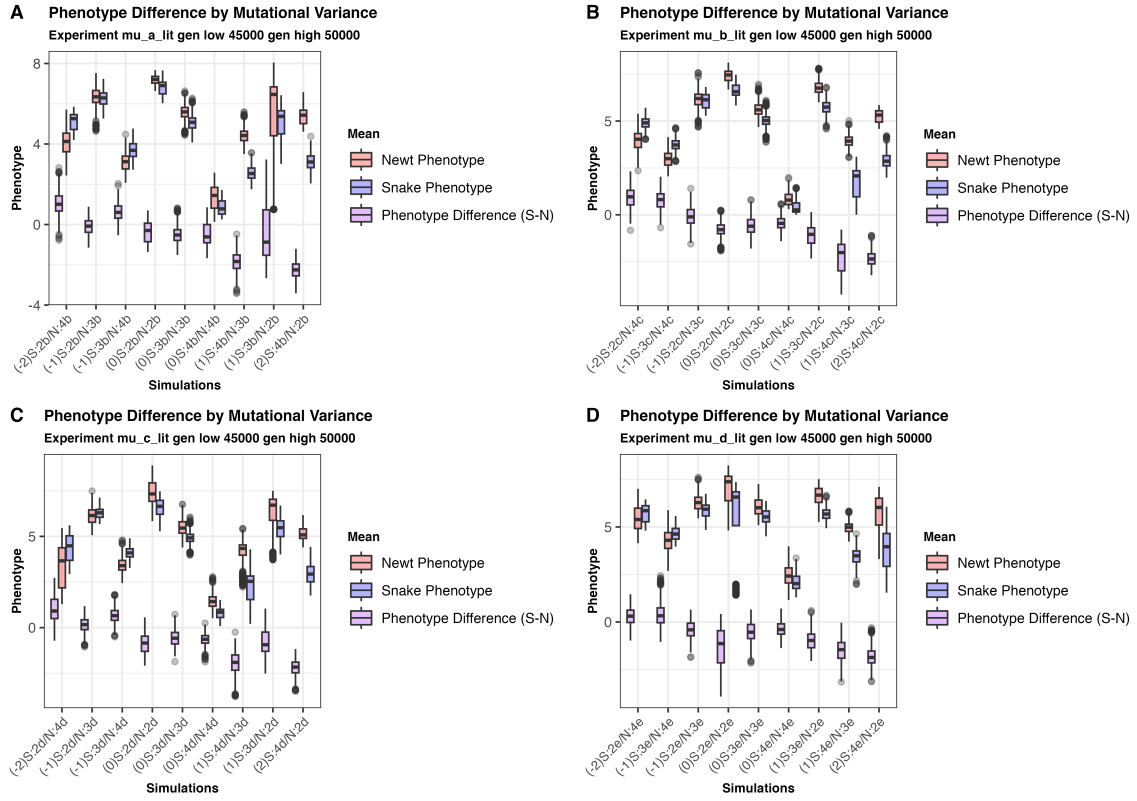

Figure A.4. Distribution of mean newt and snake phenotypes and mean (snake minus newt) phenotype differences from a regularly spaced set of generations between 45,000 and 50,000. in **Experiment 2**, except for those simulations including a genetic architecture with the lowest mutational variance (1b, 1c, 1d, or 1e). (A) shows all nine combinations of mutational variances in the first shaded section of Experiment 2 of Table 2.1 (i.e., genetic architectures 2b, 3b, and 4b). The combination is shown on the x-axis labels, prepended with the difference of snake and newt $\log_{10}(V_M)$ values: for instance, the leftmost set of boxplots, labeled "(-2)S:2b/N:4b", refers to simulations in which snakes have genetic architecture 2b, newts have genetic architecture 4b; and the snake's genetic architecture has 100 times less mutational variance than does the newt's. (B, C, and D) show the same, but for combinations in the remaining three sections of Experiment 2 in Table 2.1. The arrangement is so that for boxplots on the left, snakes have lower mutational variance than newts, and on the right, newts have lower mutational variance than snakes.

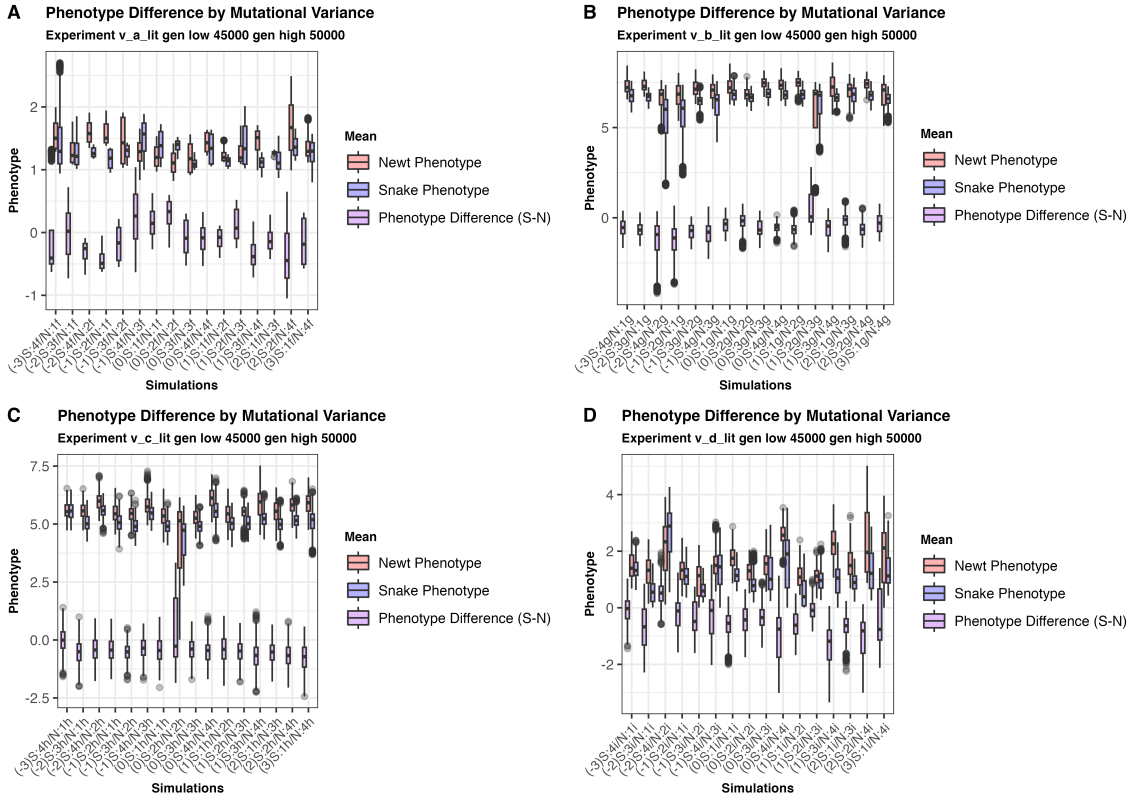

Figure A.5. Mean newt and snake phenotypes (and mean differences), much as in Figure A.4, but for **Experiment 3**. In this experiment, the different genetic architectures are grouped in Table 2.1 by mutation rate, so the boxplots are ordered by difference in $\log_{10}(\mu)$, so that on the left, snakes have lower mutation rate than newts, and on the right, newts have lower mutation rate than snakes.

APPENDIX B

CICHLID APPENDIX

Table B.1. Samples used in this study (from Malinsky, Svardal, et al. (2018))

Sample	MeanCov	Accession num	Accession num2	Accession num3
Champsochromis caeruelus	14.1176	ERR715493		
Champsochromis caeruelus	14.0589	ERR715494		
Chilotilapia rhoadesii	13.2928	ERR715491		
Chilotilapia rhoadesii	4.96914	ERR702296		
Chilotilapia rhoadesii	4.65948	ERR702297		
Chilotilapia rhoadesii	5.23728	ERR702308		
Copadichromis virginalis	13.4079	ERR299213	ERR299204	ERR299204
Copadichromis virginalis	5.83318	ERR702292		
Copadichromis virginalis	5.54251	ERR702293		
Copadichromis virginalis	5.84242	ERR702294		
Copadichromis virginalis	5.45189	ERR702295		
Copadichromis virginalis	4.3907	ERR715482		
Copadichromis virginalis	4.40469	ERR715483		
Ctenopharynx intermedius	4.69935	ERR702298		
Ctenopharynx intermedius	13.1535	ERR715492		
Ctenopharynx intermedius	4.64397	ERR702299		
Dimidiochromis strigatus	13.1328	ERR1081380		
Dimidiochromis strigatus	12.9357	ERR715523		

Continues on the next page ...

Table B.1. The table *(continued from previous page)*

Sample	MeanCov	Accession num	Accession num2	Accession num3
Protomelas ornatus	14.4113	ERR715496		
Protomelas ornatus	13.7368	ERR715497		
Fossorochromis rostratus	4.54727	ERR715476		
Fossorochromis rostratus	5.20952	ERR702309		
Fossorochromis rostratus	12.5285	ERR715512		
Fossorochromis rostratus	13.0272	ERR715516		
Hemitilapia oxyrhynchus	12.9473	ERR1081376		
Hemitilapia oxyrhynchus	12.0616	ERR715518		
Lethrinops lethrinus	11.5057	ERR271663	ERR271655	ERR271655
Lethrinops lethrinus	12.6041	ERR315230	ERR303405	ERR303405
Mylochromis anaphyrmus	12.5896	ERR266461	ERR266493	ERR266493
Mylochromis anaphyrmus	5.41922	ERR702288		
Mylochromis anaphyrmus	5.42683	ERR702289		
Mylochromis anaphyrmus	5.46714	ERR702290		
Mylochromis anaphyrmus	5.24887	ERR702291		
Otopharynx speciosus	13.8435	ERR715488		
Otopharynx speciosus	12.8562	ERR715489		
Placidochromis subocularis	4.92934	ERR702304		

Continues on the next page ...

Table B.1. The table *(continued from previous page)*

Sample	MeanCov	Accession num	Accession num2	Accession num3
Placidochromis subocularis	4.42822	ERR715487		
Placidochromis subocularis	4.66737	ERR715486		
Placidochromis subocularis	4.34062	ERR715485		
Placidochromis subocularis	4.34459	ERR715484		
Placidochromis subocularis	5.1356	ERR702307		
Placidochromis subocularis	5.12527	ERR702306		
Placidochromis cf longimanus	12.6469	ERR715524		
Placidochromis cf longimanus	4.63765	ERR715478		
Placidochromis cf longimanus	4.54243	ERR715479		
Placidochromis cf longimanus	4.50002	ERR715480		
Placidochromis cf longimanus	4.49001	ERR715481		
Protomelas ornatus	14.0089	ERR715514		
Protomelas ornatus	13.0325	ERR1081372		
Tremitochranus placodon	4.91619	ERR702300		
Tremitochranus placodon	4.97781	ERR702301		
Tremitochranus placodon	4.82966	ERR702302		
Tremitochranus placodon	5.18536	ERR702303		

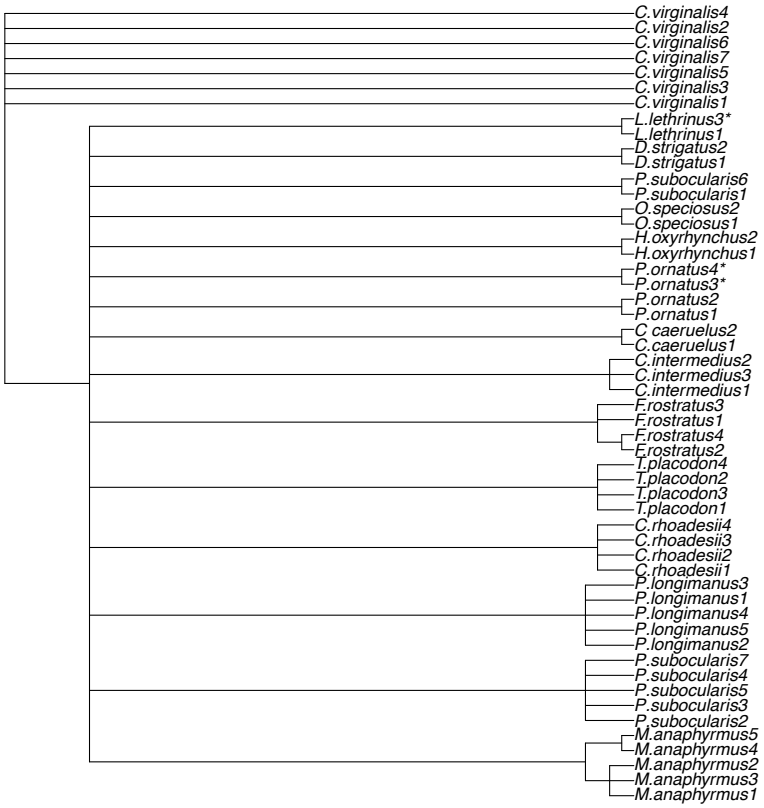

Figure B.1. Phylogenetic tree of all individuals (see text for details).

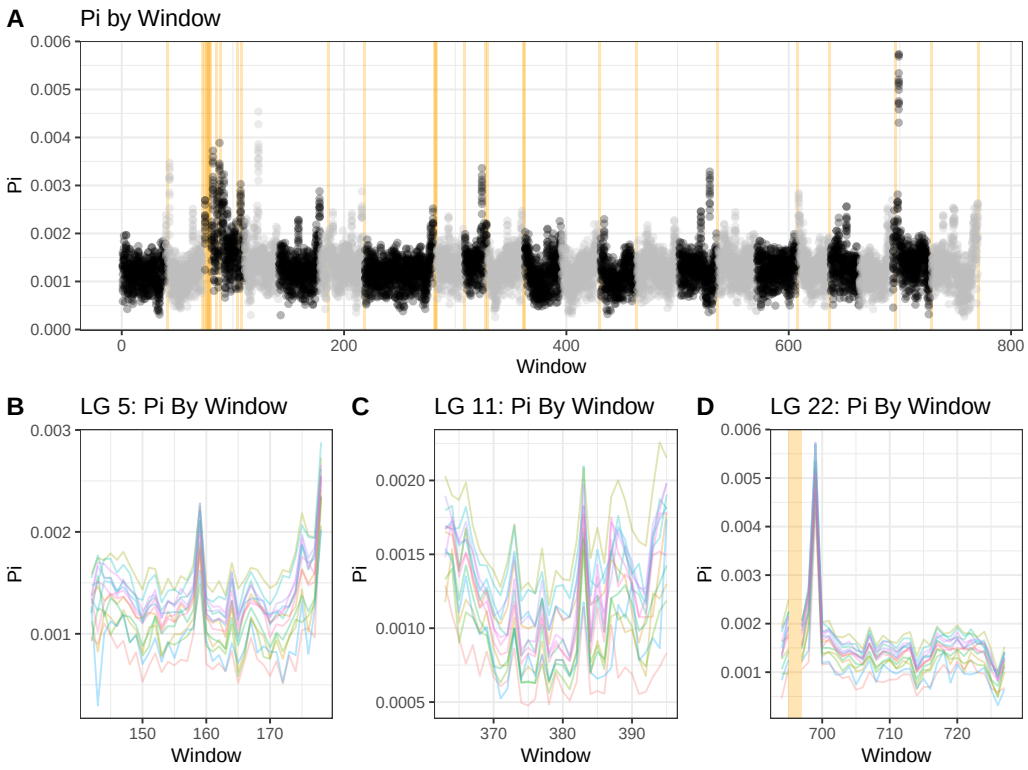

Figure B.2. **(A)** Genetic diversity along the genome in 1Mb windows. **(B, C, D)** Diversity in 1Mb windows for linkage groups 5, 11, and 22.

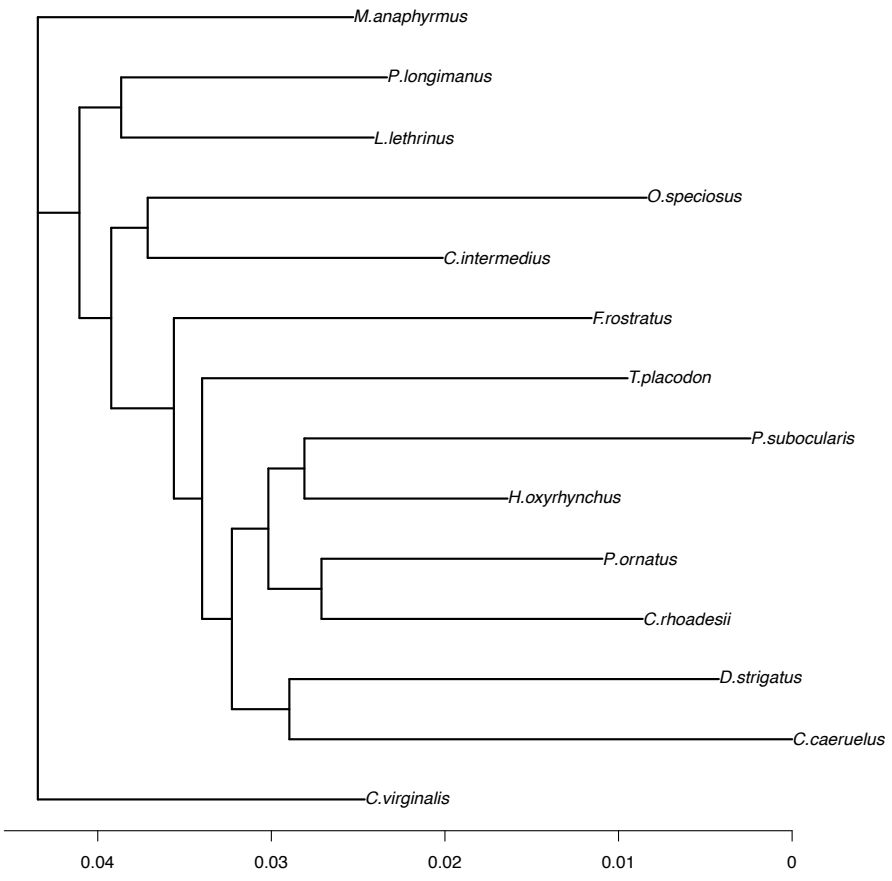

Figure B.3. Whole-genome phylogenetic tree for one individual per species (see text for details).

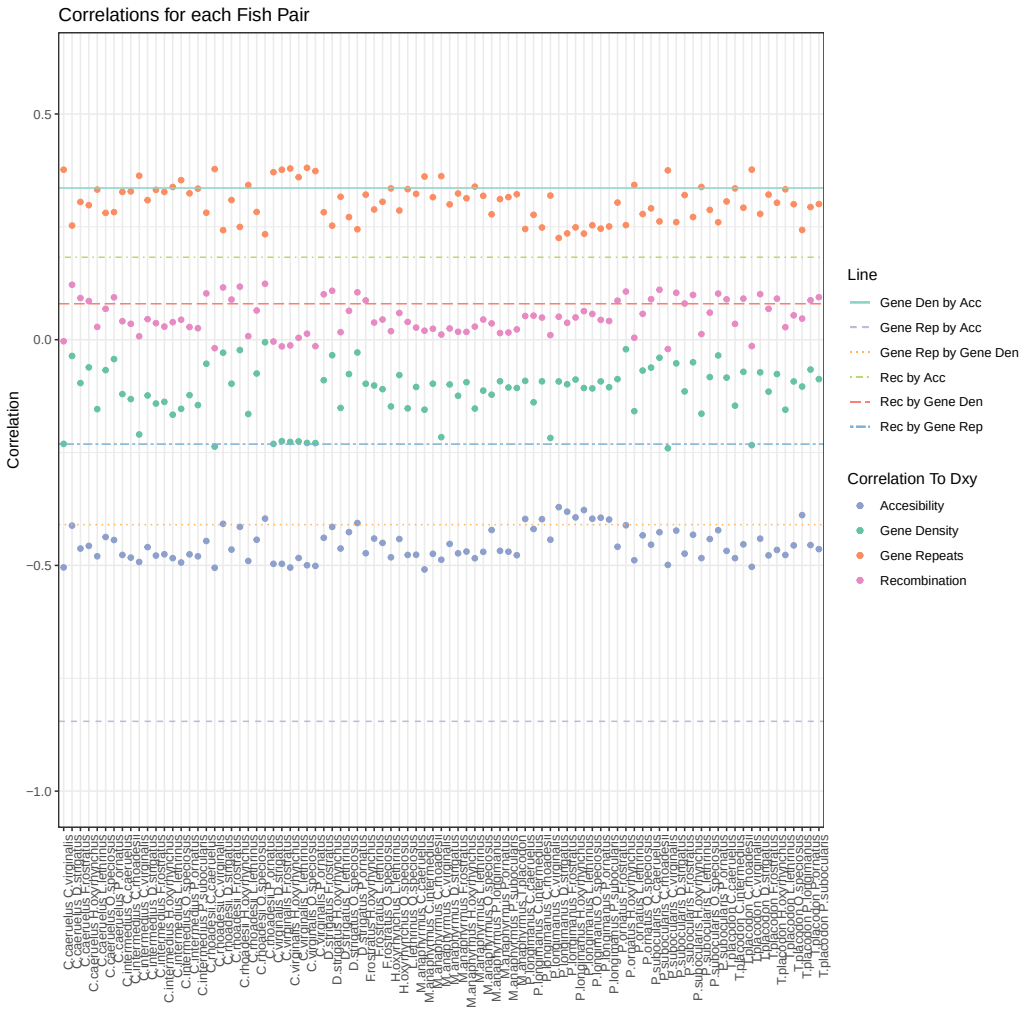

Figure B.4. Whole genome correlations for all species pairs. Each dot is a correlation between d_{XY} of a species pair and accessibility (blue), gene density (green), gene repeat content (orange), or recombination rate (pink). Each line is a correlation between genome functions (gene density by accessibility is a light green solid line, gene repeat by accessibility is a dashed purple line, gene repeat by gene density is an orange dotted line, recombination rate by accessibility is light green dashed and dotted line, recombination rate by gene density is red long dashed line, and recombination rate by gene repeat is a blue two dashed line).

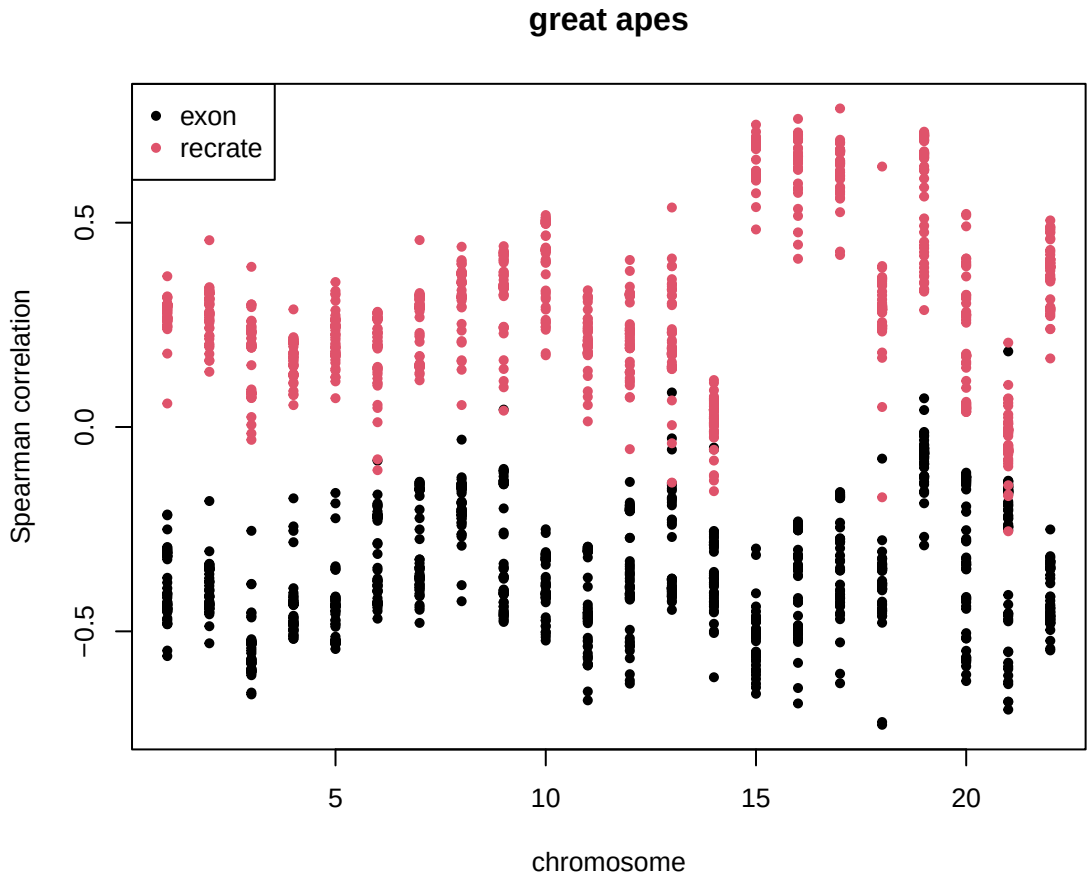

Figure B.5. Correlations of landscapes of diversity and divergence in the great apes with recombination rate ("recrate", red) and exon density ("exon"), by chromosome. Data are re-plotted from Rodrigues, Kern, and Ralph (2024): correlations are Spearman correlations for 1Mb windows, and are shown for genetic diversity and divergence within and between humans, bonobo, chimpanzee, gorilla, and orangutan.

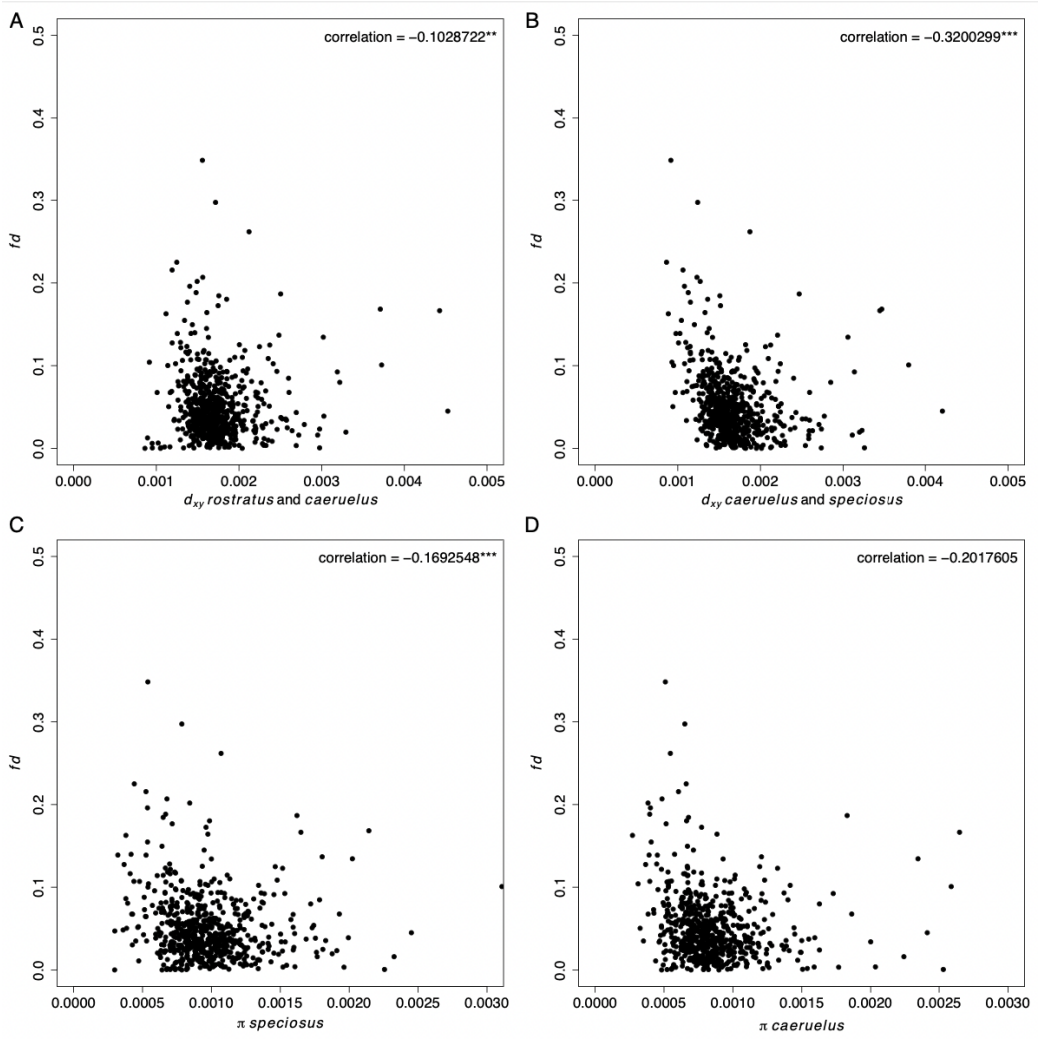

Figure B.6. Each point is a 1Mb window of admixture proportions (*fd*, trios in 3.5). (A) Whole genome *fd* values by d_{XY} between *rostratus* and *caeruelus*. (B) Whole genome *fd* values by d_{XY} between *caeruelus* and *speciosus*. (C) Whole genome *fd* values by π *speciosus*. (D) Whole genome *fd* values by π *caeruelus*.

APPENDIX C

WINDOW DXY BY SPECIES

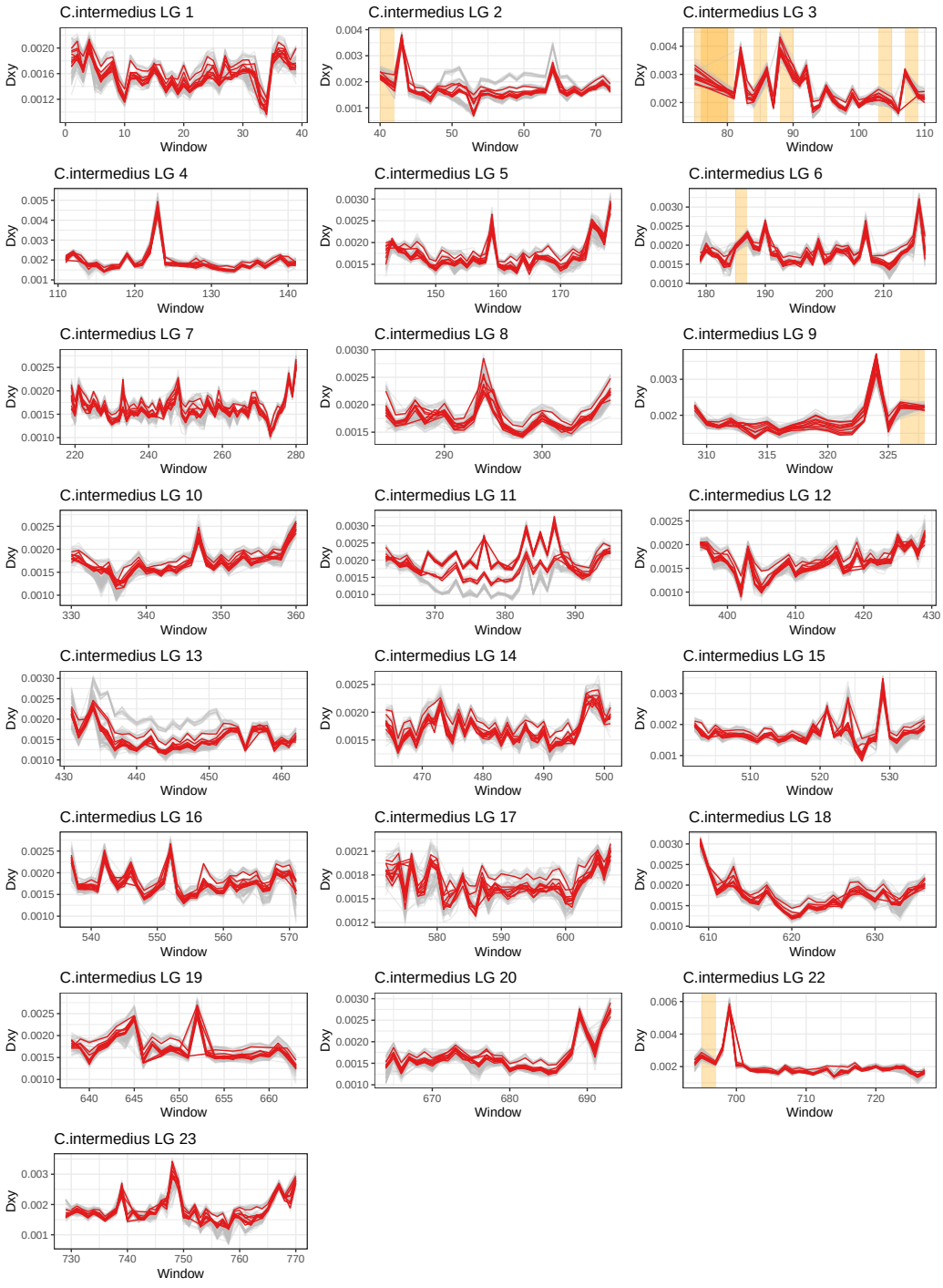

Figure C.1. 1Mb window d_{XY} for all linkage groups, C.intermedius

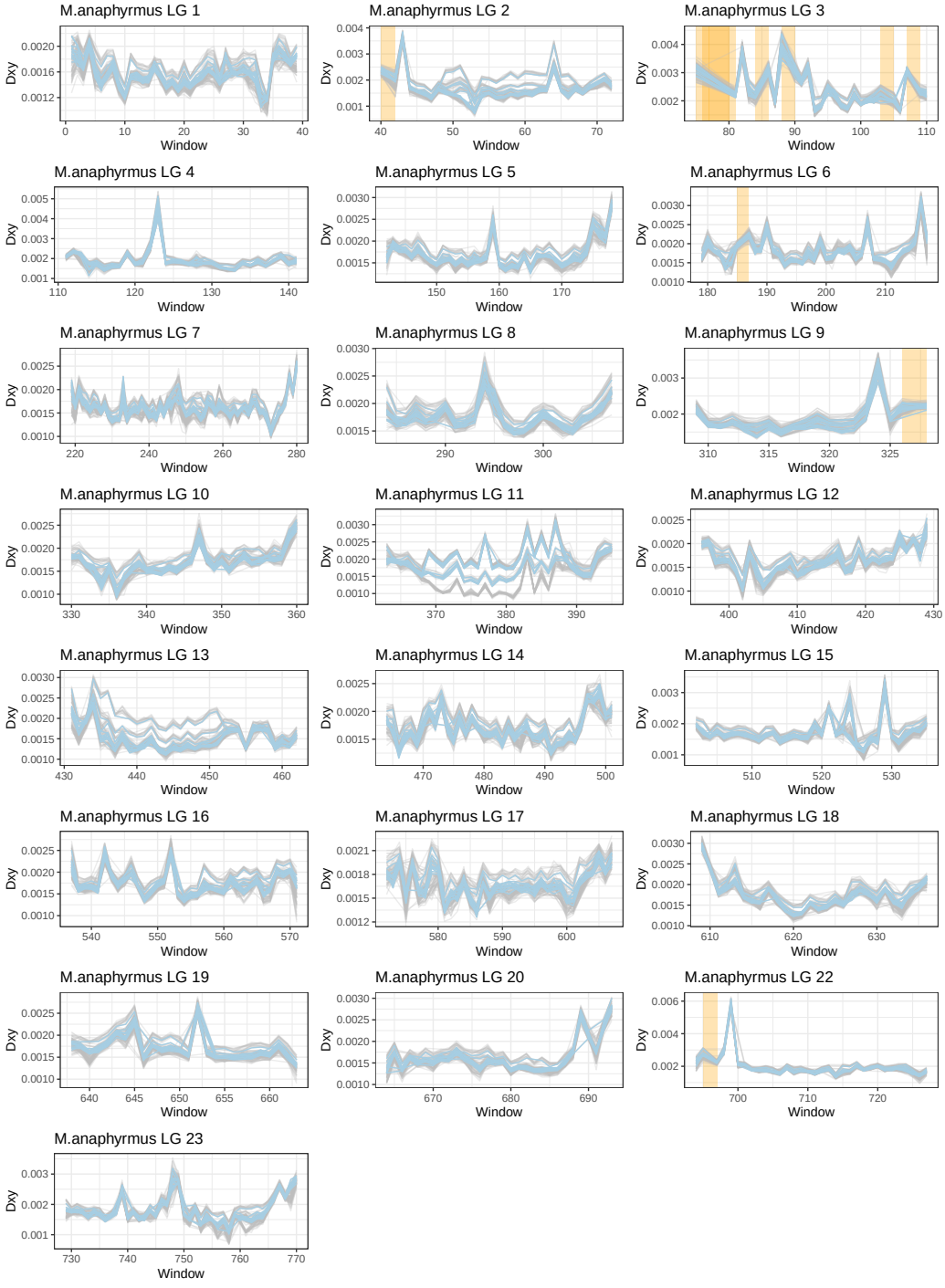

Figure C.2. 1Mb window d_{XY} for all linkage groups, M.anaphyrmus

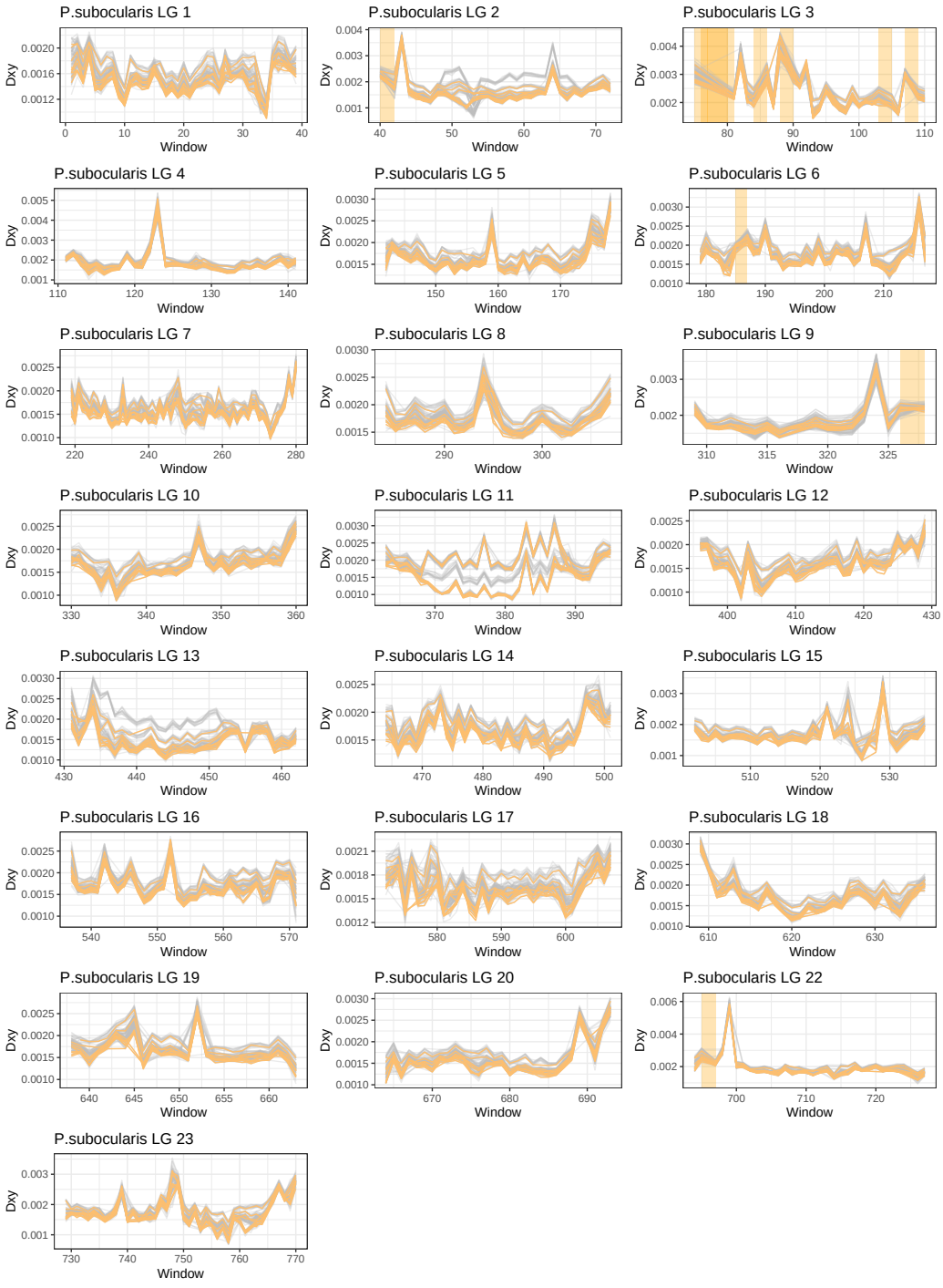

Figure C.3. 1Mb window d_{XY} for all linkage groups, P.subocularis

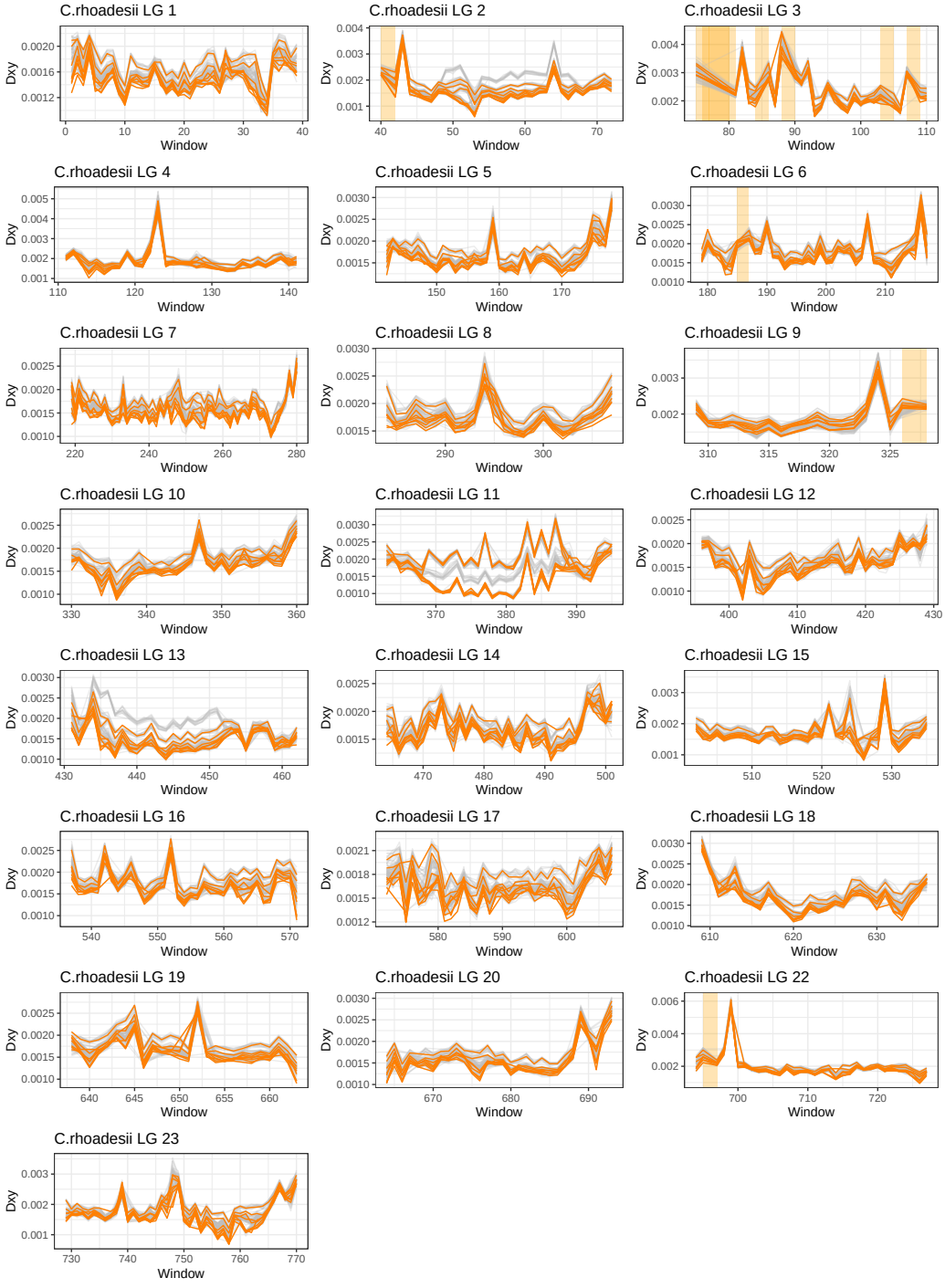

Figure C.4. 1Mb window d_{XY} for all linkage groups, C.rhoadesii

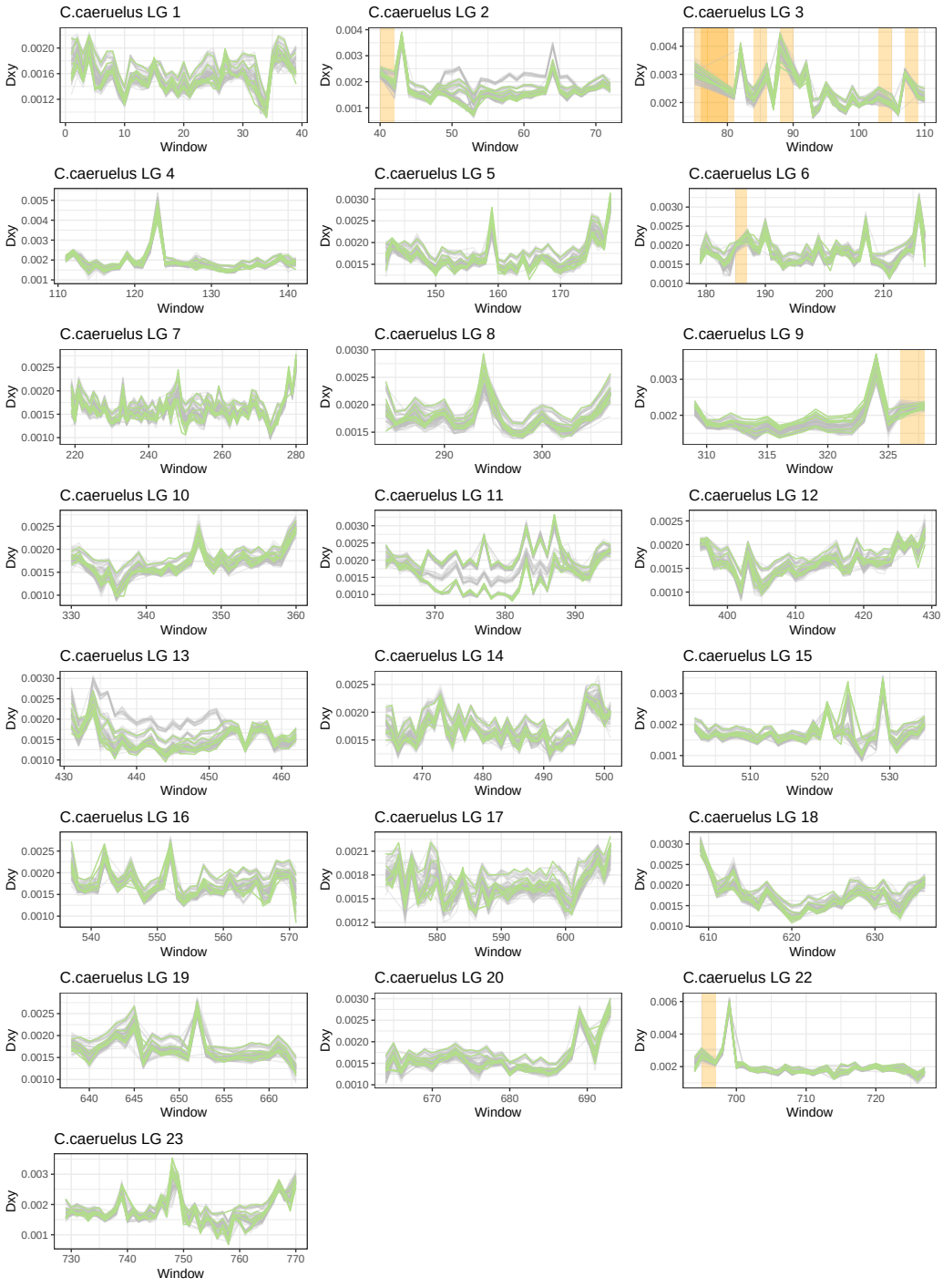

Figure C.5. 1Mb window d_{XY} for all linkage groups, C.caeruelus

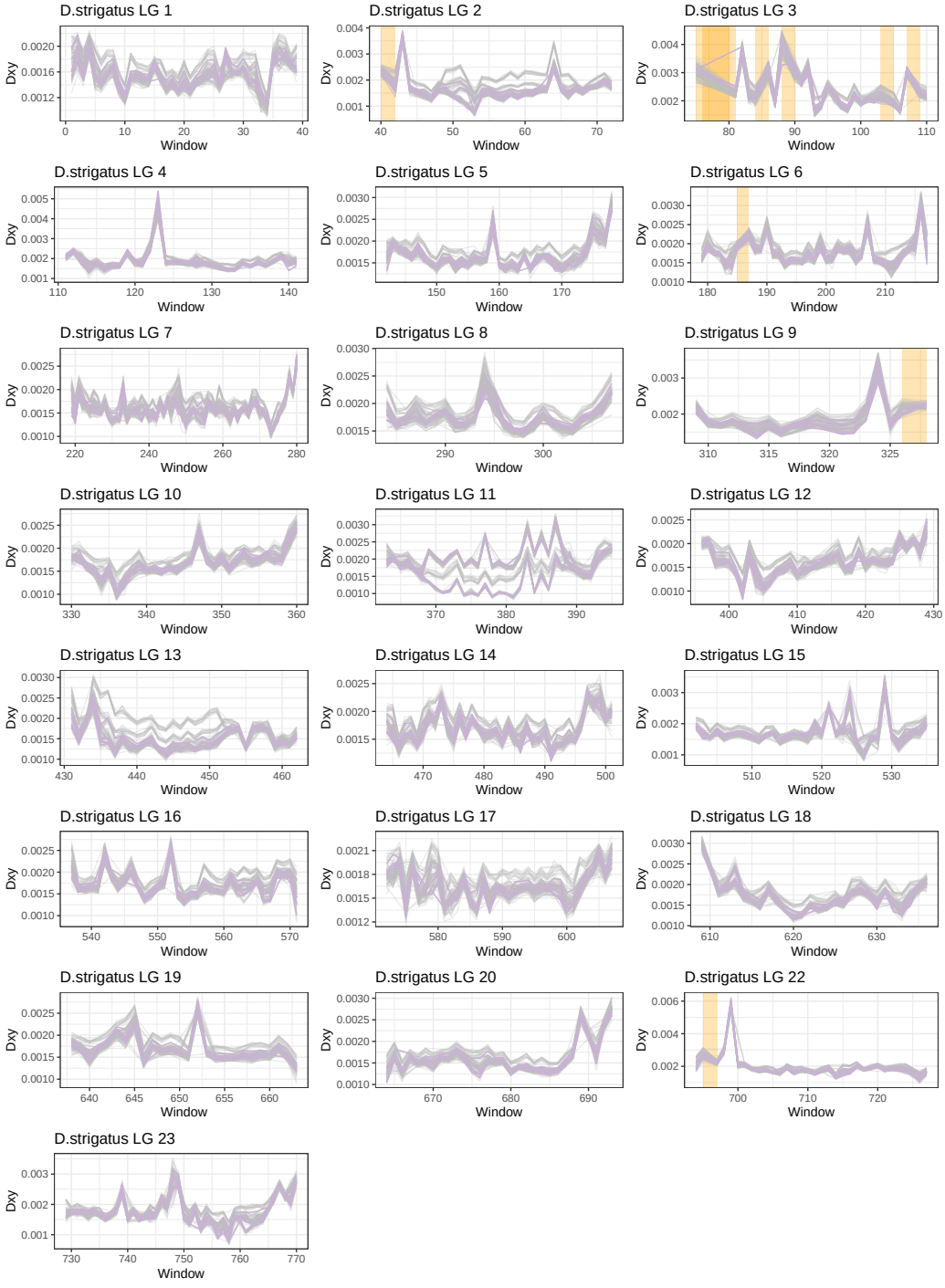

Figure C.6. 1Mb window d_{XY} for all linkage groups, D.strigatus

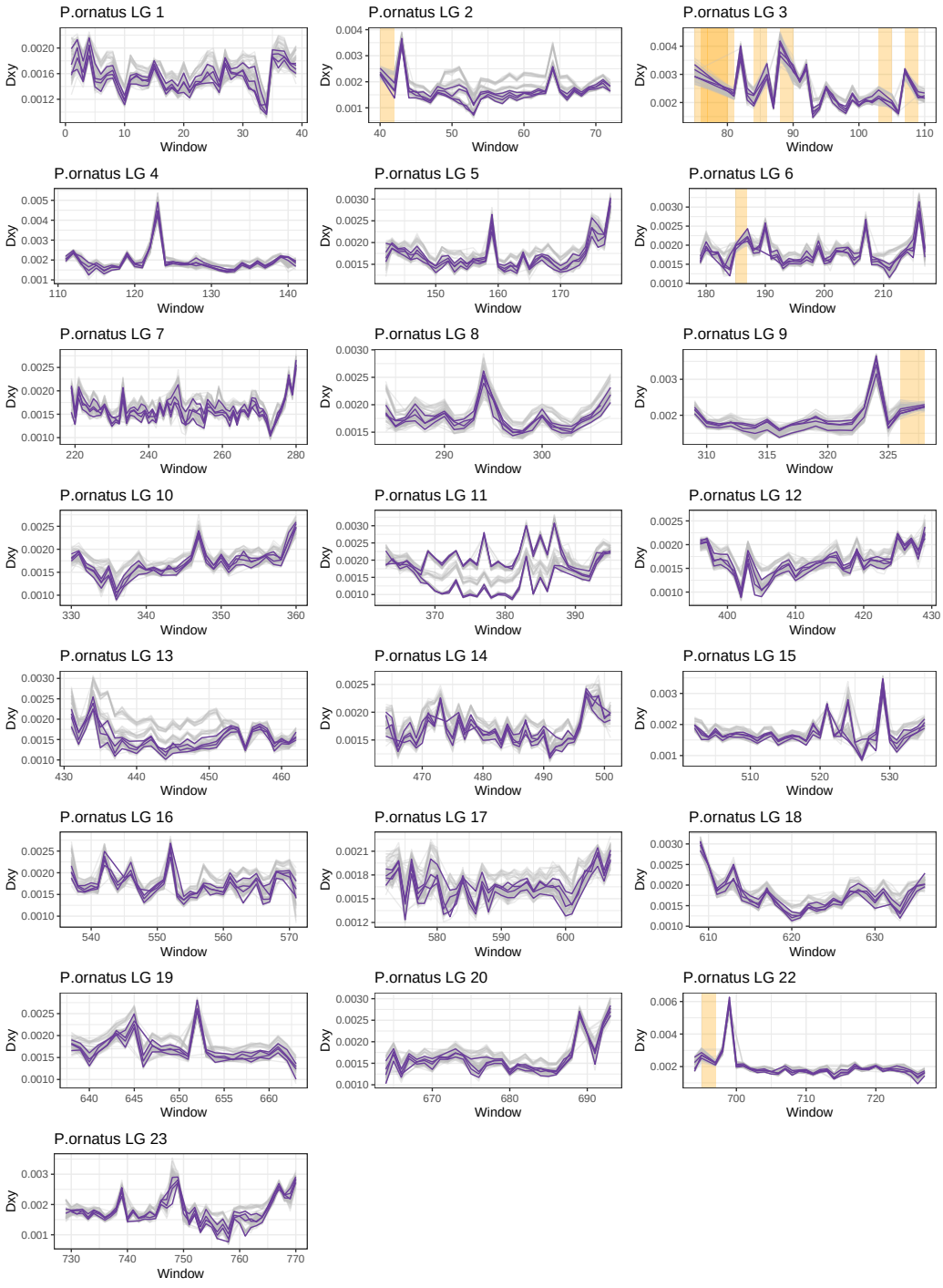

Figure C.7. 1Mb window d_{XY} for all linkage groups, P.ornatus

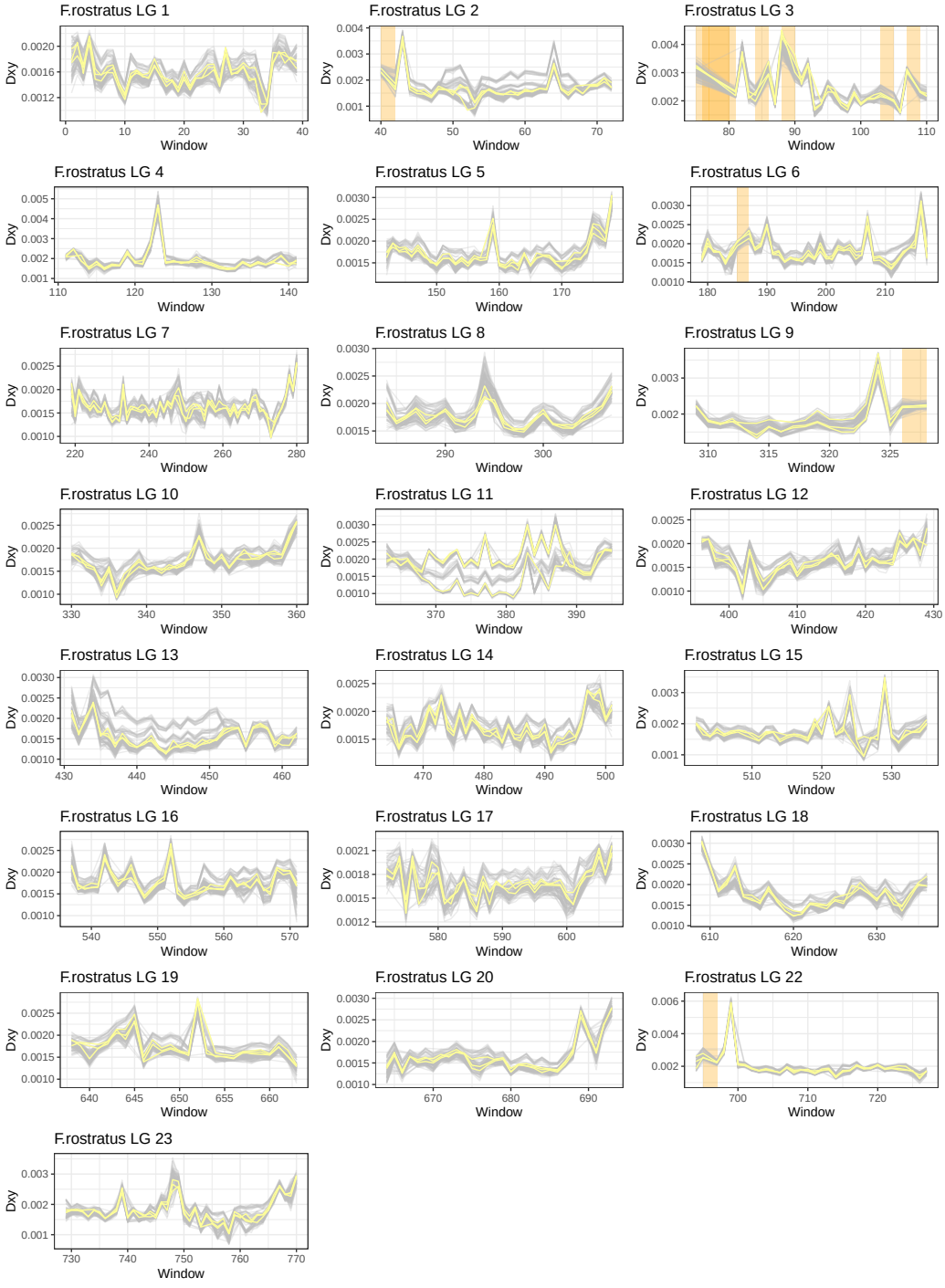

Figure C.8. 1Mb window d_{XY} for all linkage groups, F.rostratus

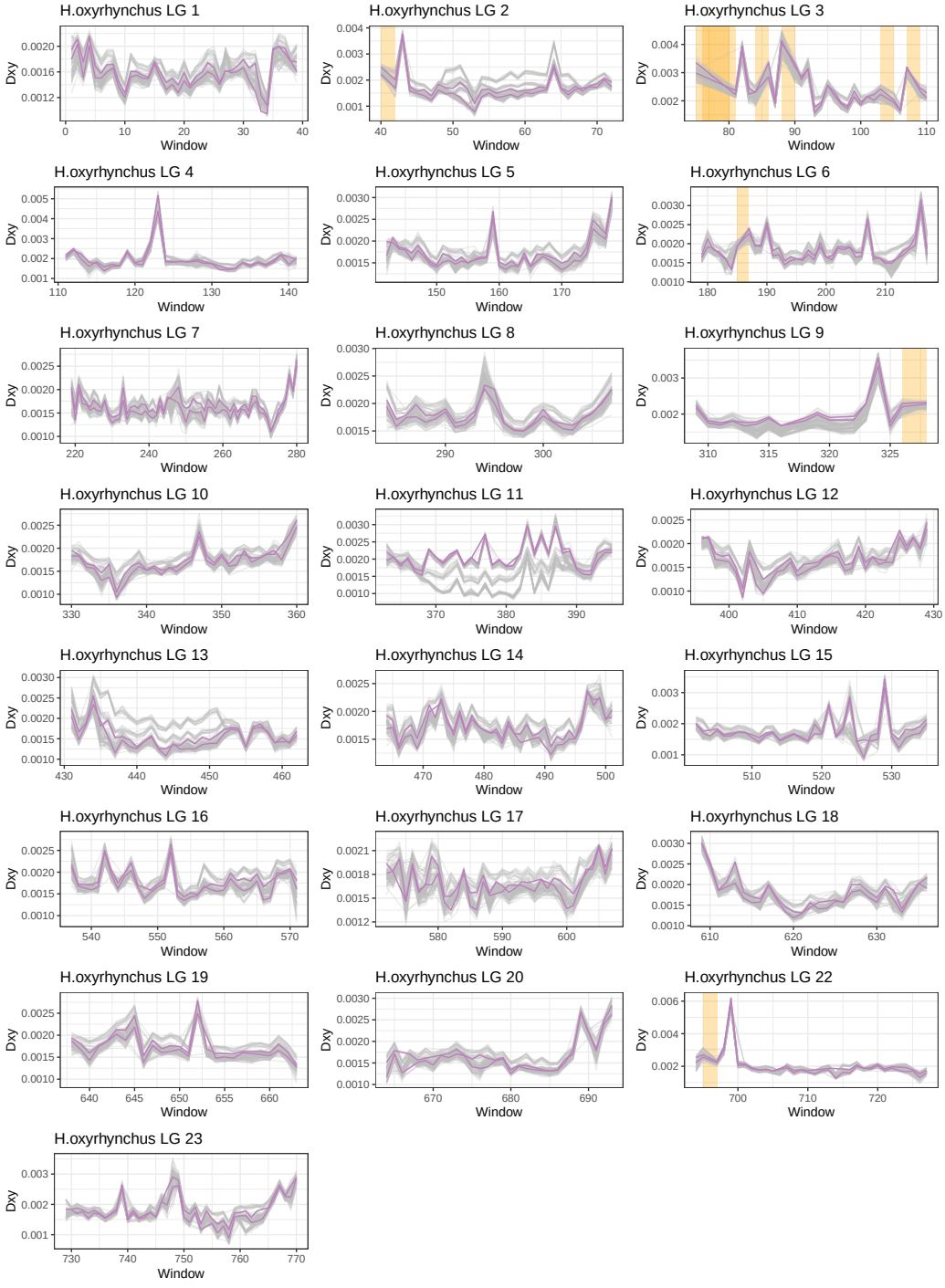

Figure C.9. 1Mb window d_{XY} for all linkage groups, H.oxyrhynchus

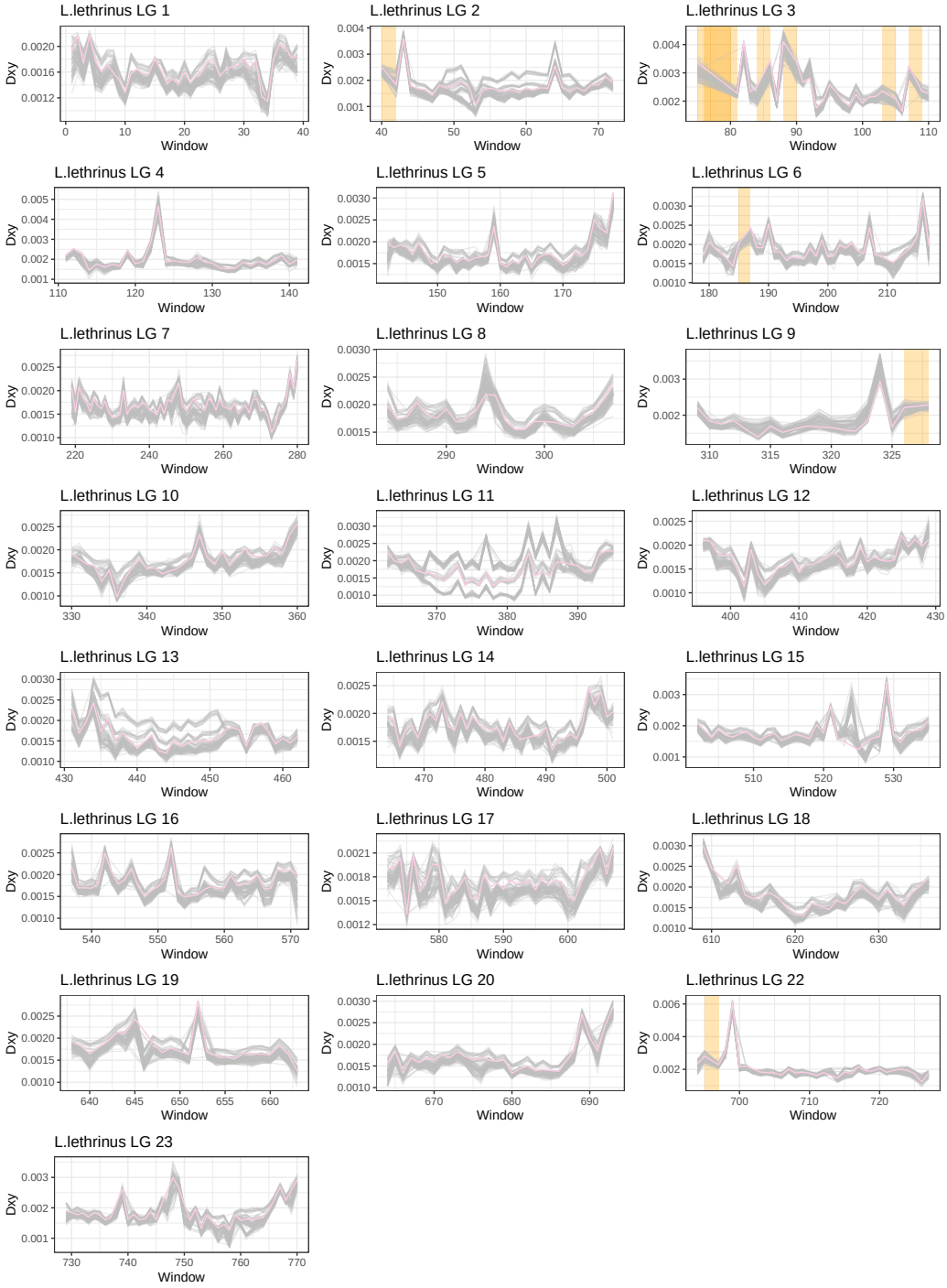

Figure C.10. 1Mb window d_{XY} for all linkage groups, L.lethrinus

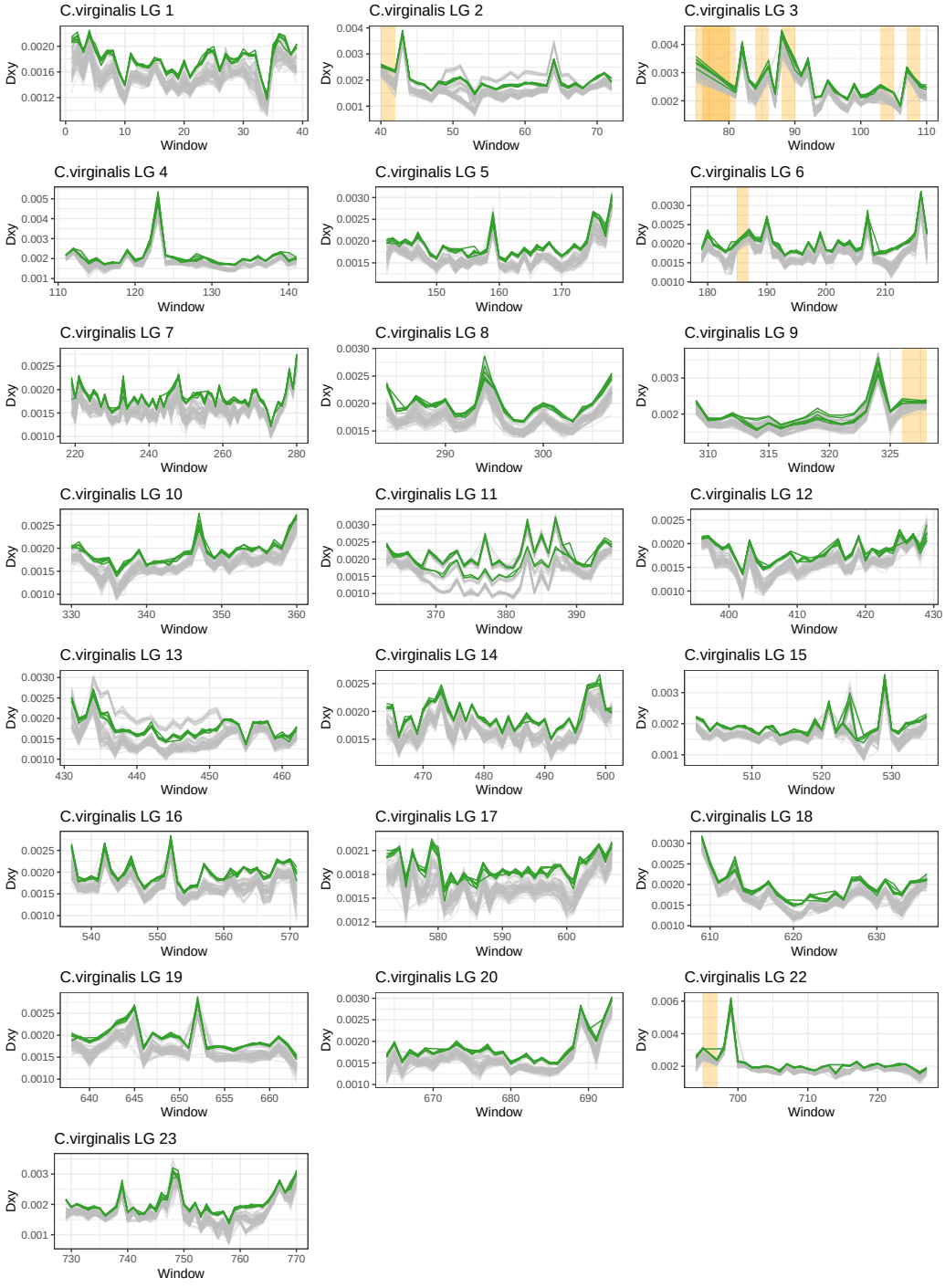

Figure C.11. 1Mb window d_{XY} for all linkage groups, C.virginalis

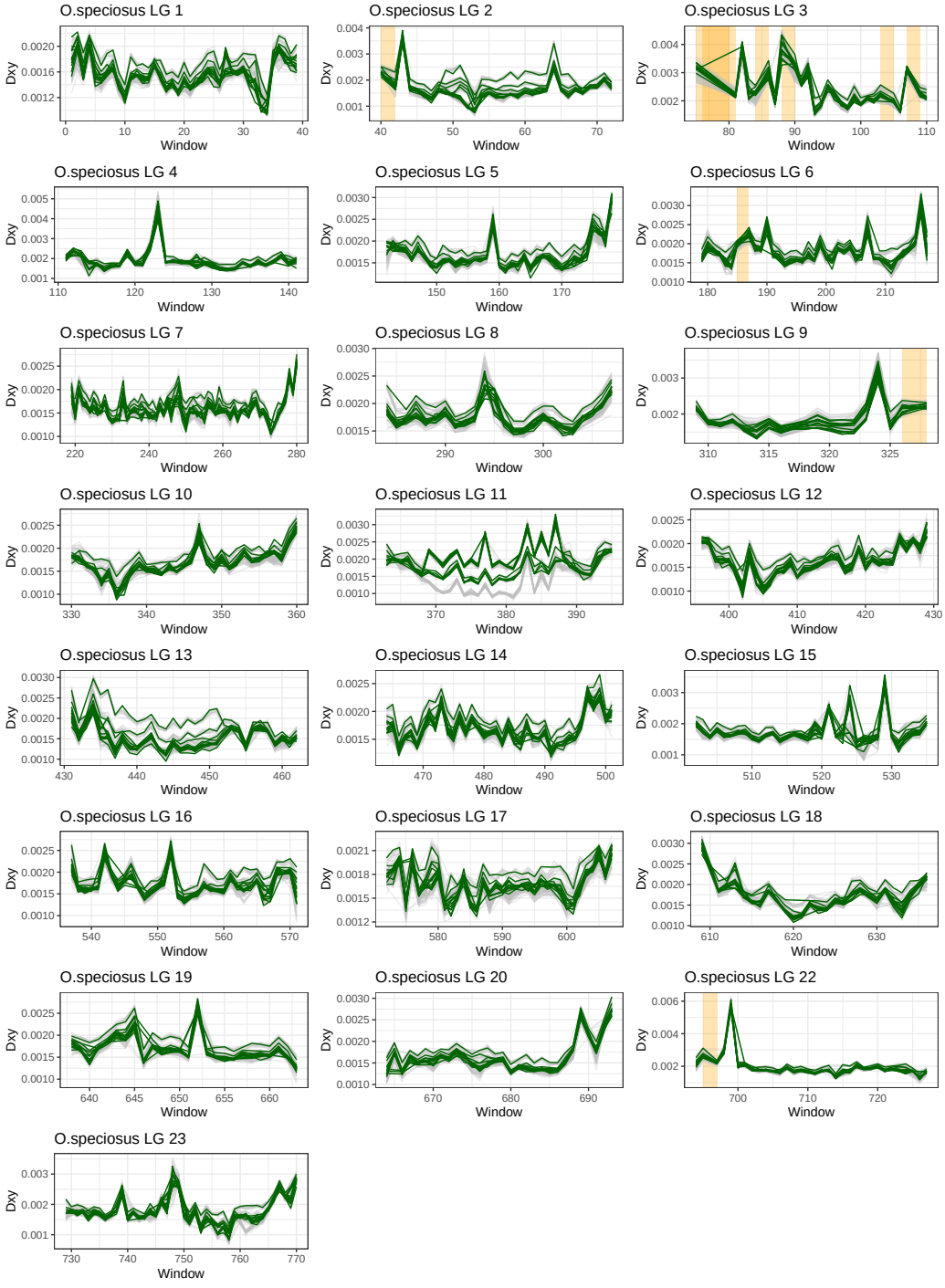

Figure C.12. 1Mb window d_{XY} for all linkage groups, O.speciosus

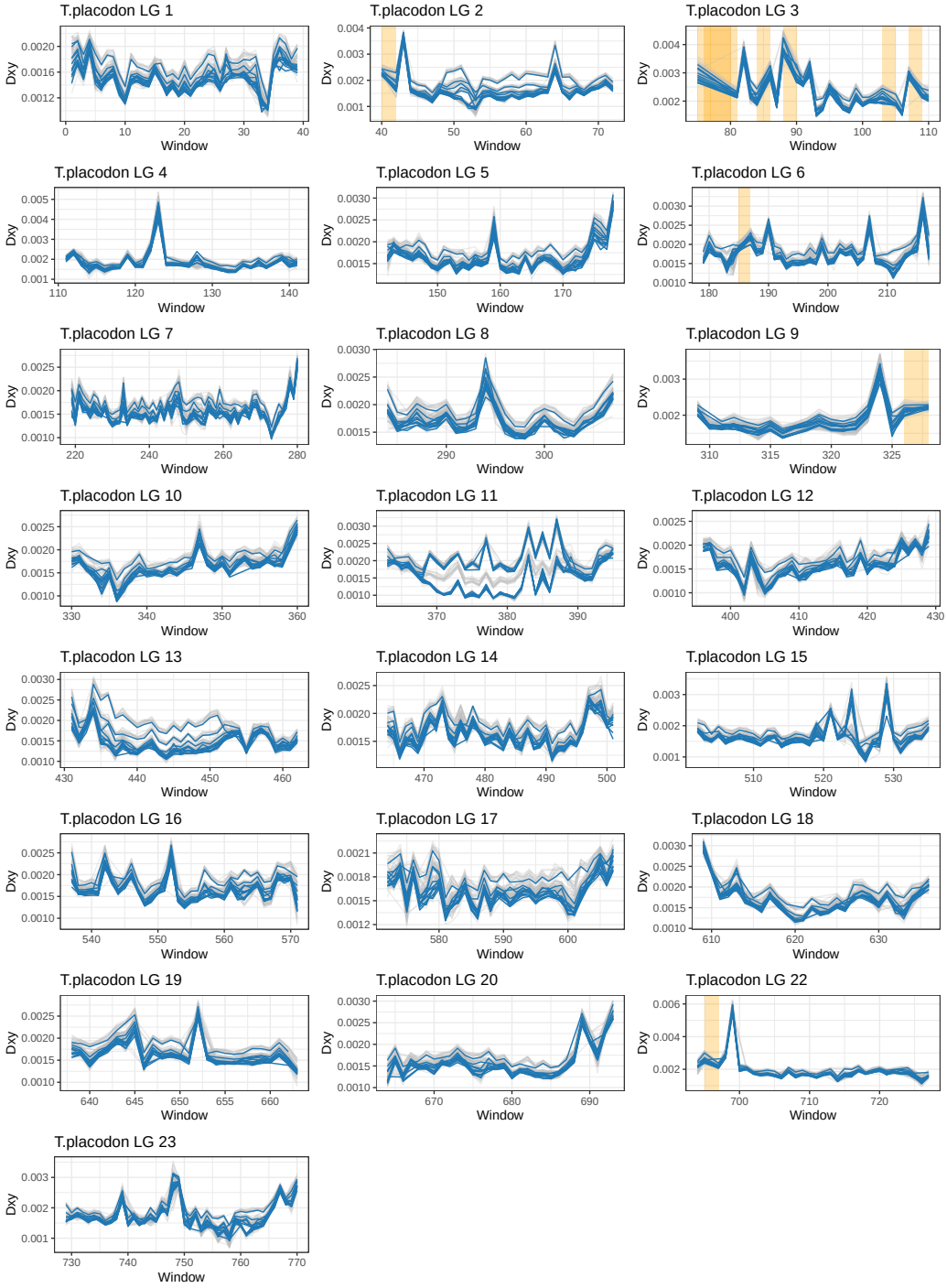

Figure C.13. 1Mb window d_{XY} for all linkage groups, T.placodon

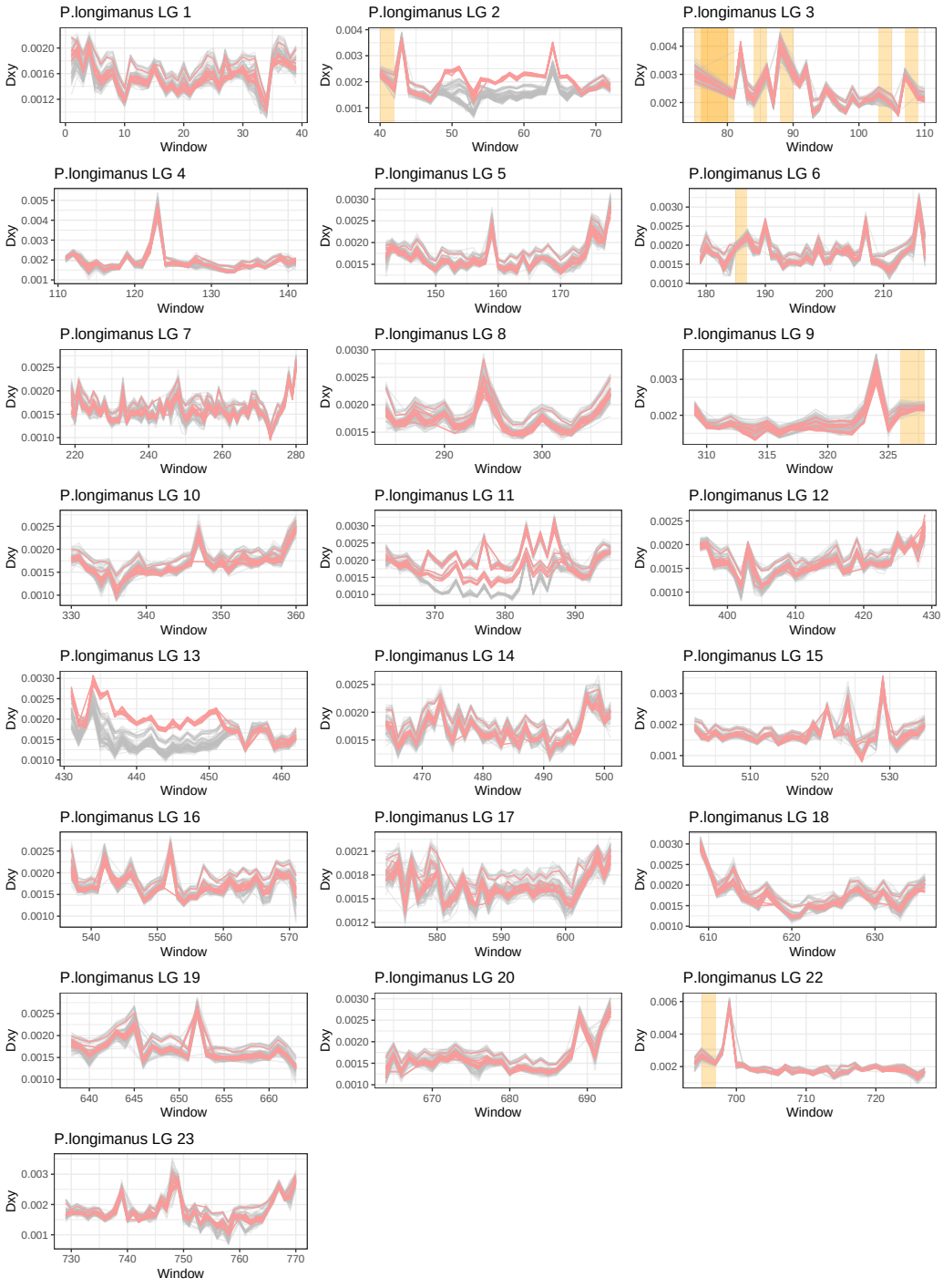

Figure C.14. 1Mb window d_{XY} for all linkage groups, P.longimanus

APPENDIX D

WINDOW π BY SPECIES

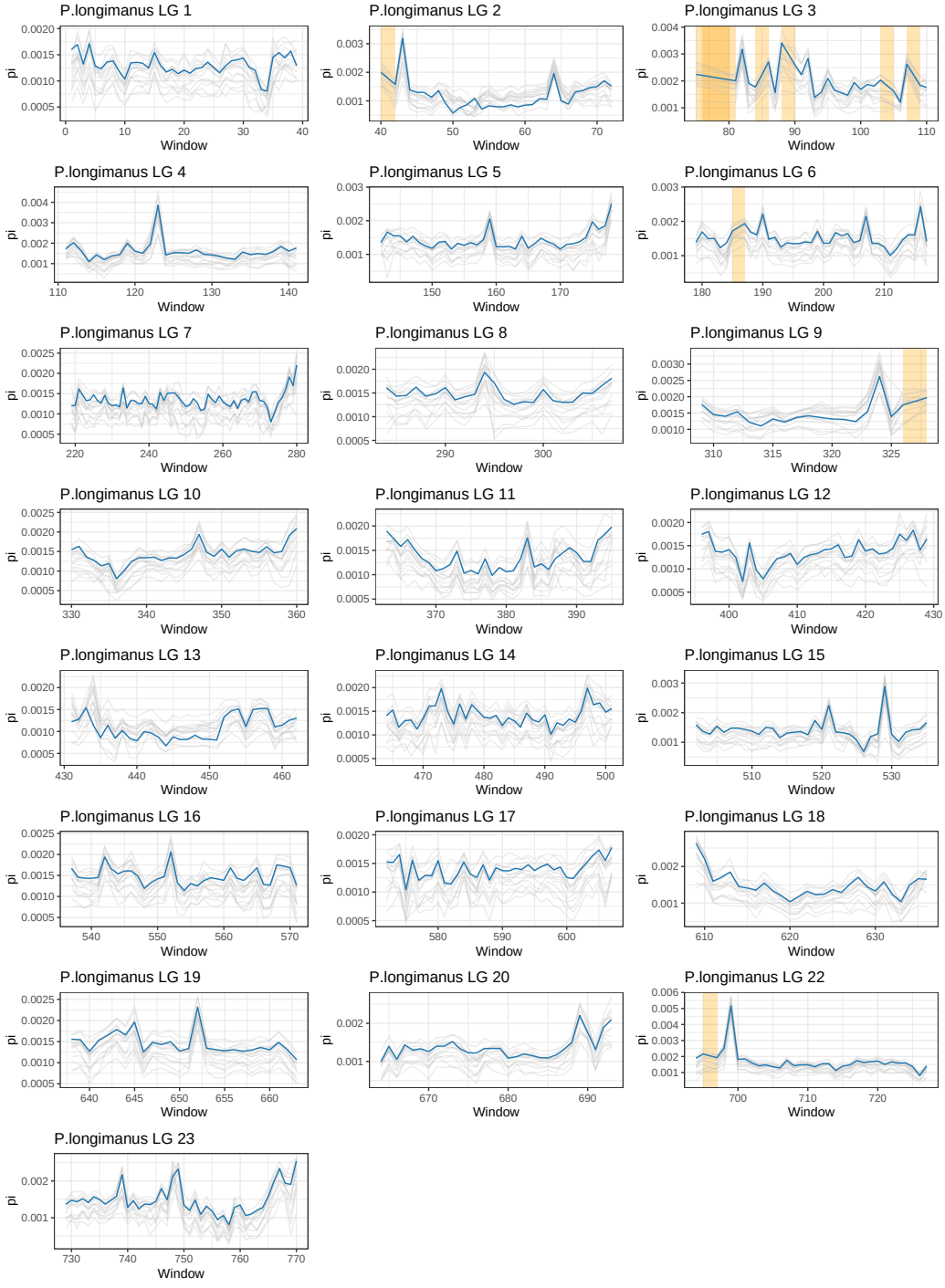

Figure D.1. 1Mb window π for all linkage groups, C.intermedius

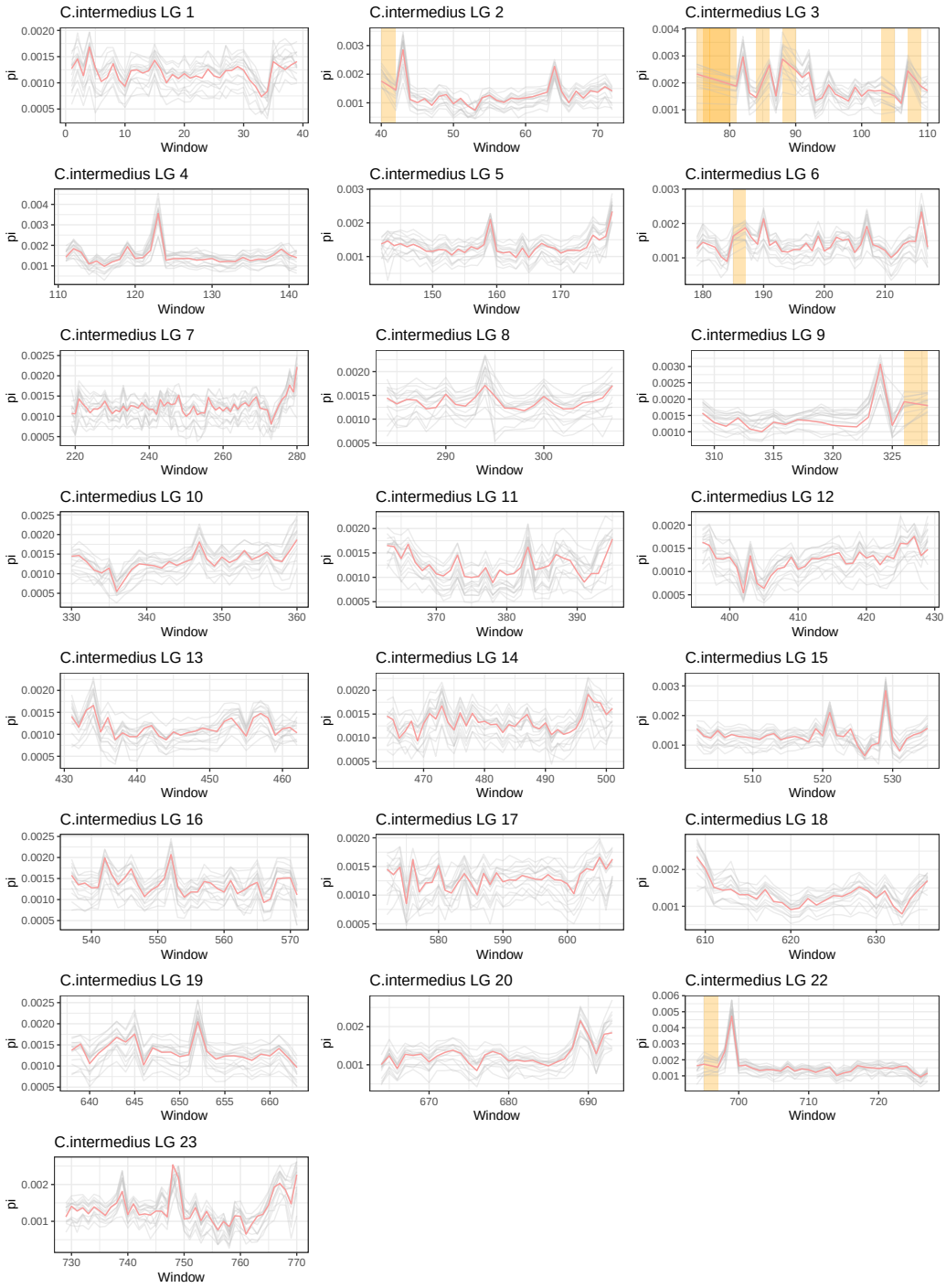

Figure D.2. 1Mb window π for all linkage groups, M.anaphyrmus

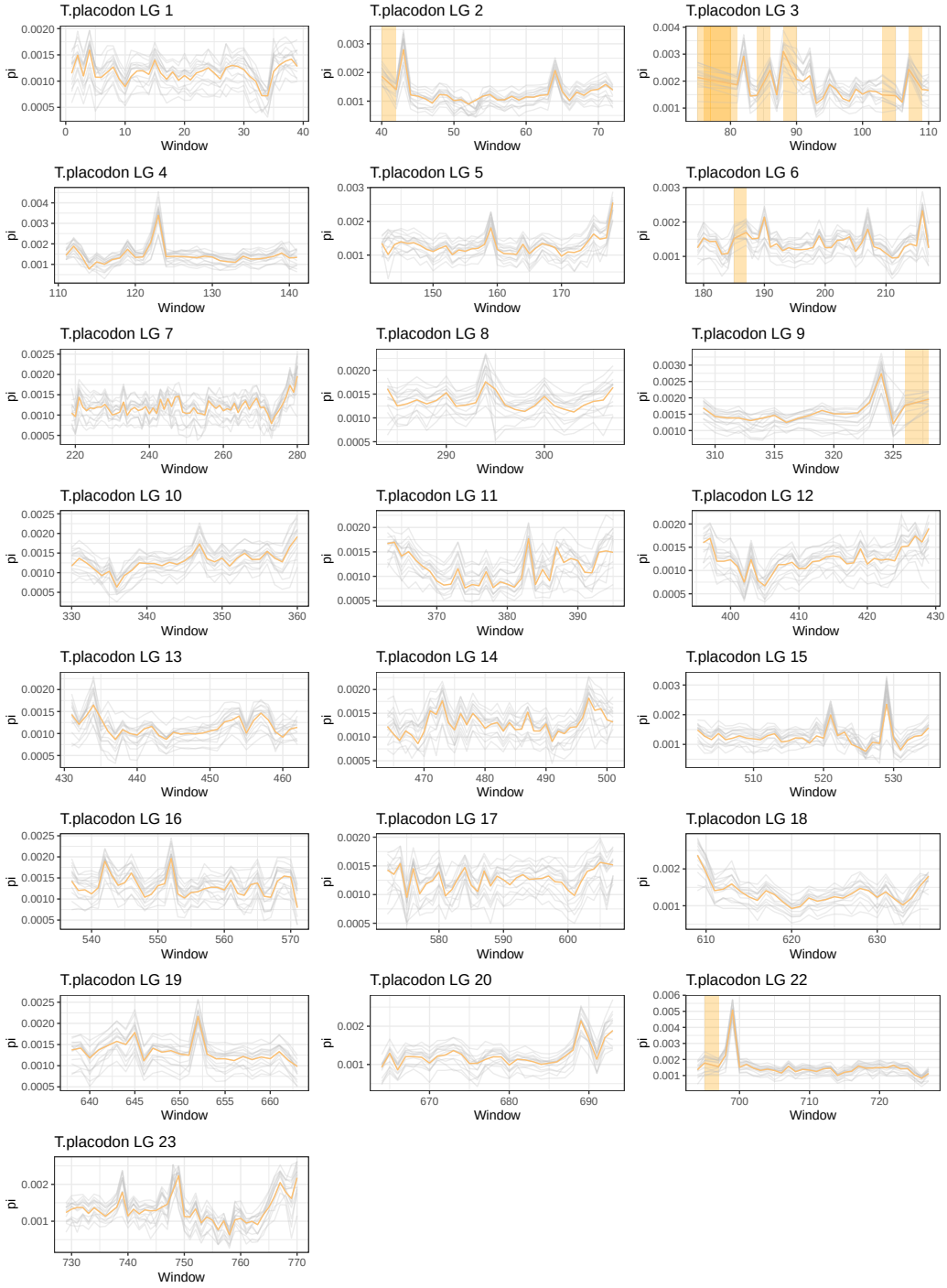

Figure D.3. 1Mb window π for all linkage groups, P.subocularis

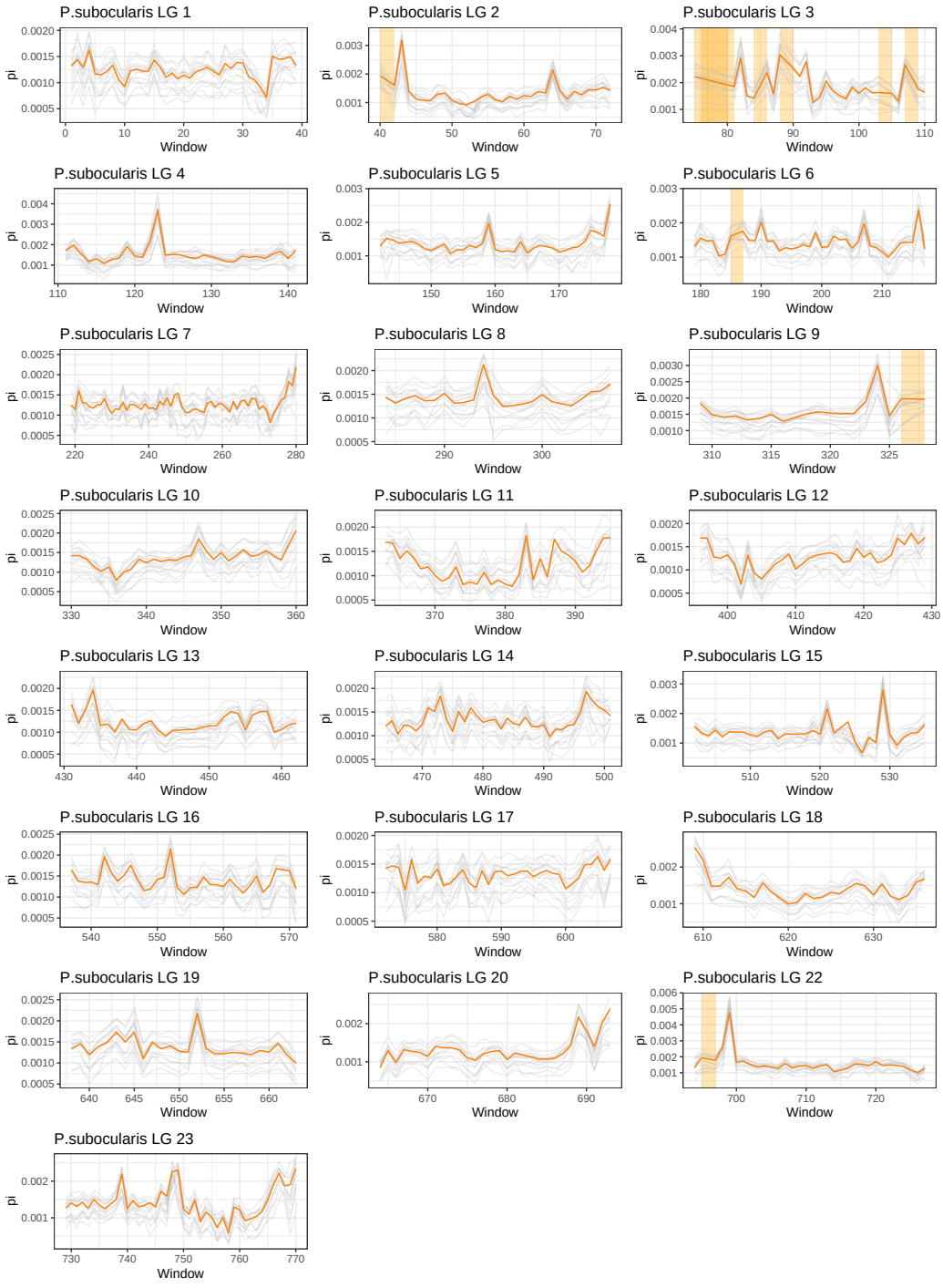

Figure D.4. 1Mb window π for all linkage groups, C.rhoadesii

Figure D.5. 1Mb window π for all linkage groups, C.caeruelus

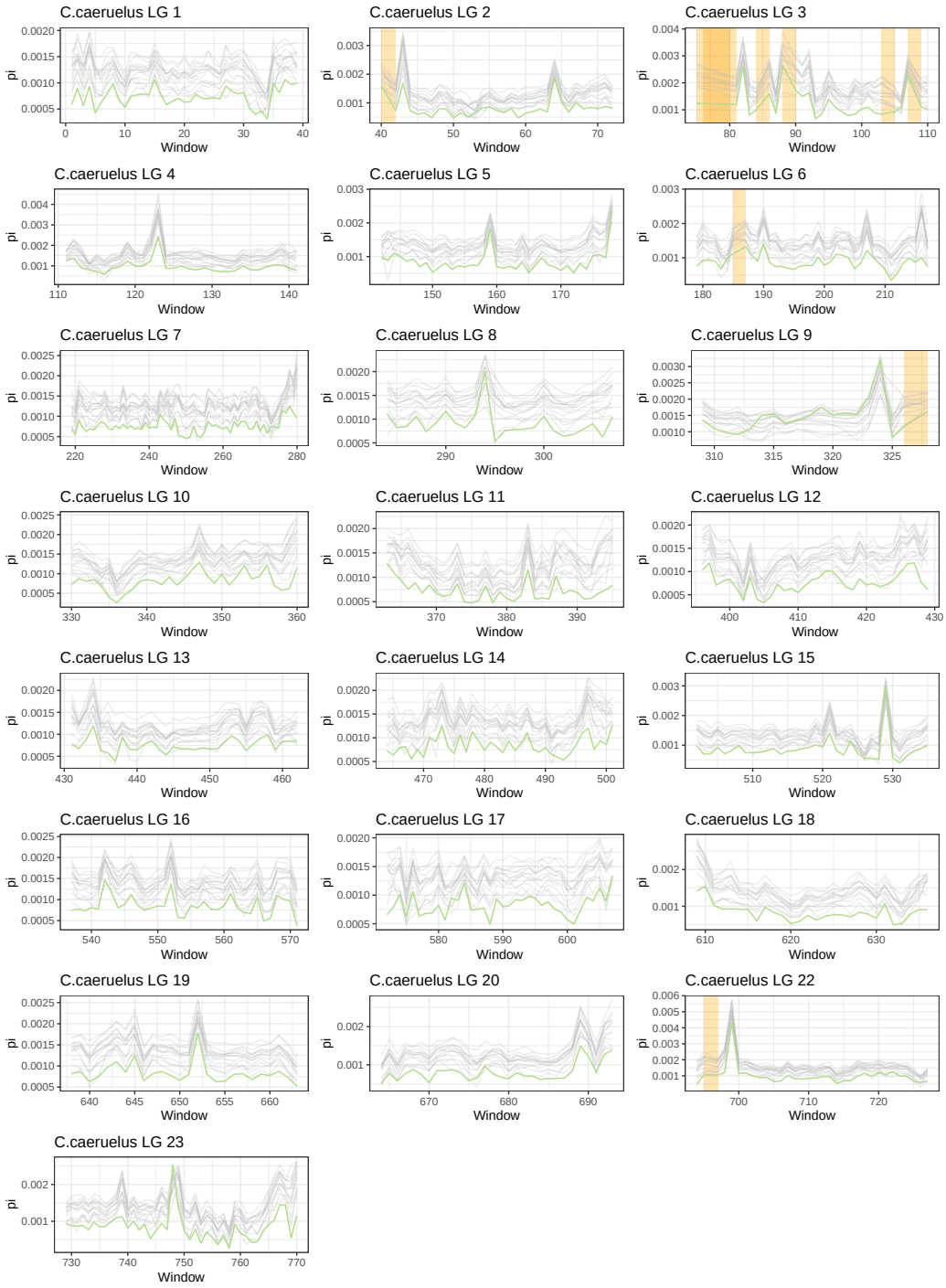

Figure D.6. 1Mb window π for all linkage groups, D.strigatus

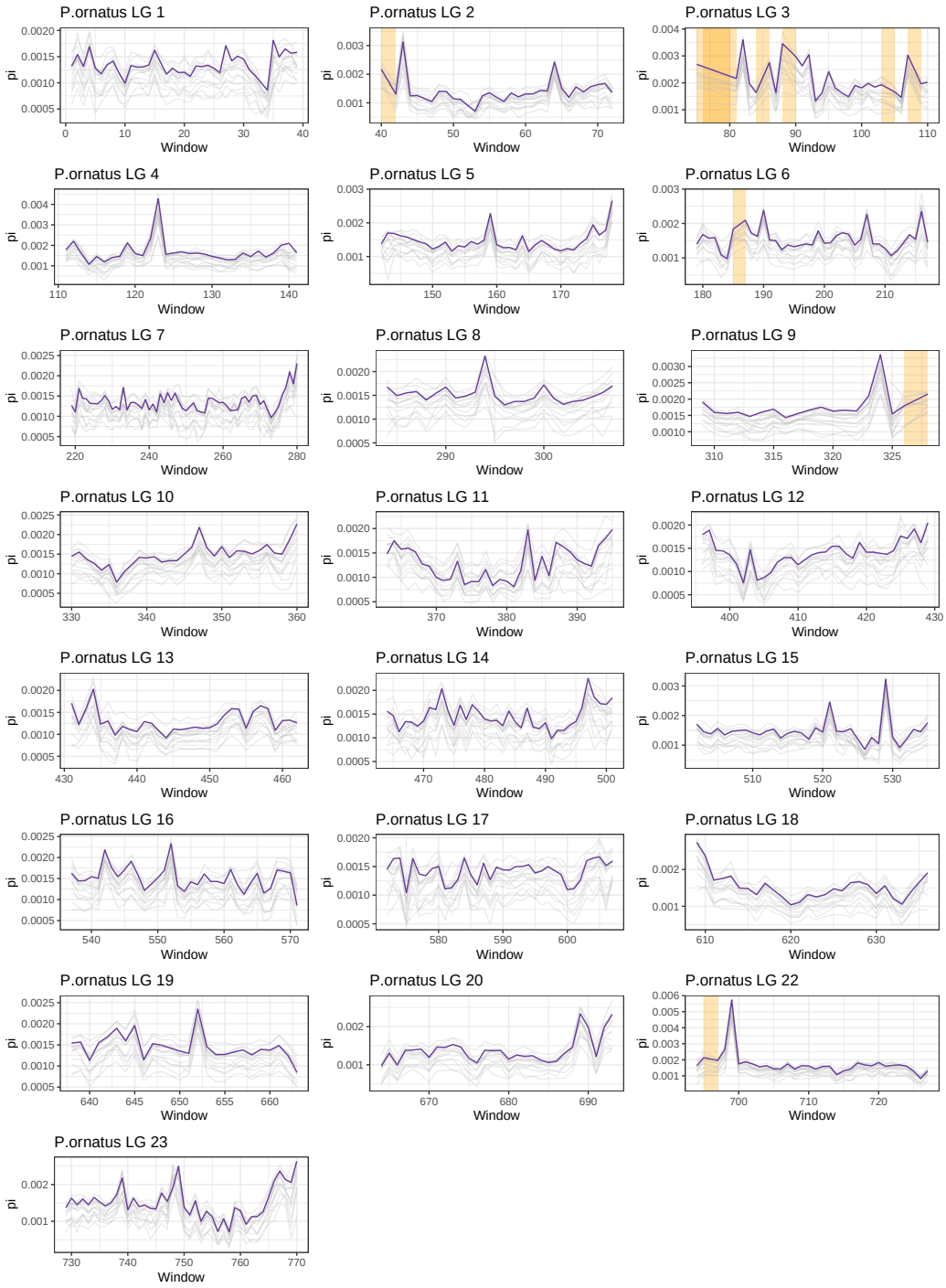

Figure D.7. 1Mb window π for all linkage groups, P.ornatus

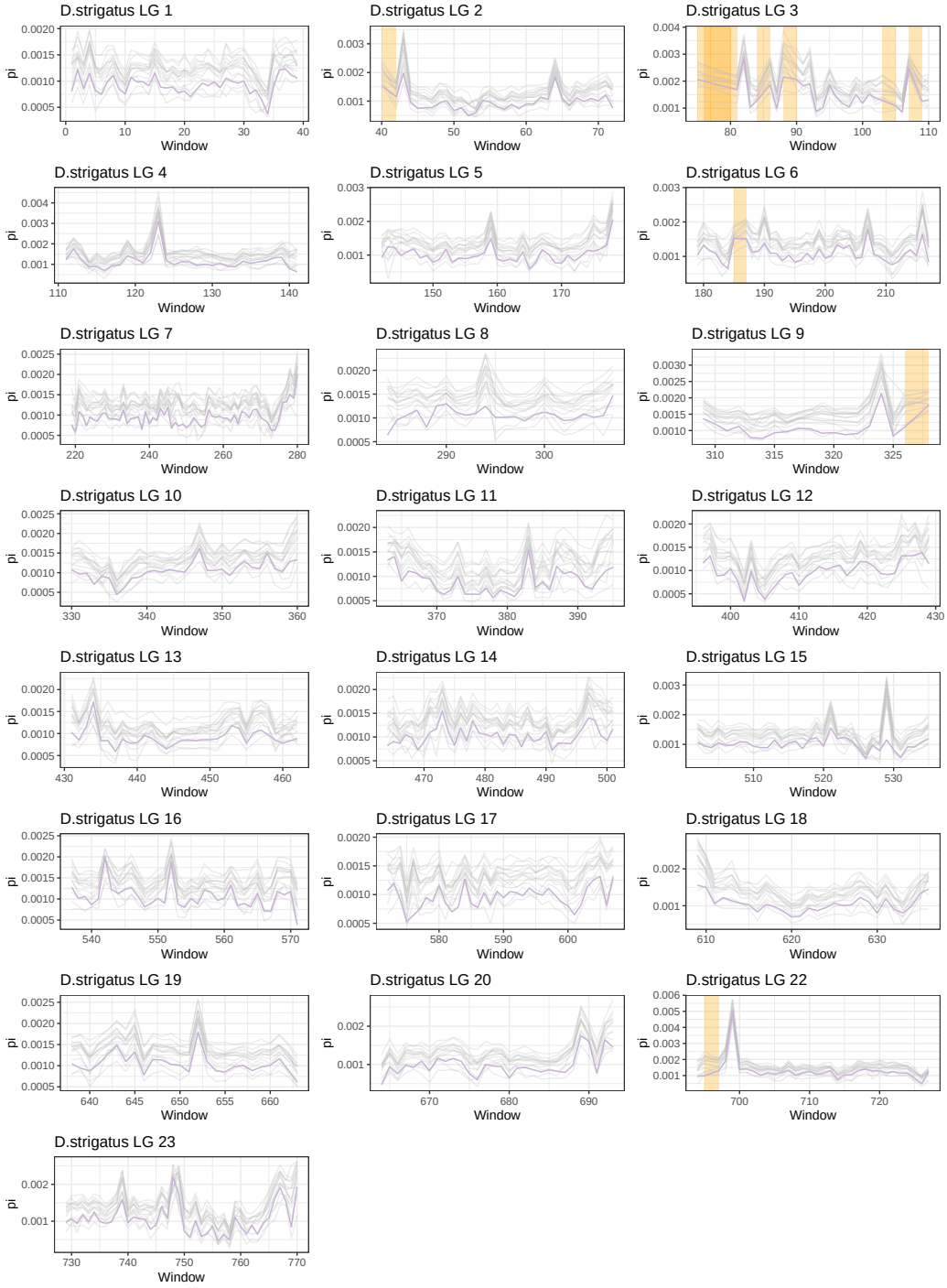

Figure D.8. 1Mb window π for all linkage groups, F.rostratus

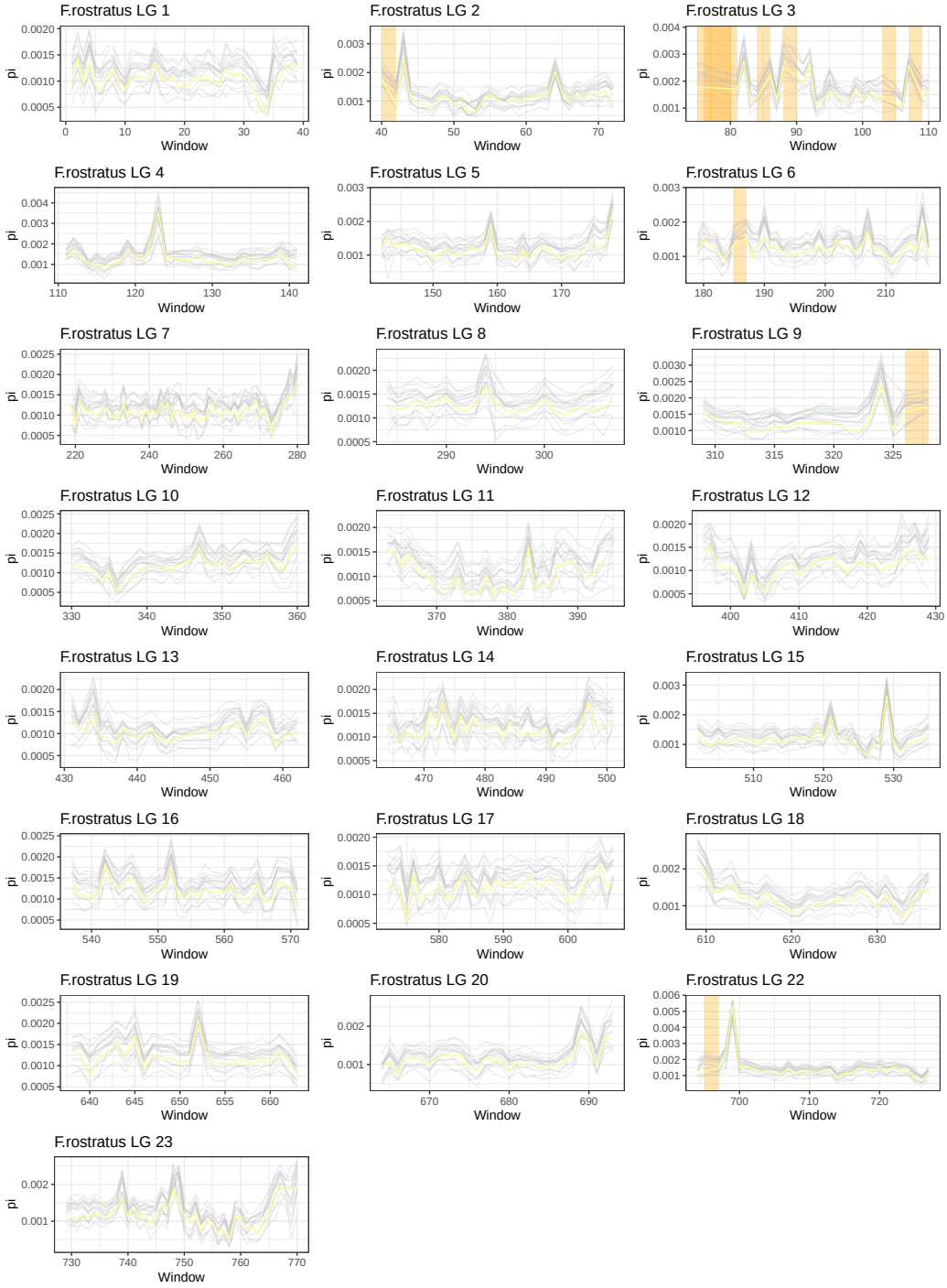

Figure D.9. 1Mb window π for all linkage groups, H.oxyrhynchus

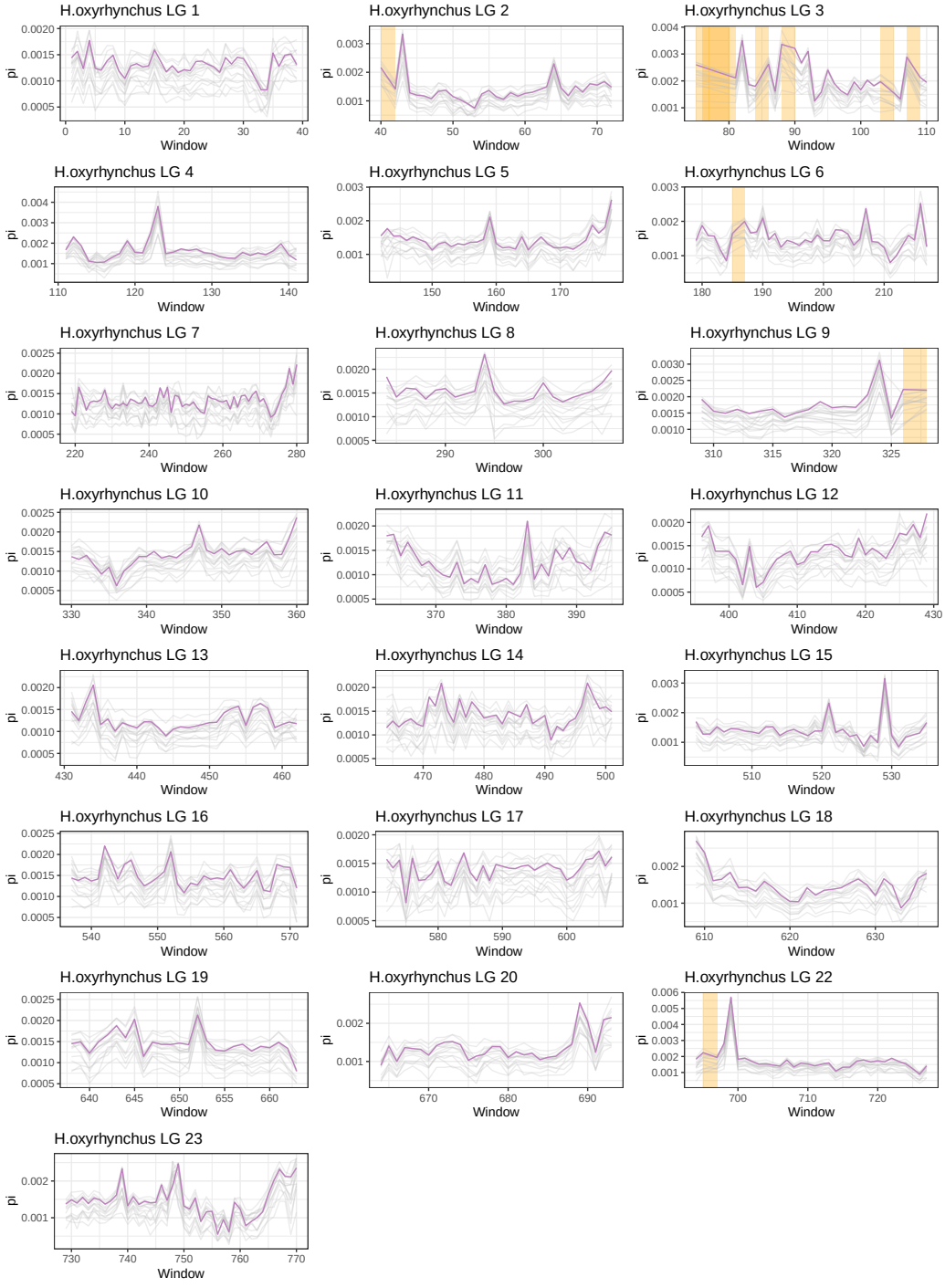

Figure D.10. 1Mb window π for all linkage groups, L.lethrinus

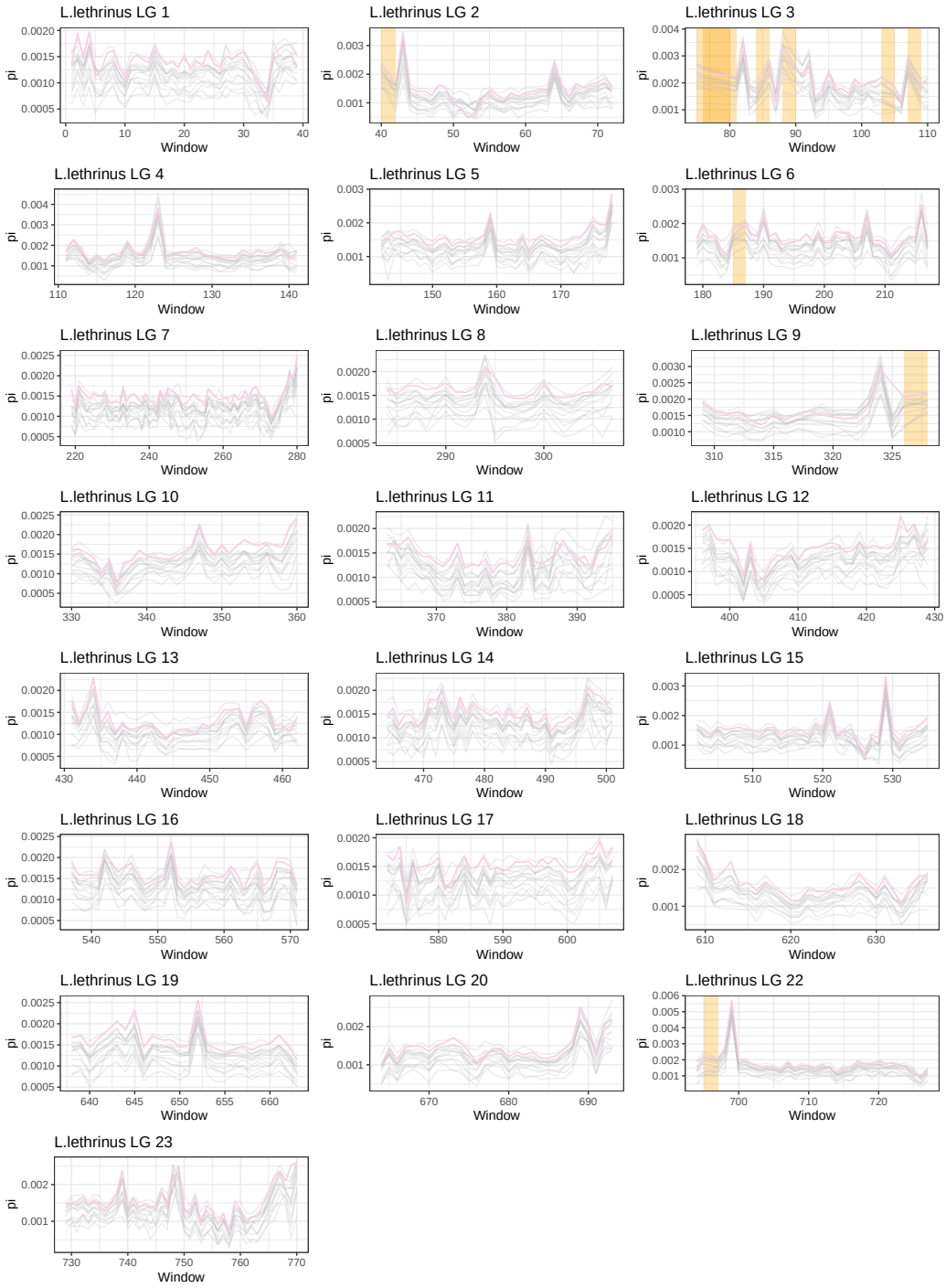

Figure D.11. 1Mb window π for all linkage groups, C.virginalis

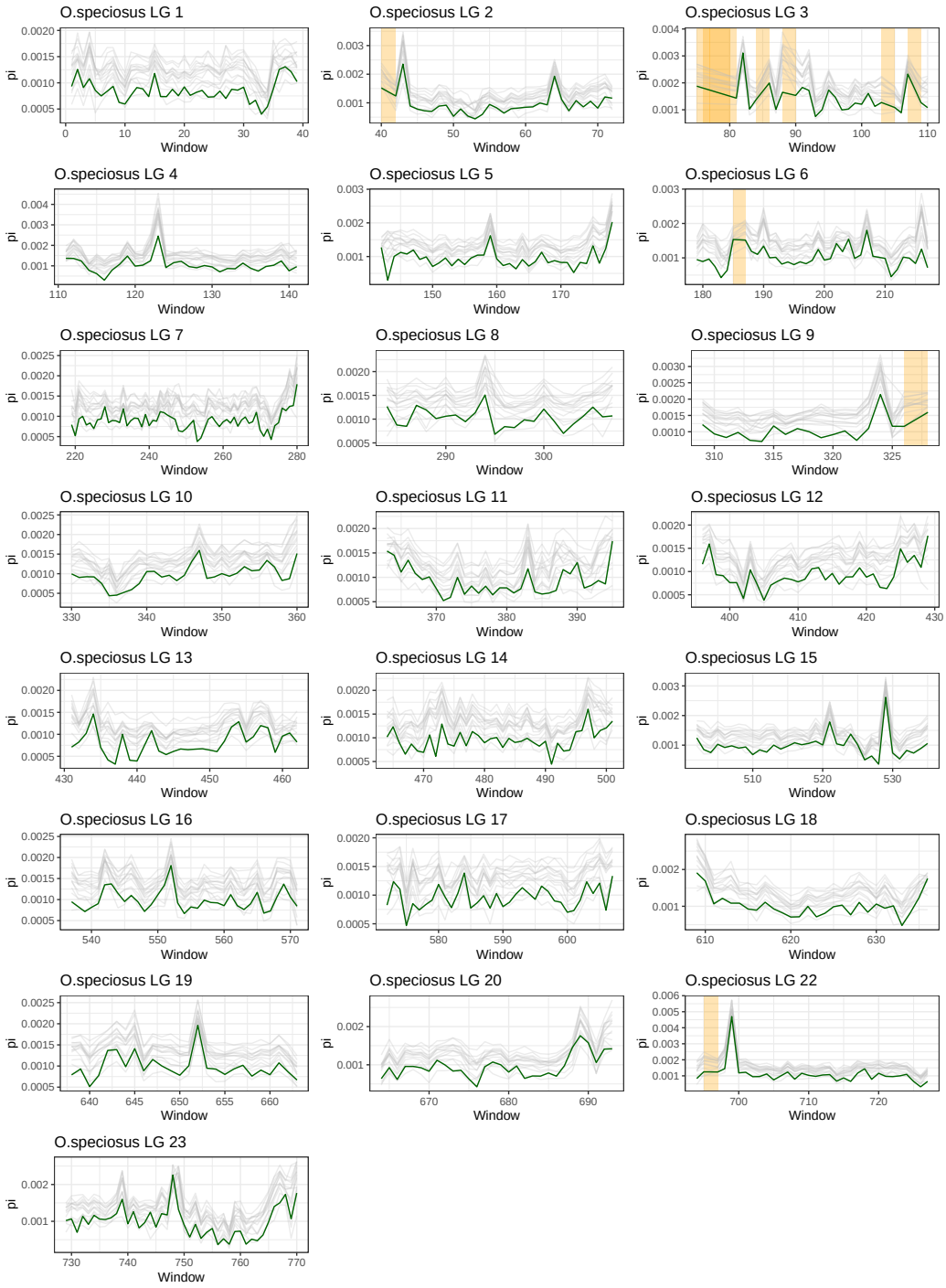

Figure D.12. 1Mb window π for all linkage groups, O.speciosus

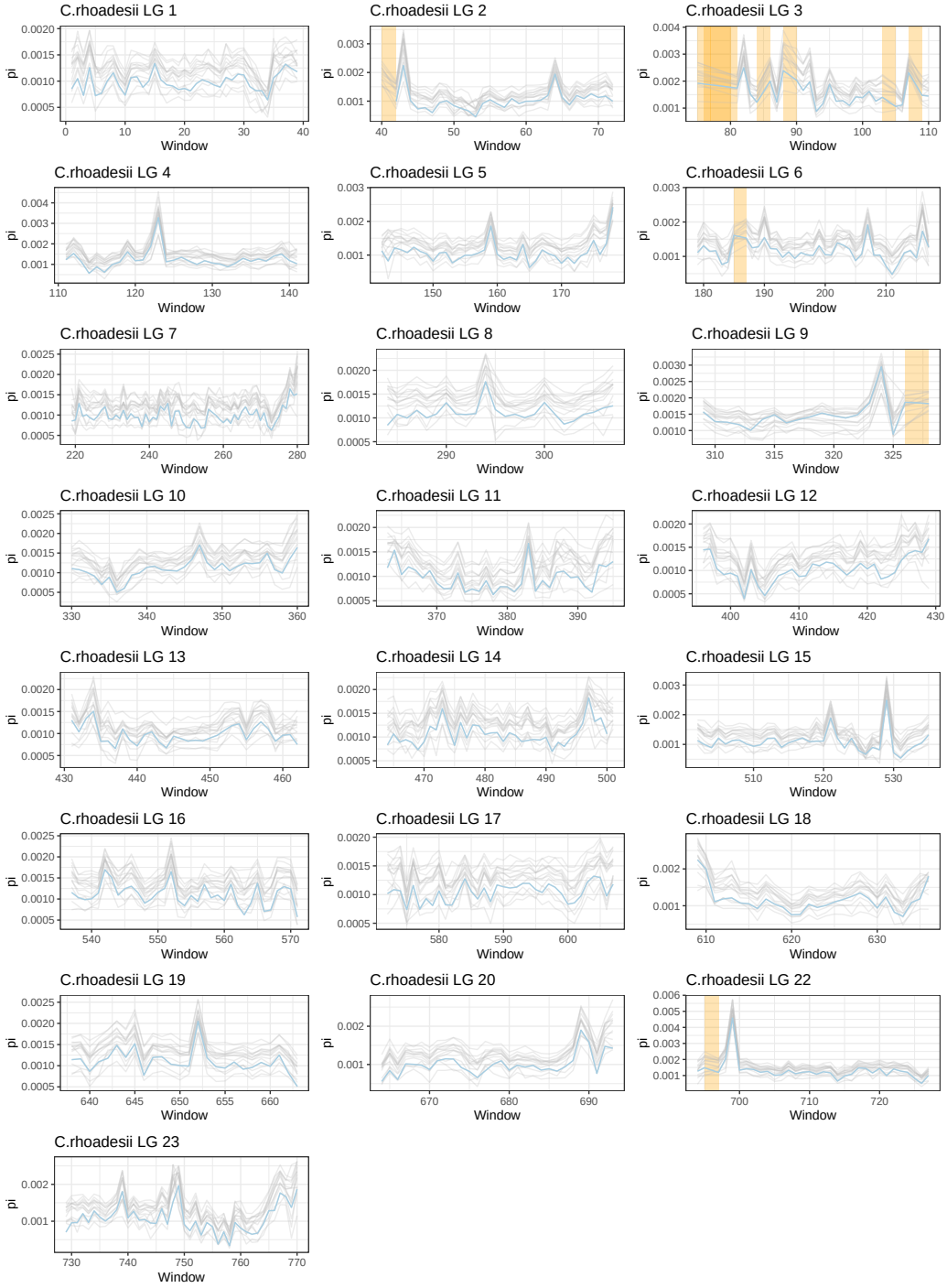

Figure D.13. 1Mb window π for all linkage groups, T.placodon

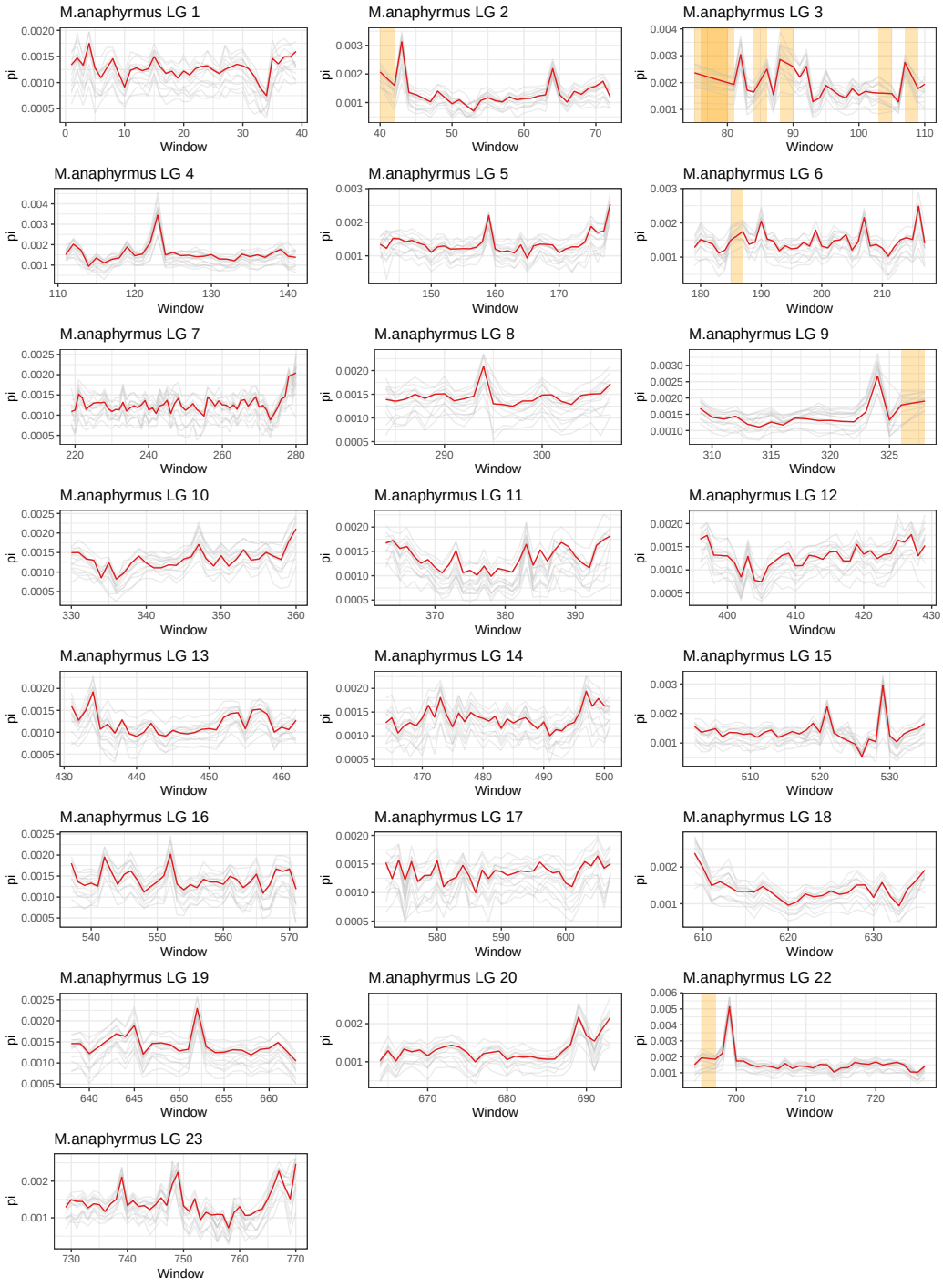

Figure D.14. 1Mb window π for all linkage groups, P.longimanus

APPENDIX E

CM BY BASE PAIR

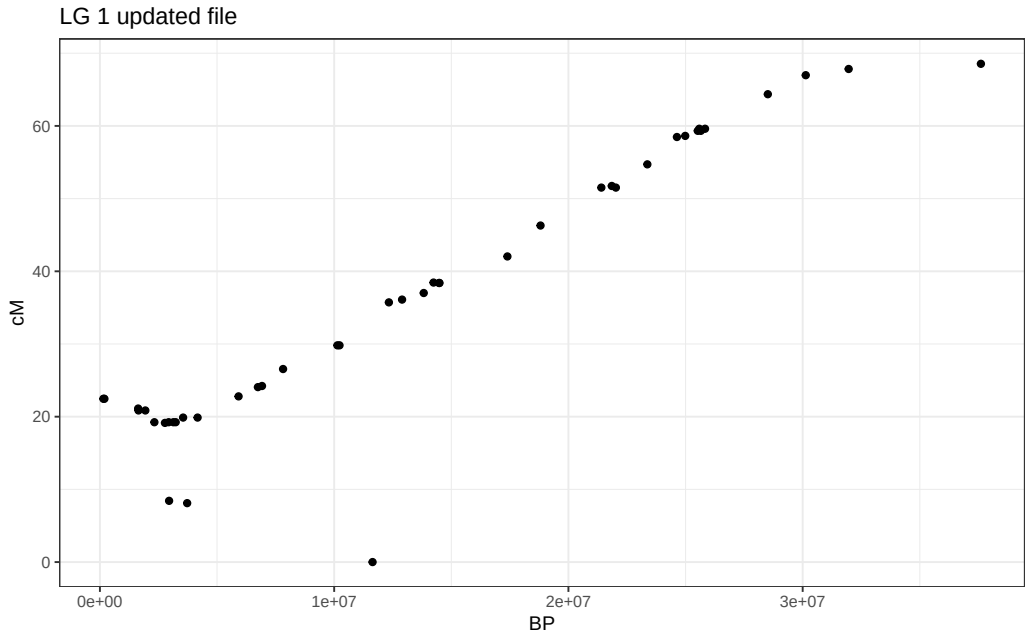

Figure E.1. cM by base pair linkage group one

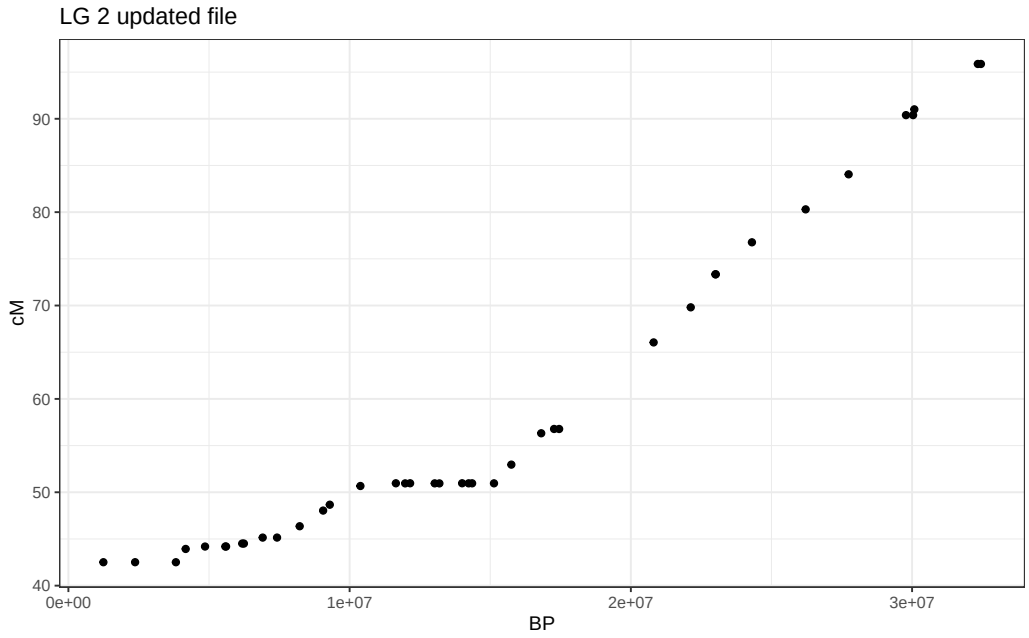

Figure E.2. cM by base pair linkage group two

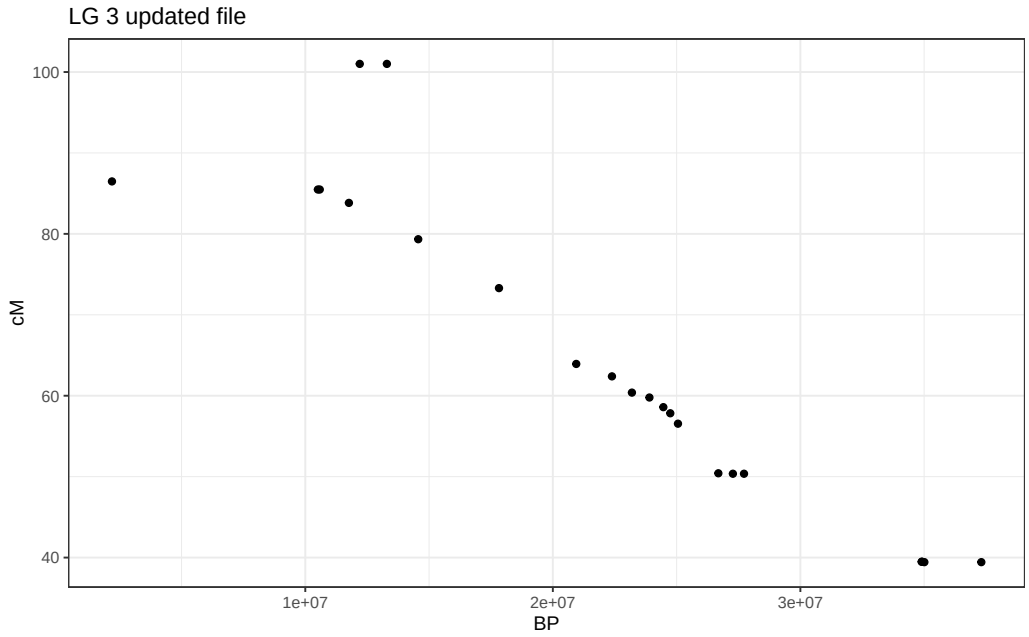

Figure E.3. cM by base pair linkage group three

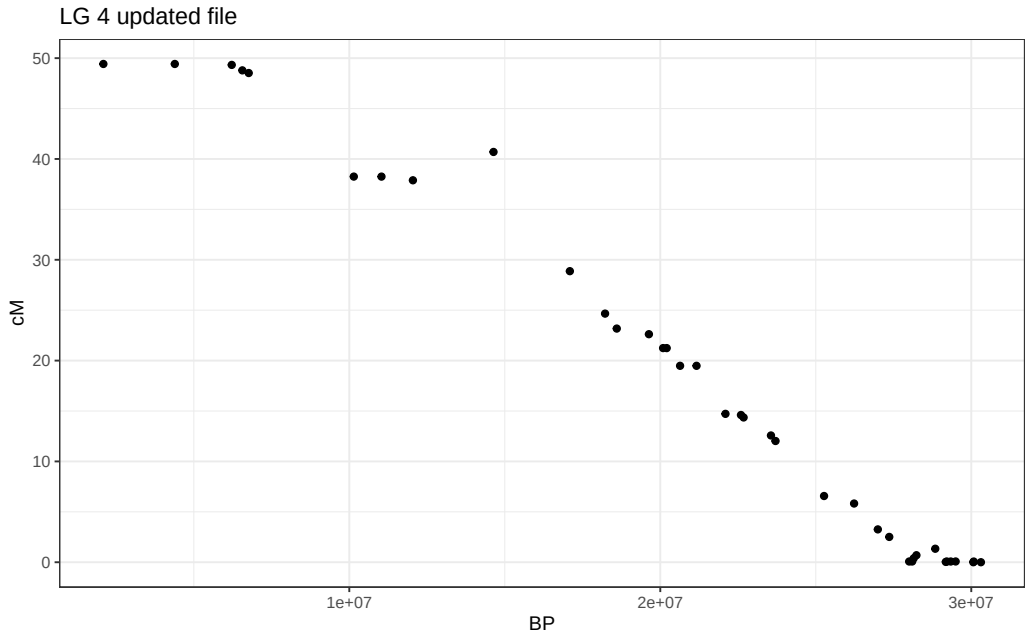

Figure E.4. cM by base pair linkage group four

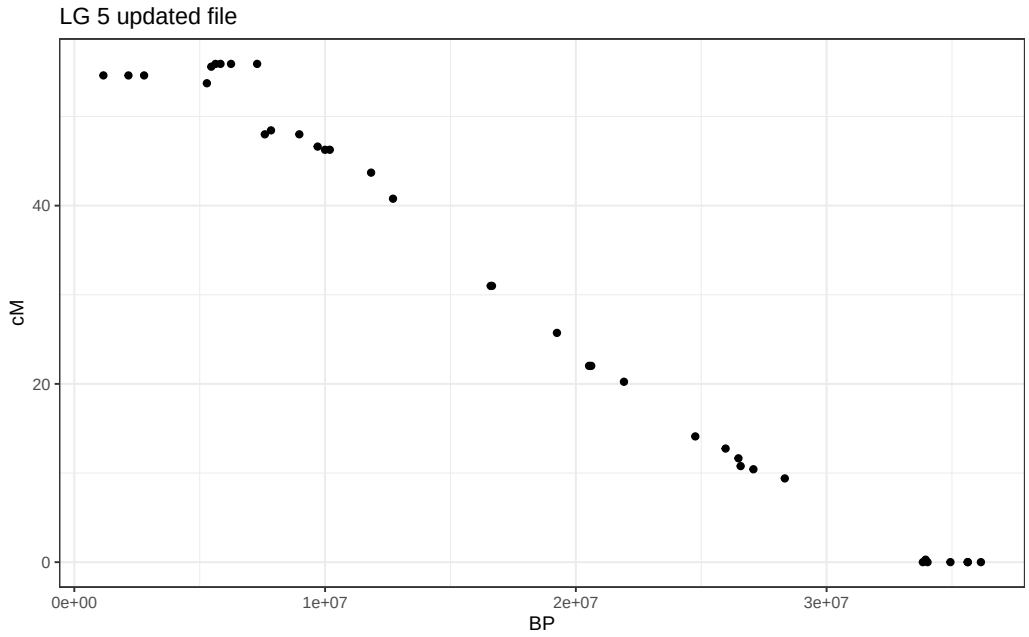

Figure E.5. cM by base pair linkage group five

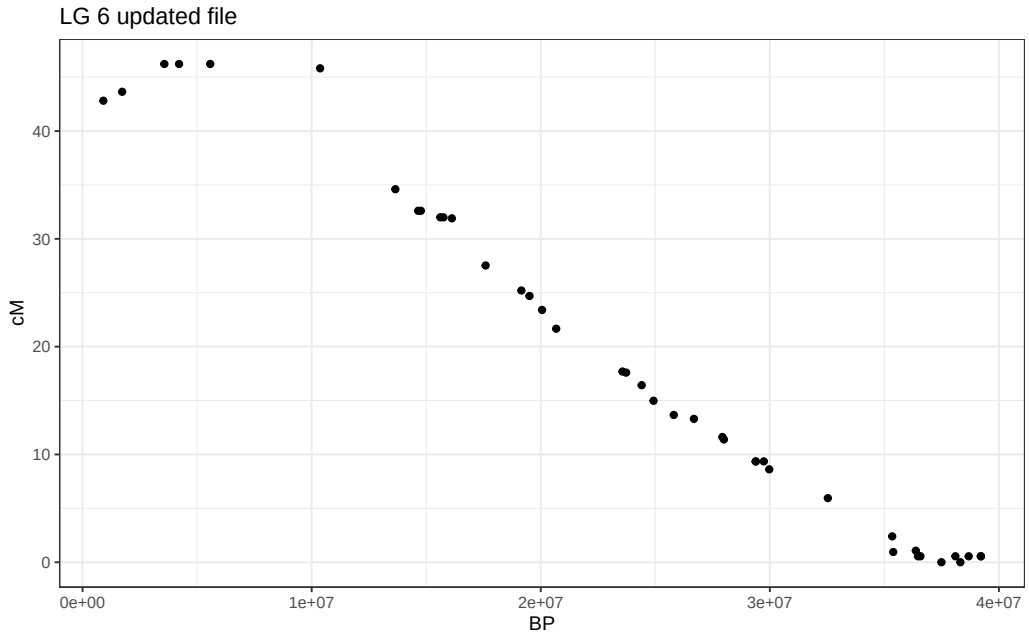

Figure E.6. cM by base pair linkage group six

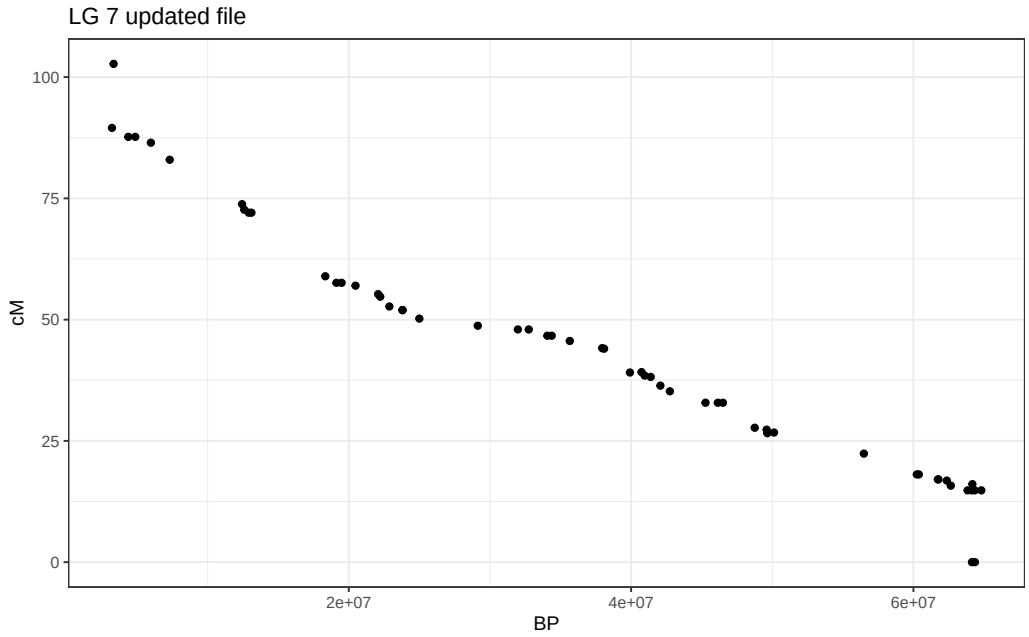

Figure E.7. cM by base pair linkage group seven

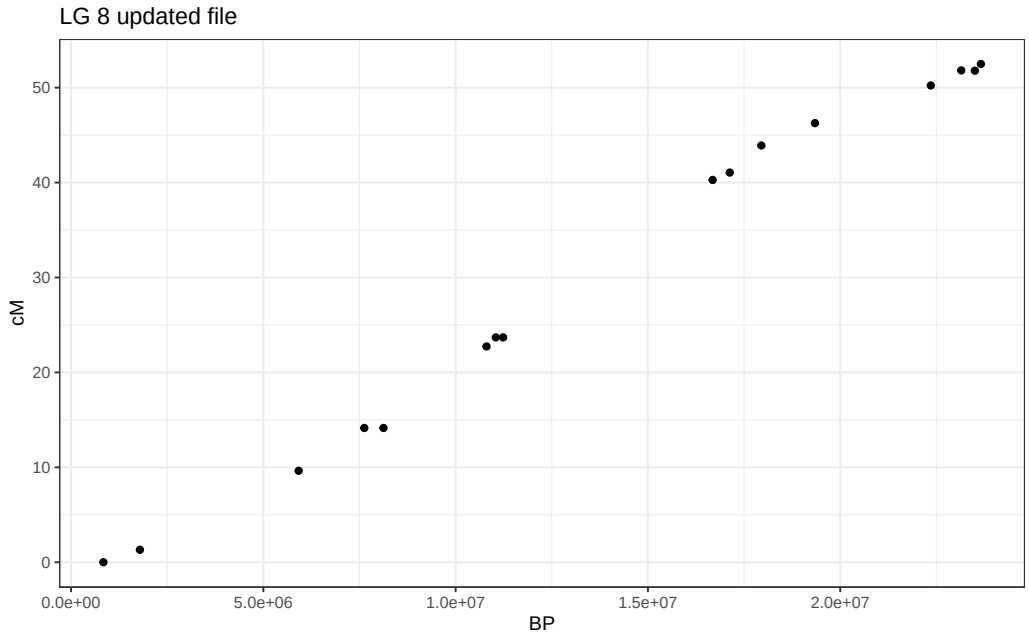

Figure E.8. cM by base pair linkage group eight

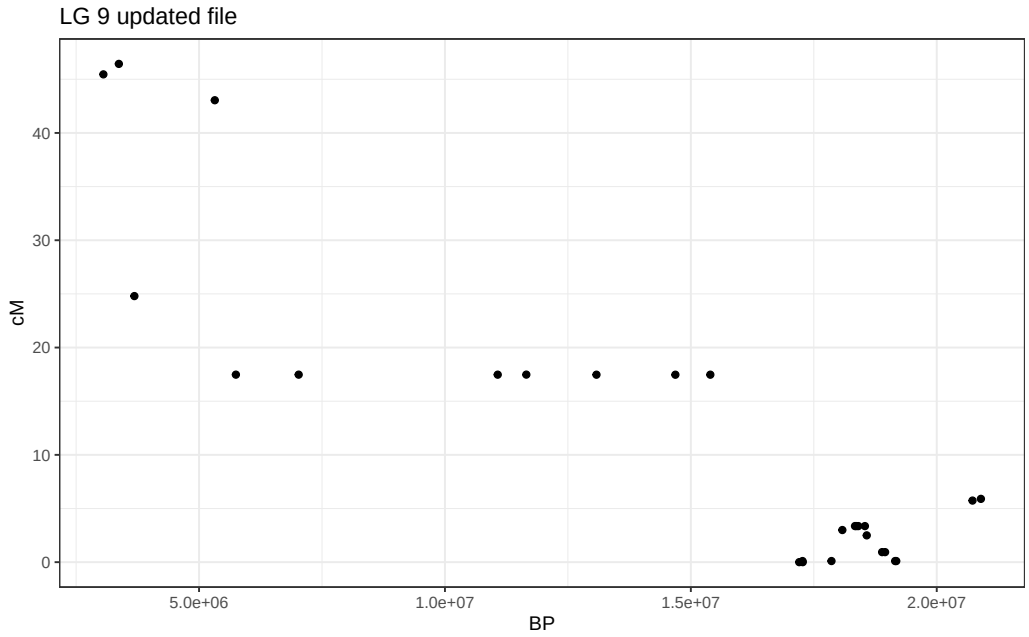

Figure E.9. cM by base pair linkage group nine

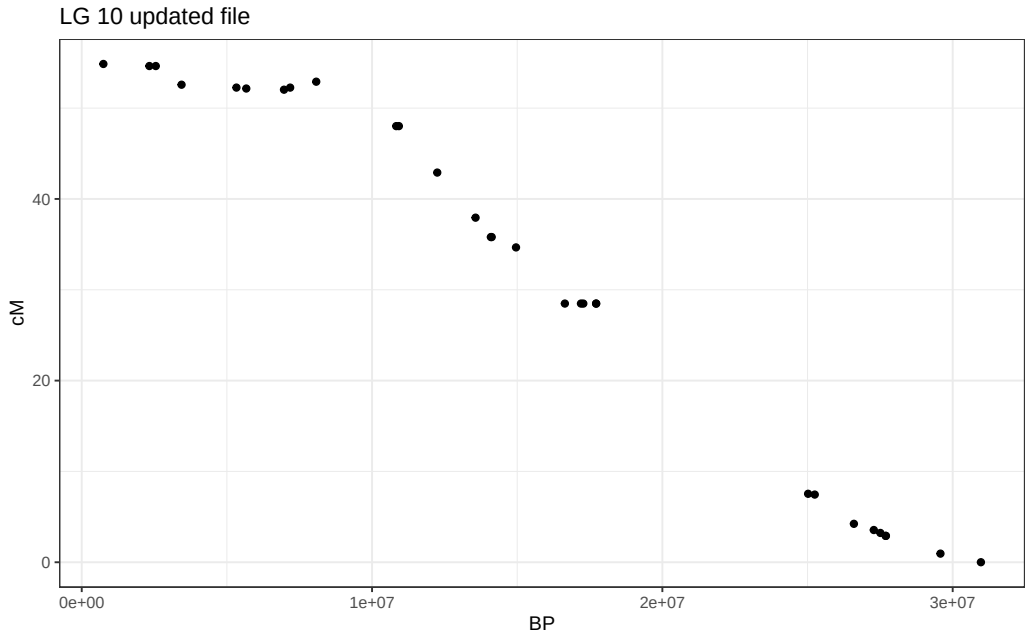

Figure E.10. cM by base pair linkage group ten

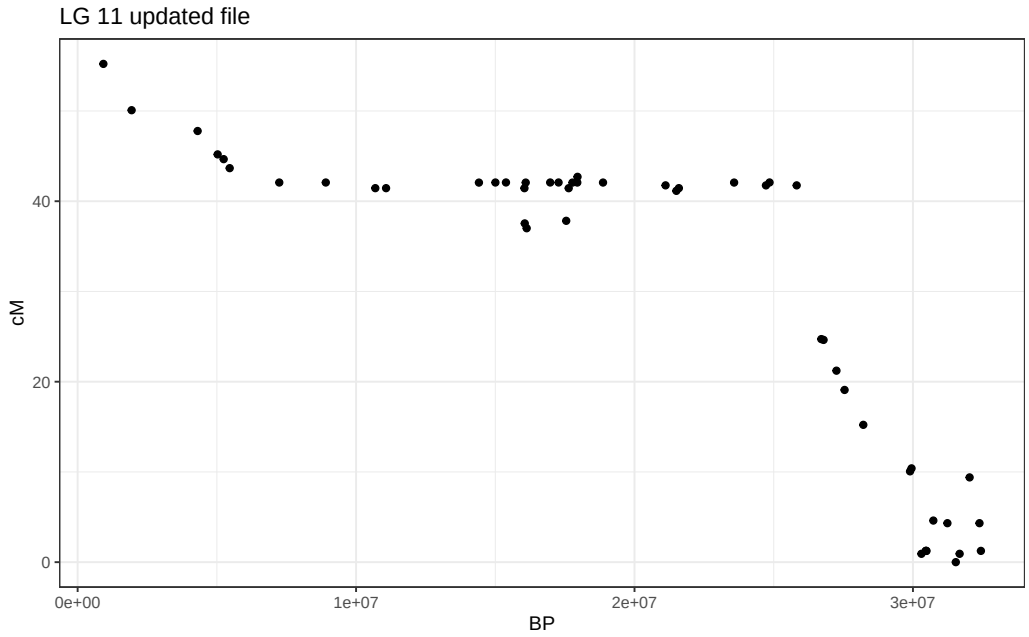

Figure E.11. cM by base pair linkage group eleven

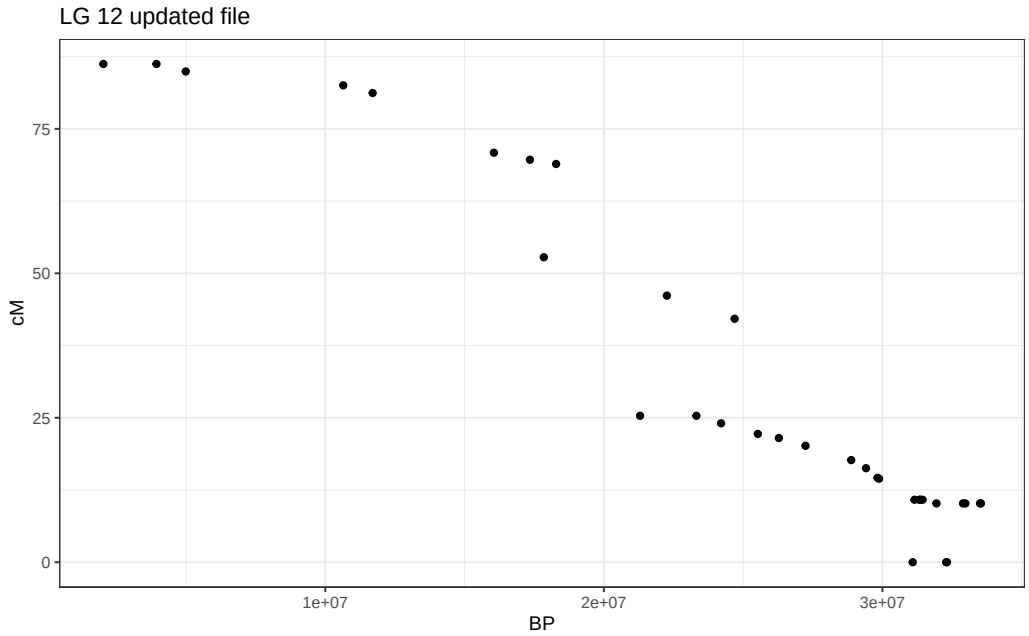

Figure E.12. cM by base pair linkage group twelve

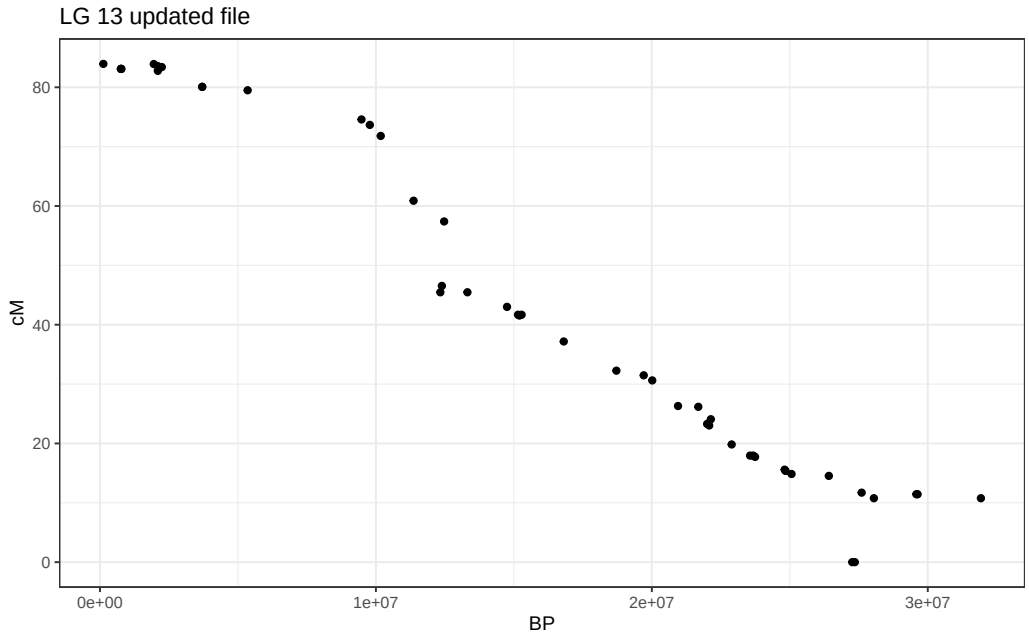

Figure E.13. cM by base pair linkage group thirteen

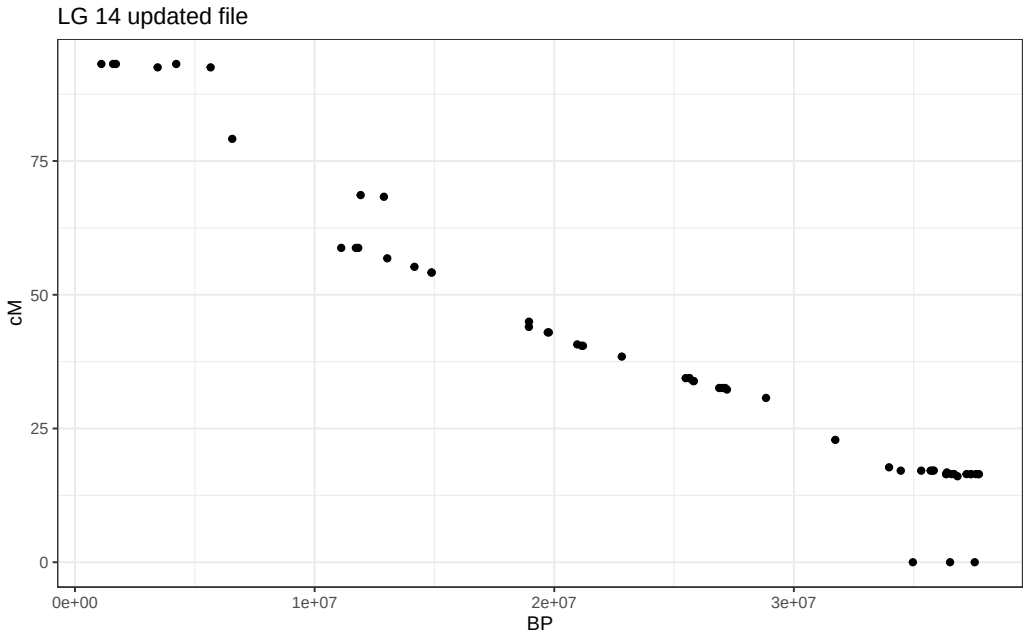

Figure E.14. cM by base pair linkage group fourteen

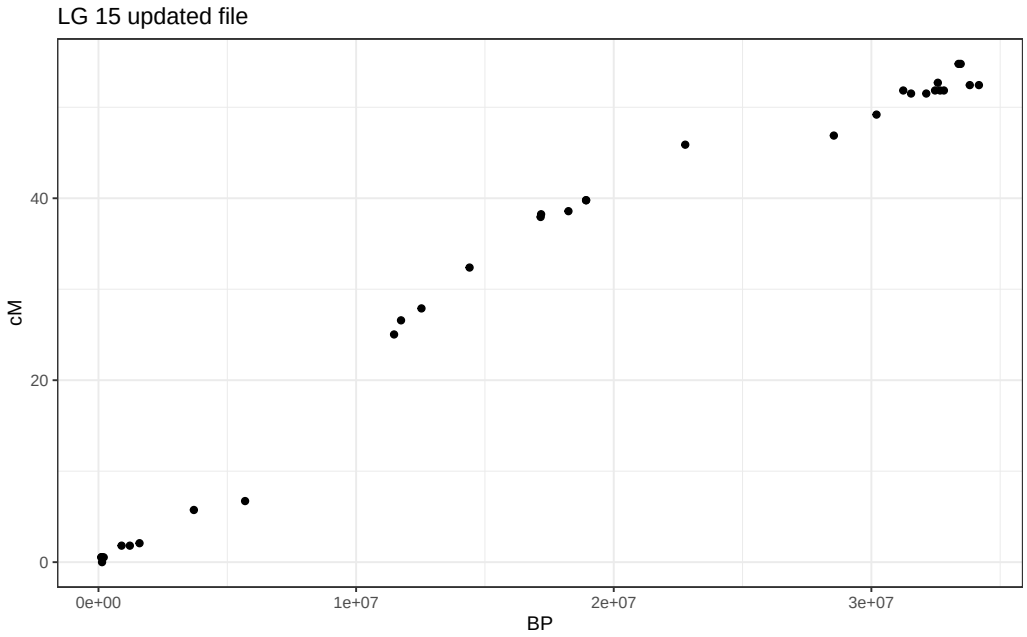

Figure E.15. cM by base pair linkage group fifteen

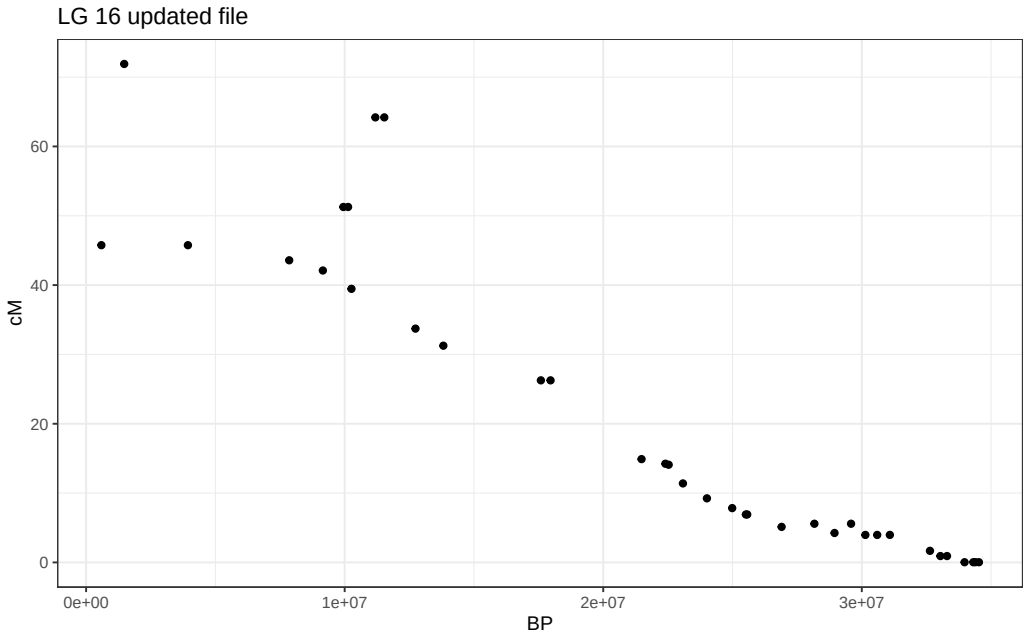

Figure E.16. cM by base pair linkage group sixteen

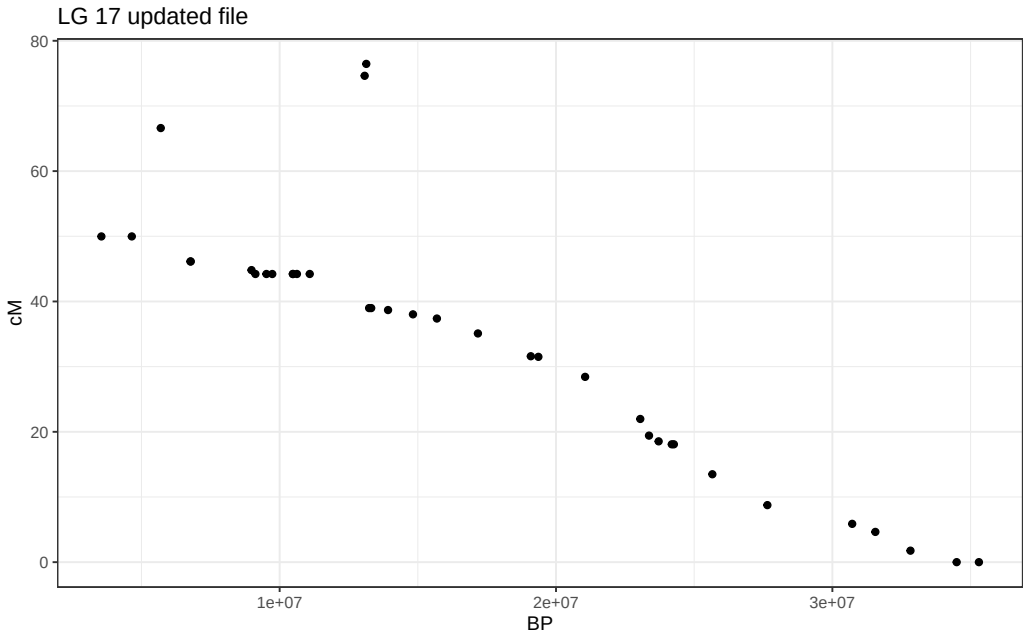

Figure E.17. cM by base pair linkage group seventeen

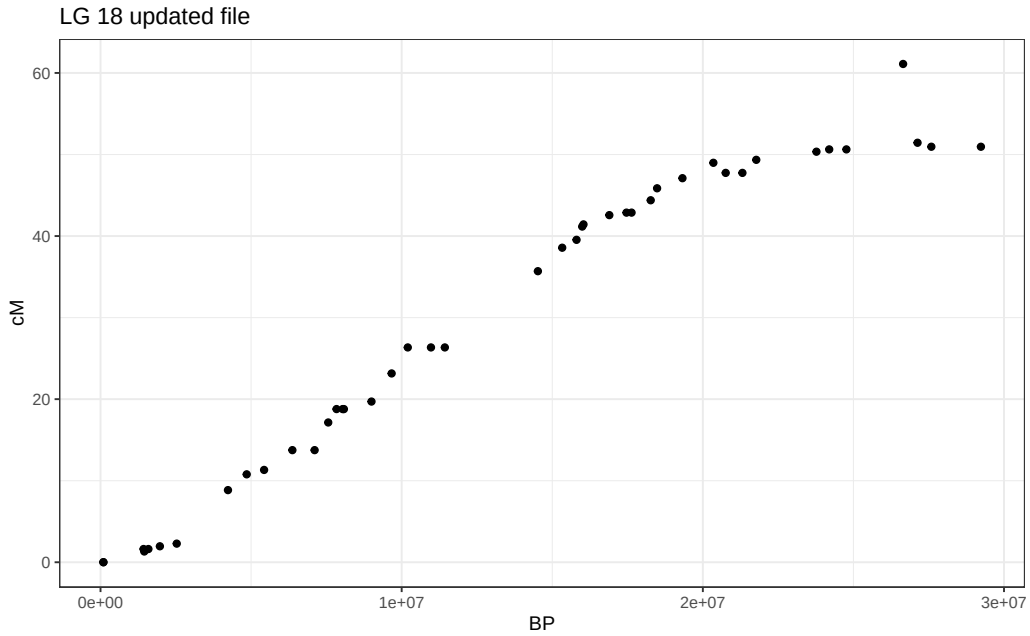

Figure E.18. cM by base pair linkage group eighteen

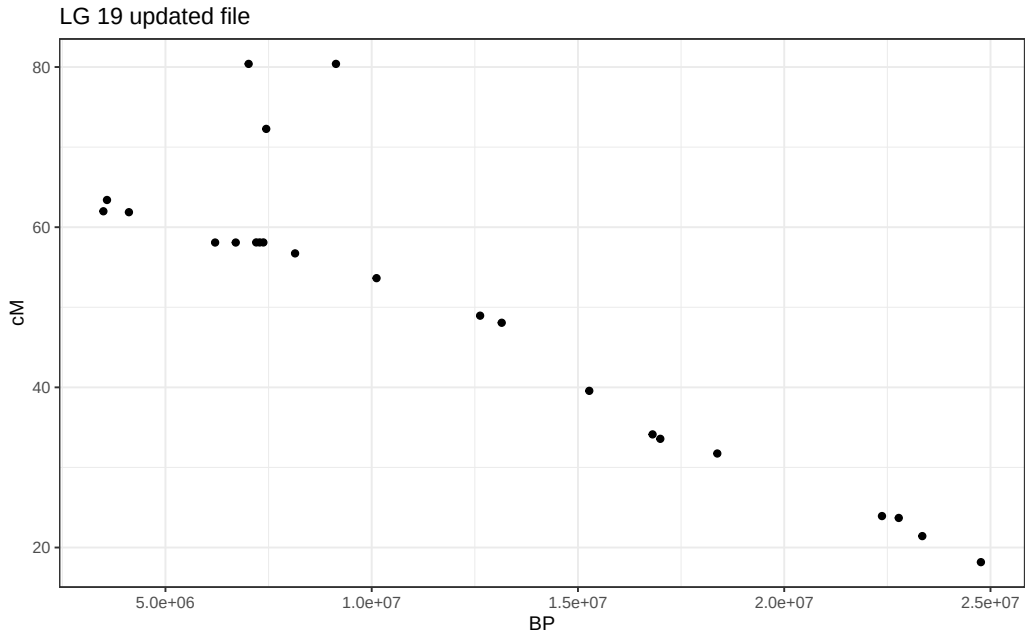

Figure E.19. cM by base pair linkage group nineteen

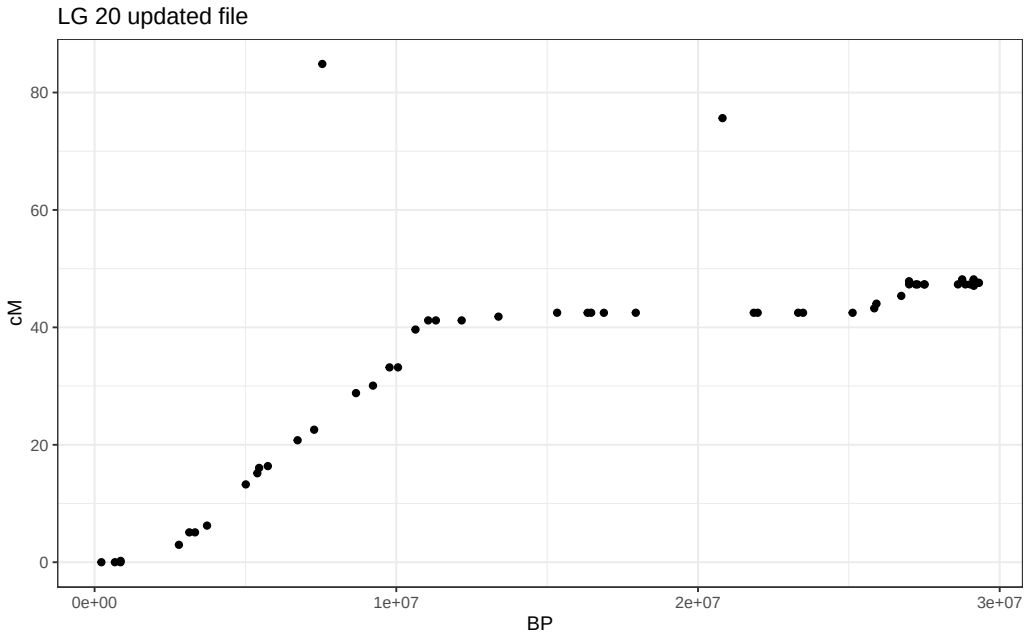

Figure E.20. cM by base pair linkage group twenty

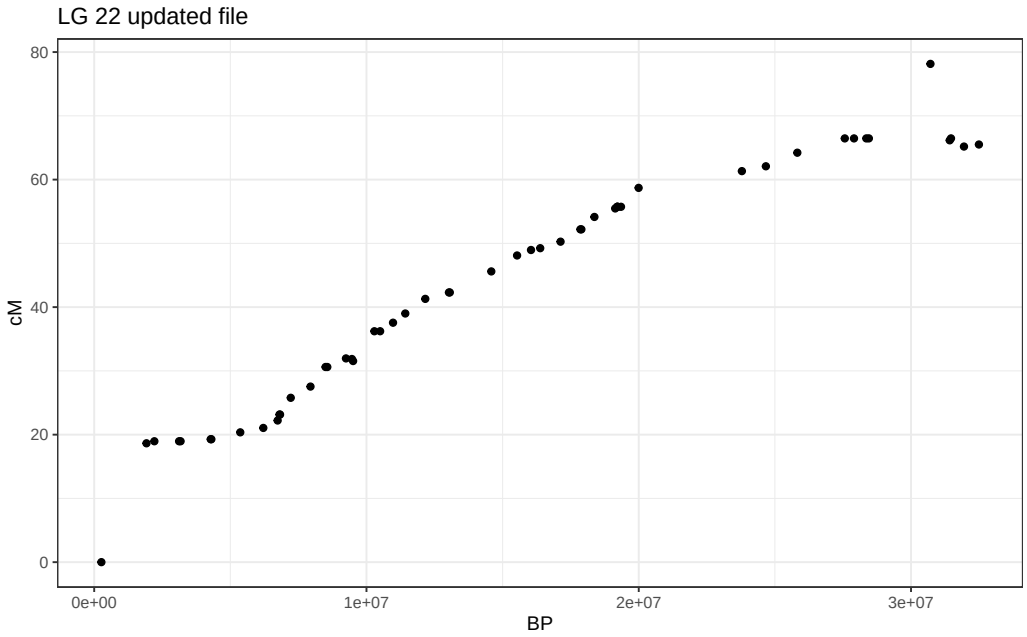

Figure E.21. cM by base pair linkage group twenty-two

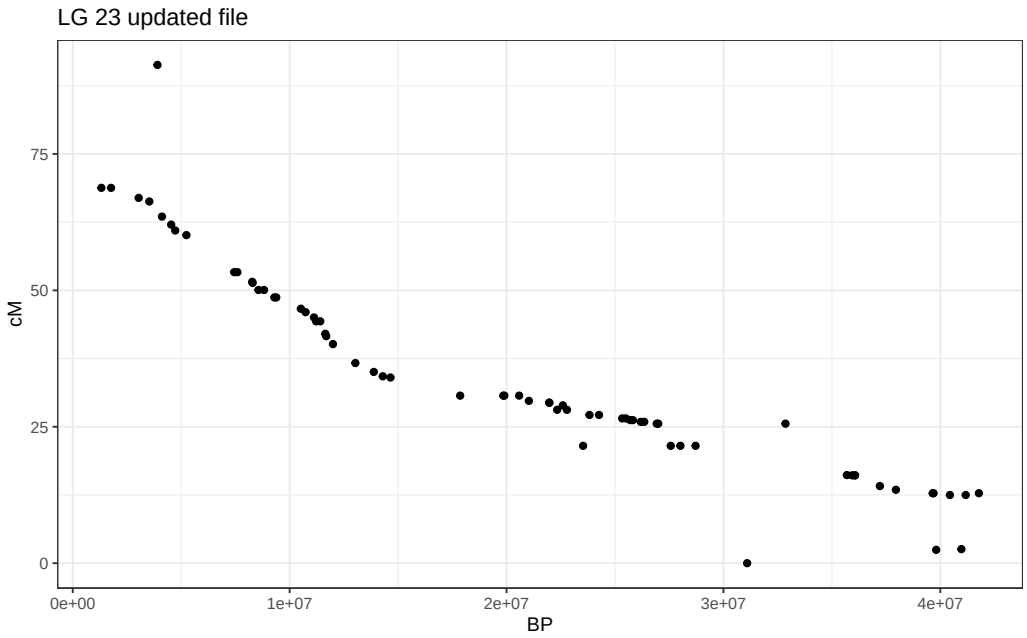

Figure E.22. cM by base pair linkage group twenty-three

APPENDIX F

CORRELATION AND COVARIANCE BY LINKAGE GROUPS

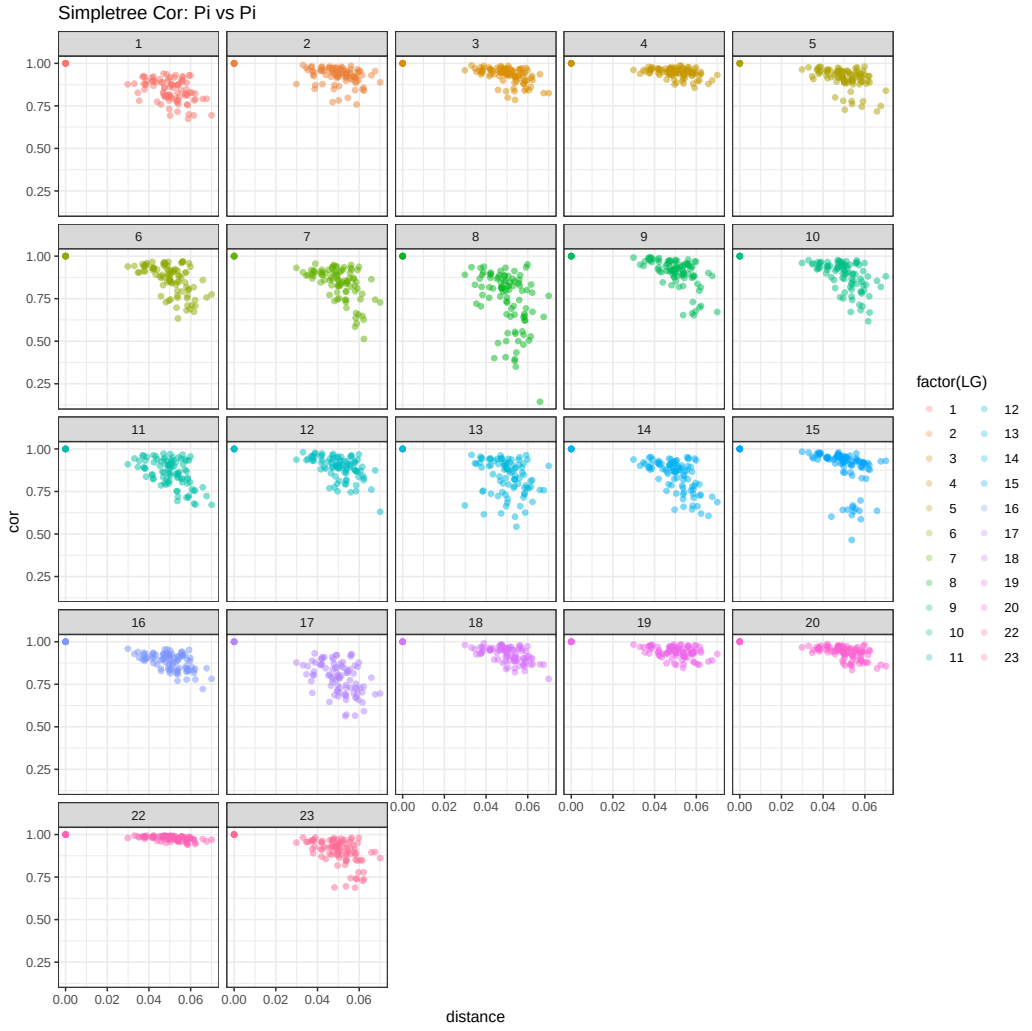

Figure F.1. Correlation between π and π in 1Mb widowed by phylogenetic time, all linkage groups

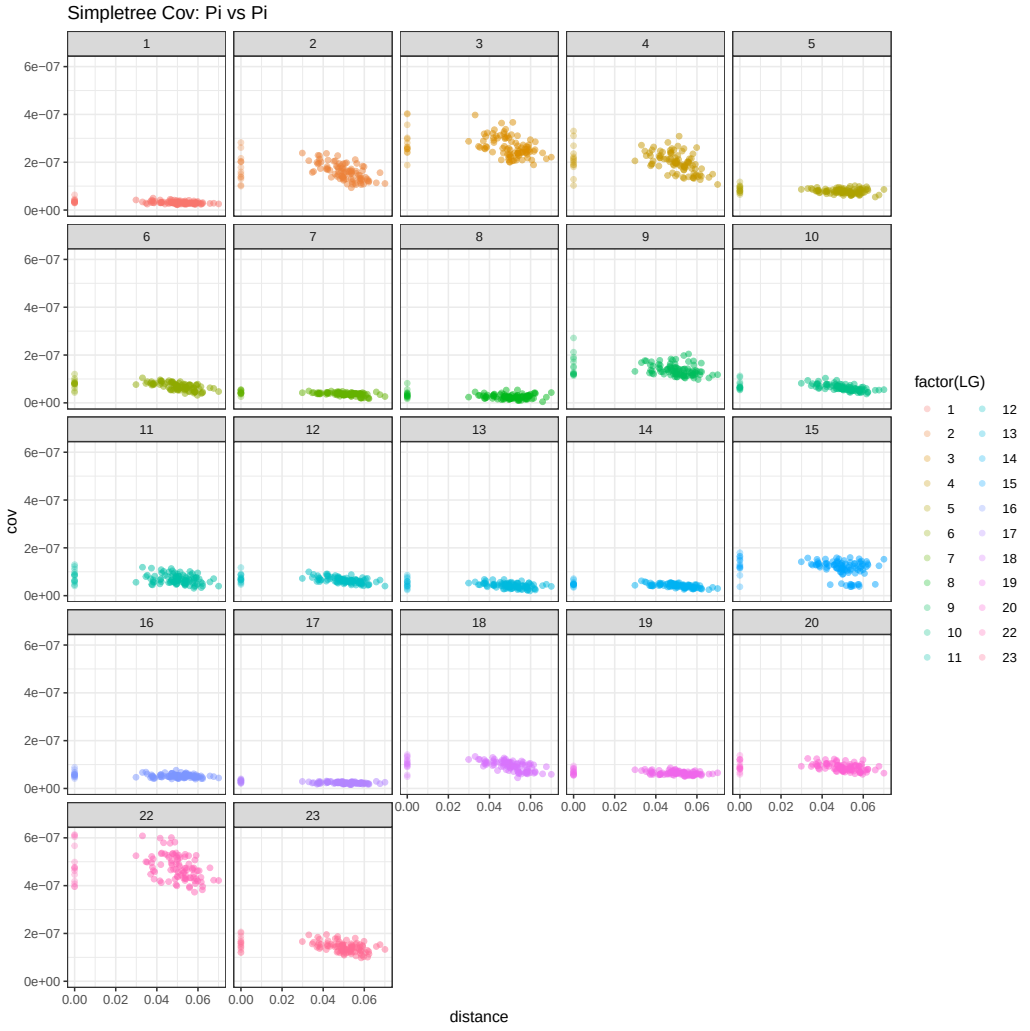

Figure F.2. Covariance between π and π in 1Mb widowed by phylogenetic time, all linkage groups

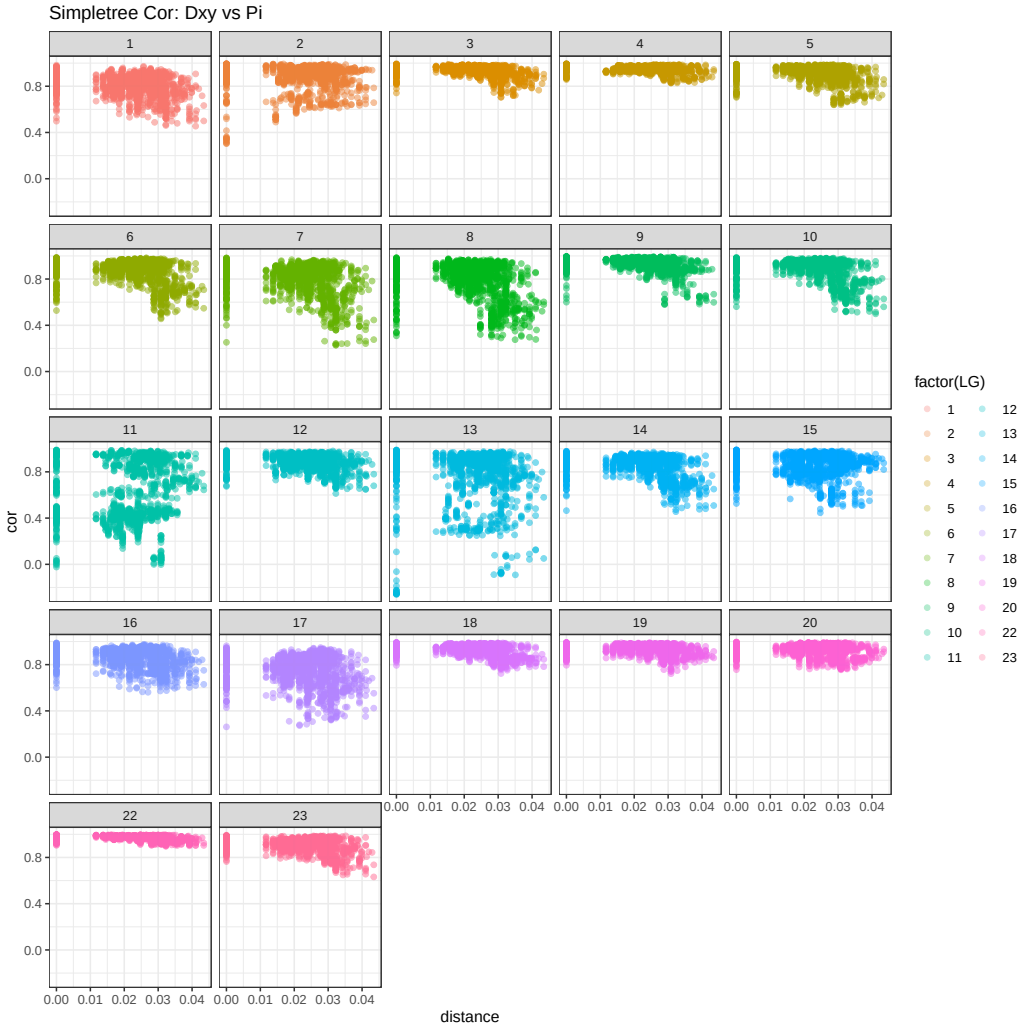

Figure F.3. Correlation between d_{XY} and π in 1Mb widowed by phylogenetic time, all linkage groups

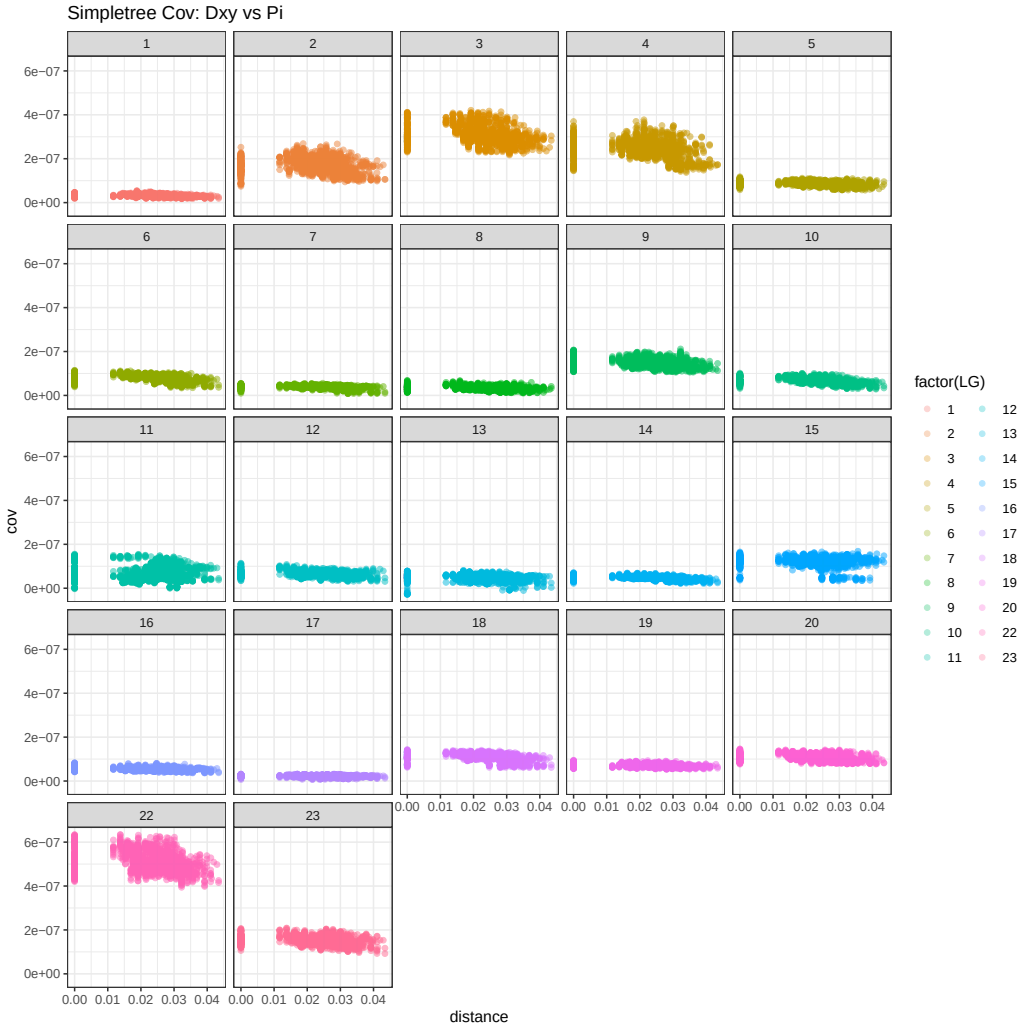

Figure F.4. Covariance between d_{XY} and π in 1Mb widowed by phylogenetic time, all linkage groups

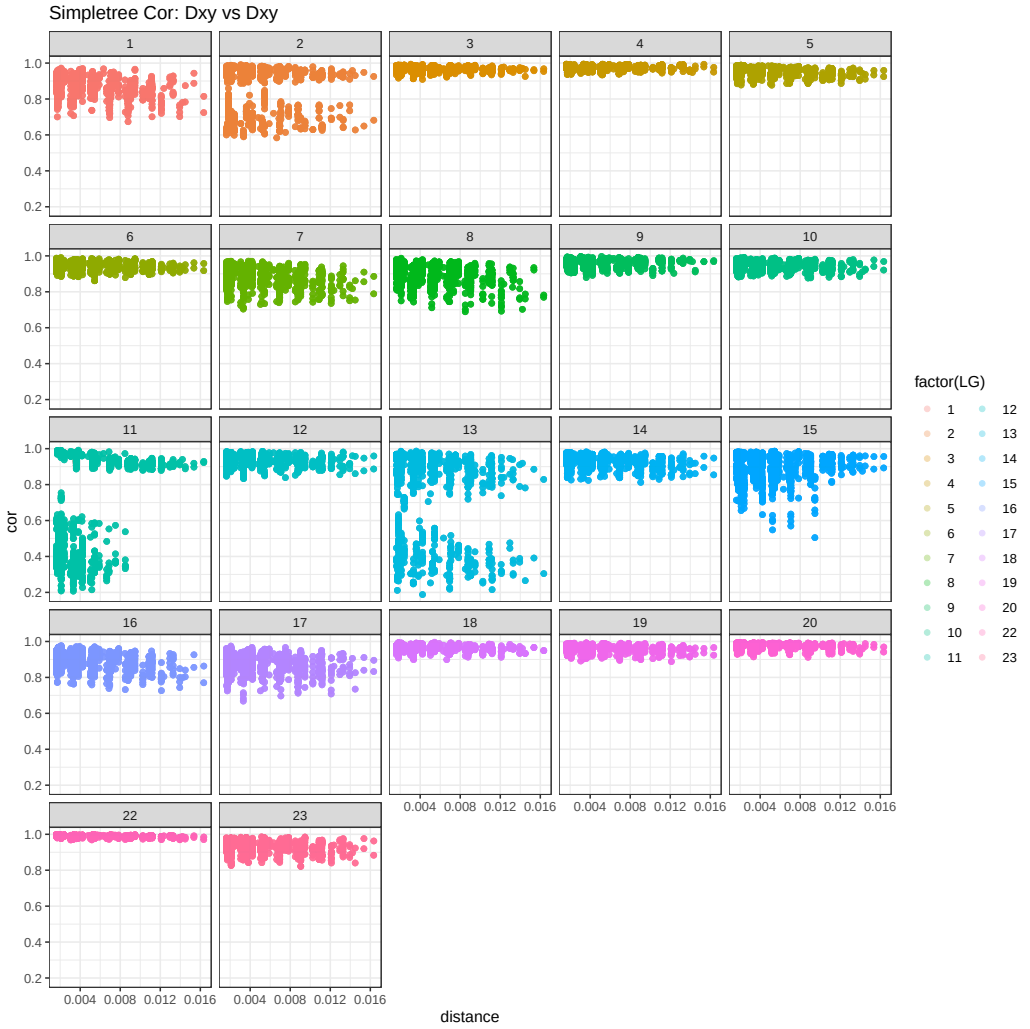

Figure F.5. Correlation between d_{XY} and d_{XY} in 1Mb widowed by phylogenetic time, all linkage groups

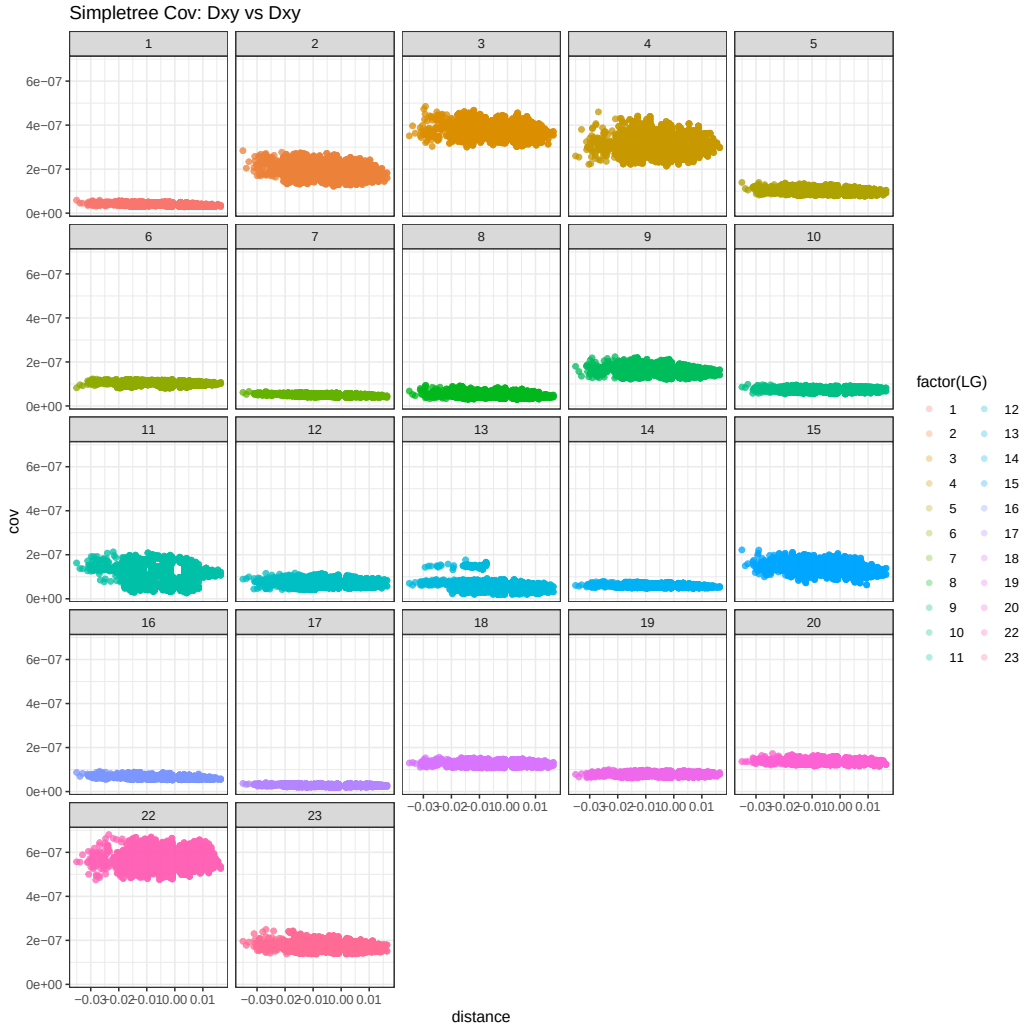

Figure F.6. Covariance between d_{XY} and d_{XY} in 1Mb widowed by phylogenetic time, all linkage groups

APPENDIX G

CORRELATION OF d_{XY} AND GENOME FUNCTIONS FOR ALL LINKAGE GROUPS

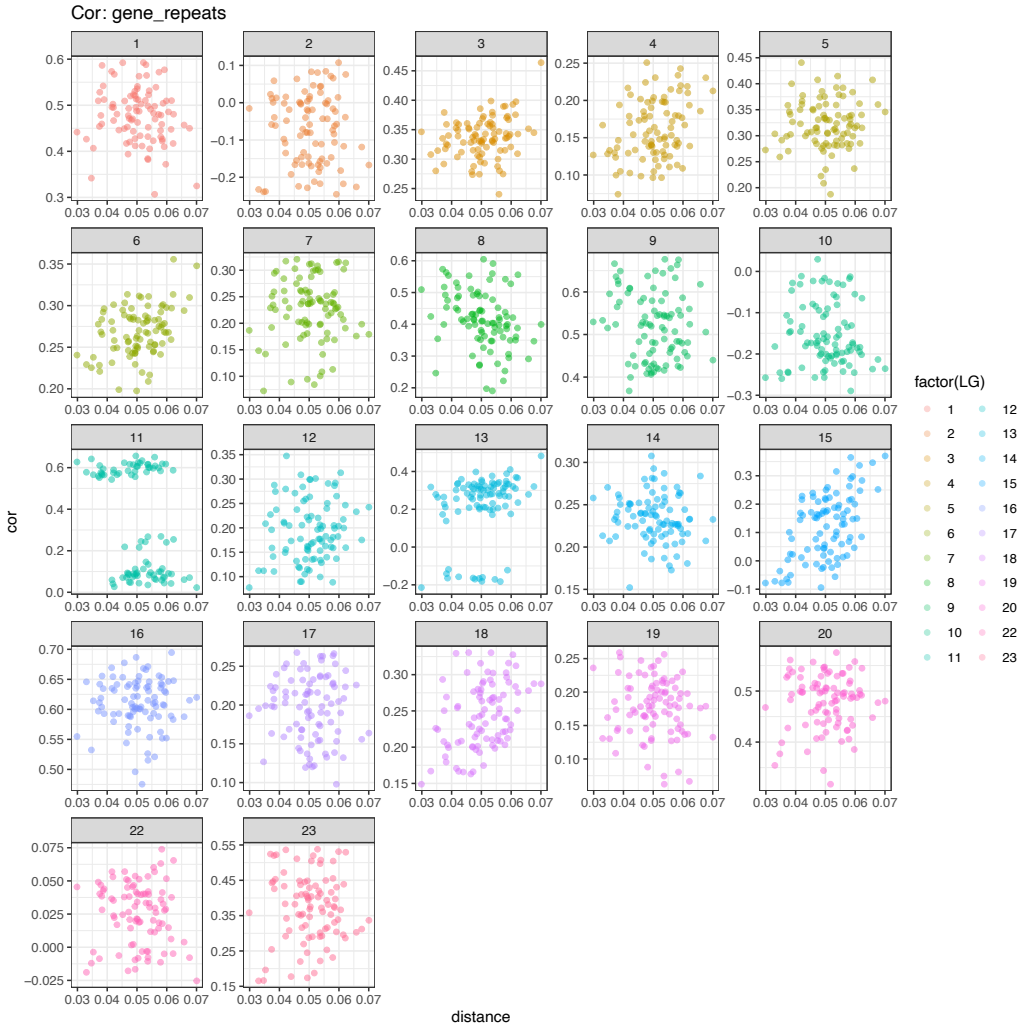

Figure G.1. Correlation between d_{XY} and gene repeats in 1Mb widowed by phylogenetic time, all linkage groups

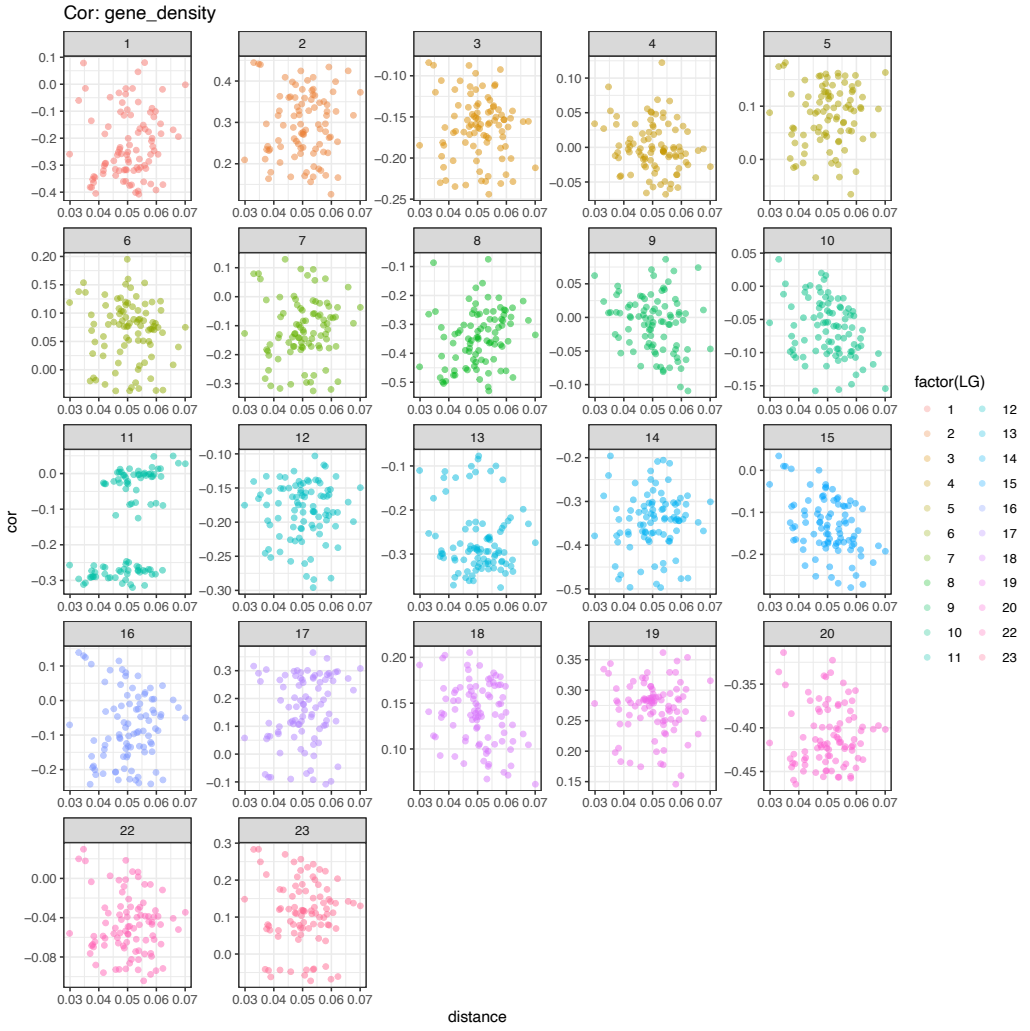

Figure G.2. Correlation between d_{XY} and gene density in 1Mb widowed by phylogenetic time, all linkage groups

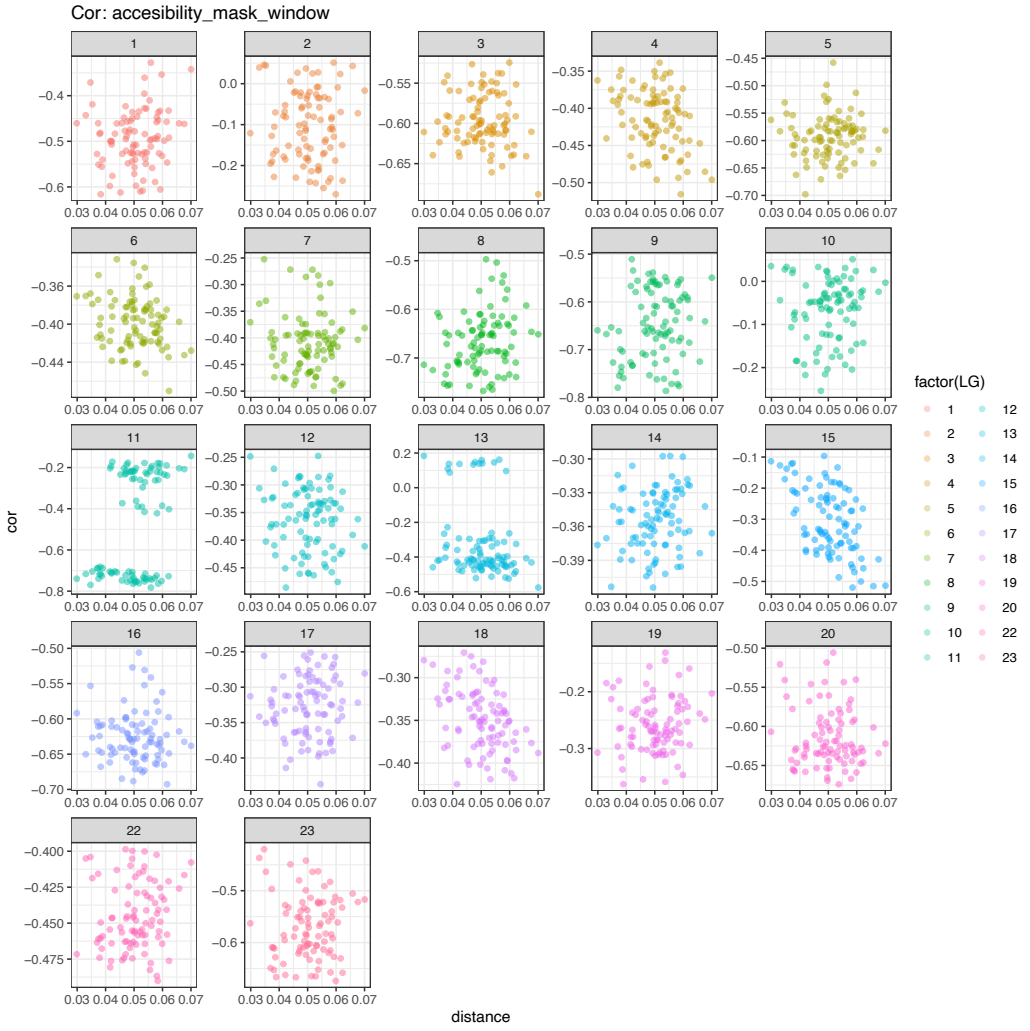

Figure G.3. Correlation between d_{XY} and accessibility in 1Mb widowed by phylogenetic time, all linkage groups

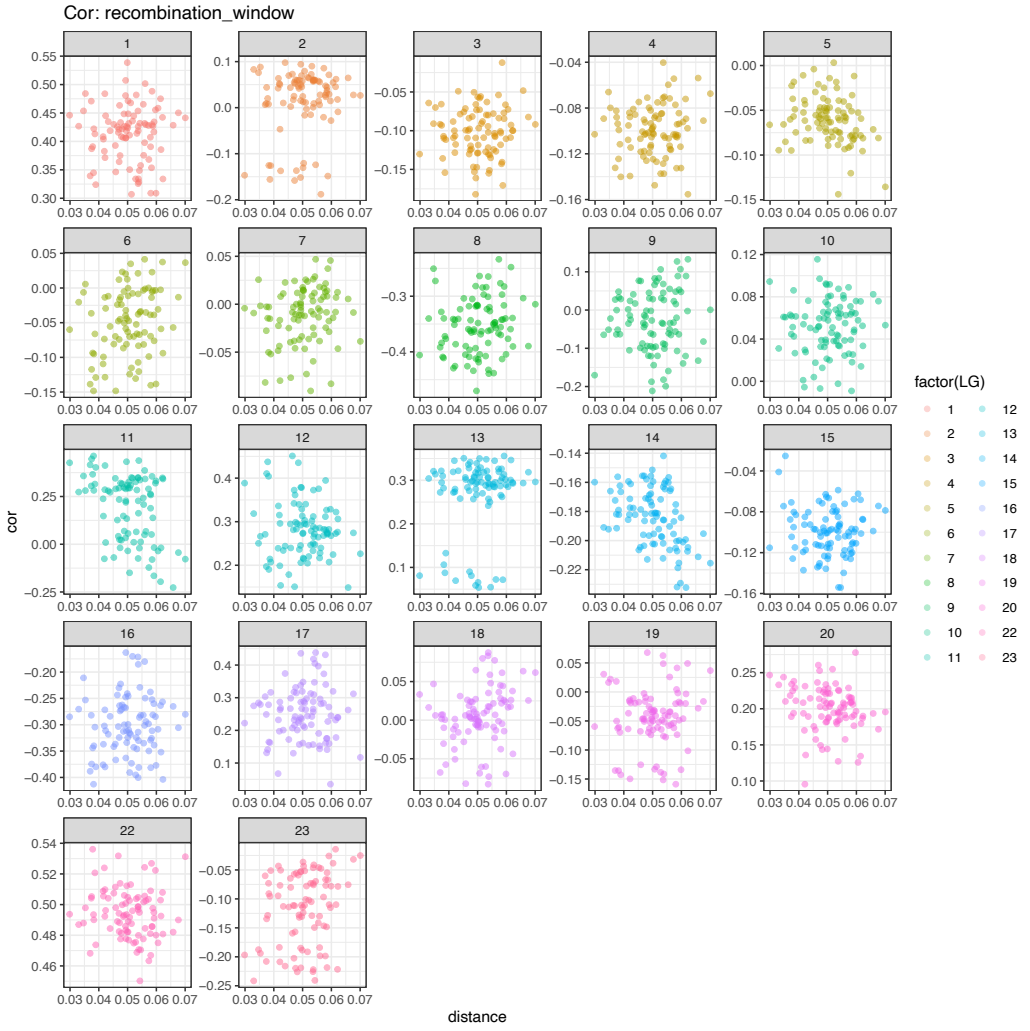

Figure G.4. Correlation between d_{XY} and recombination rate in 1Mb widowed by phylogenetic time, all linkage groups

APPENDIX H

PHYLOGENETIC TREES BY LINKAGE GROUP

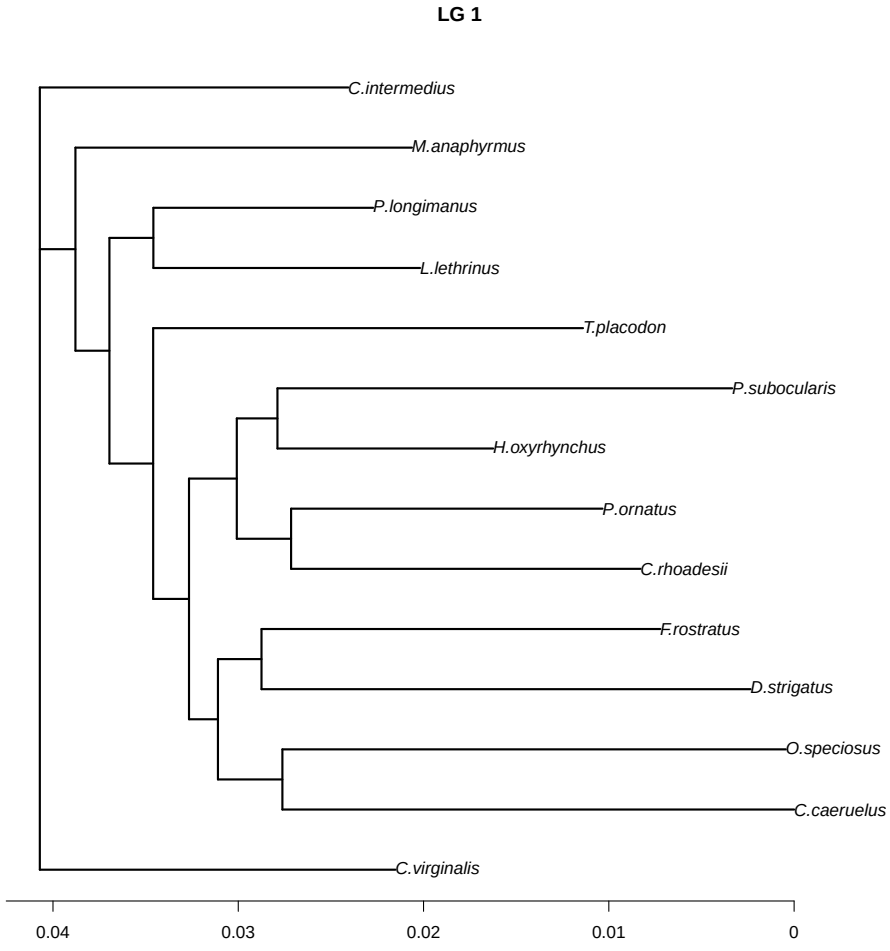

Figure H.1. Phylogenetic tree, linkage group 1

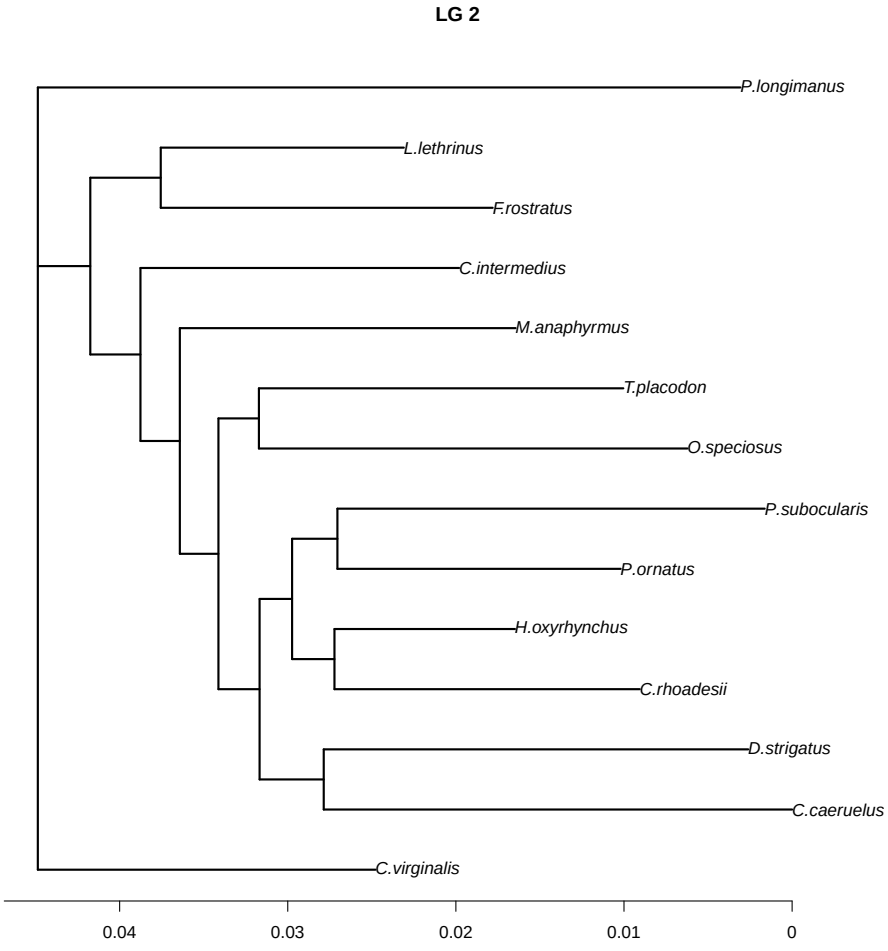

Figure H.2. Phylogenetic tree, linkage group 2

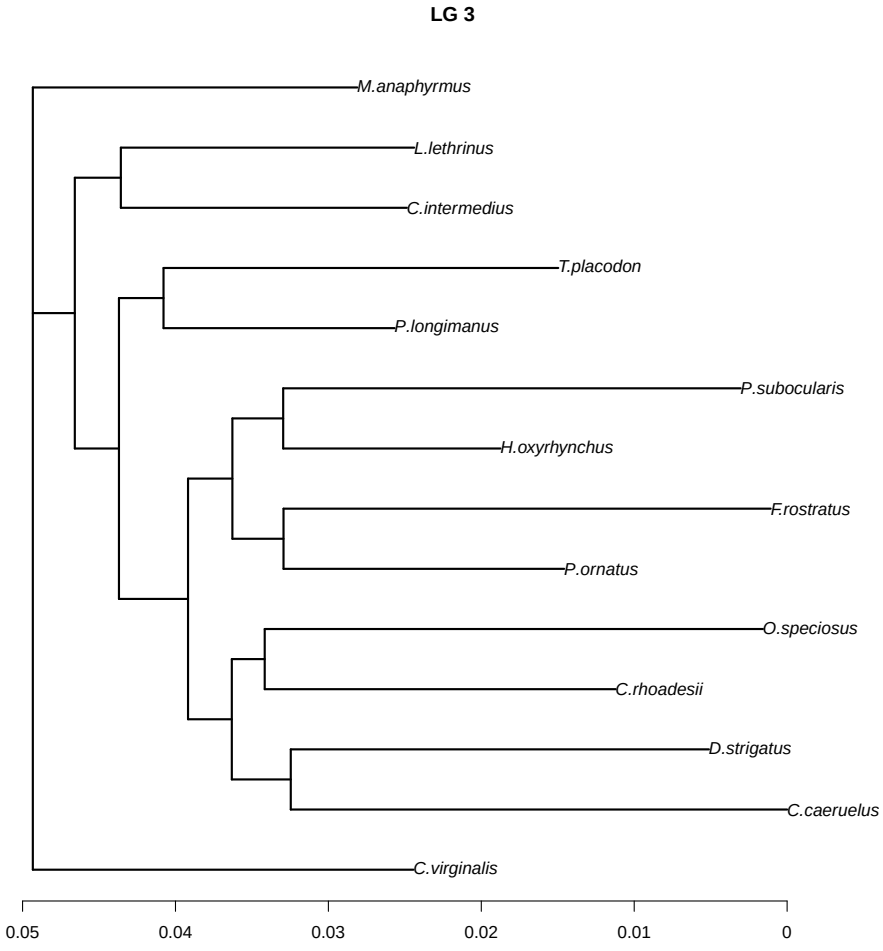

Figure H.3. Phylogenetic tree, linkage group 3

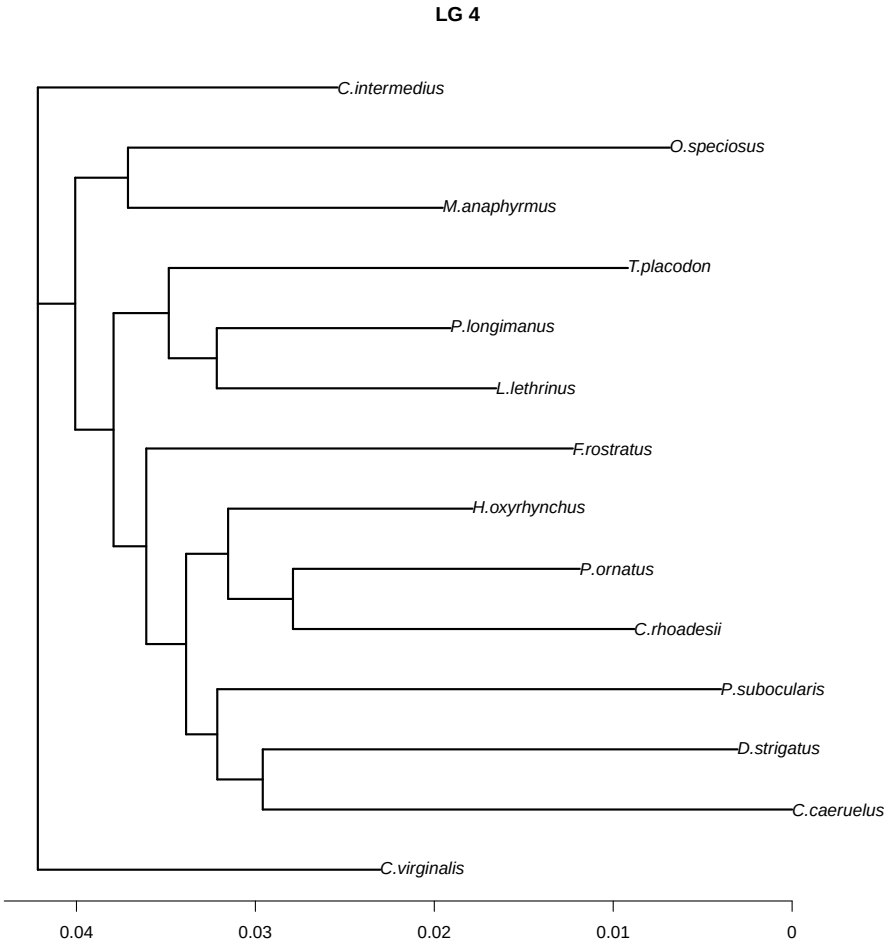

Figure H.4. Phylogenetic tree, linkage group 4

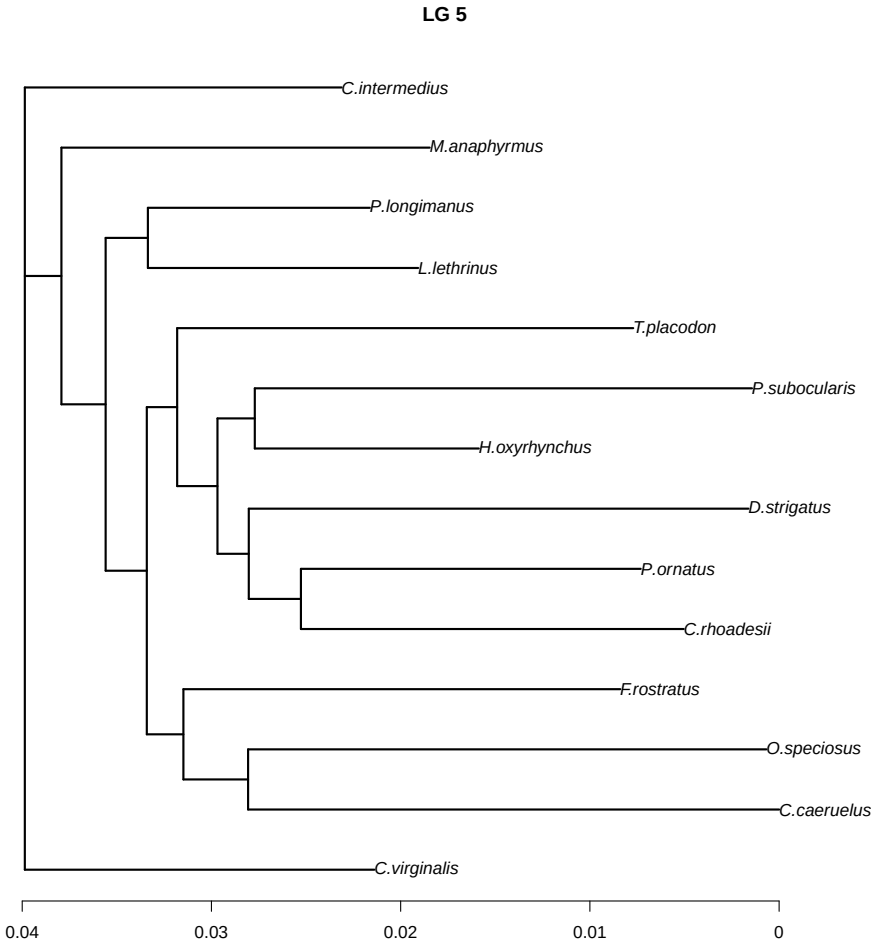

Figure H.5. Phylogenetic tree, linkage group 5

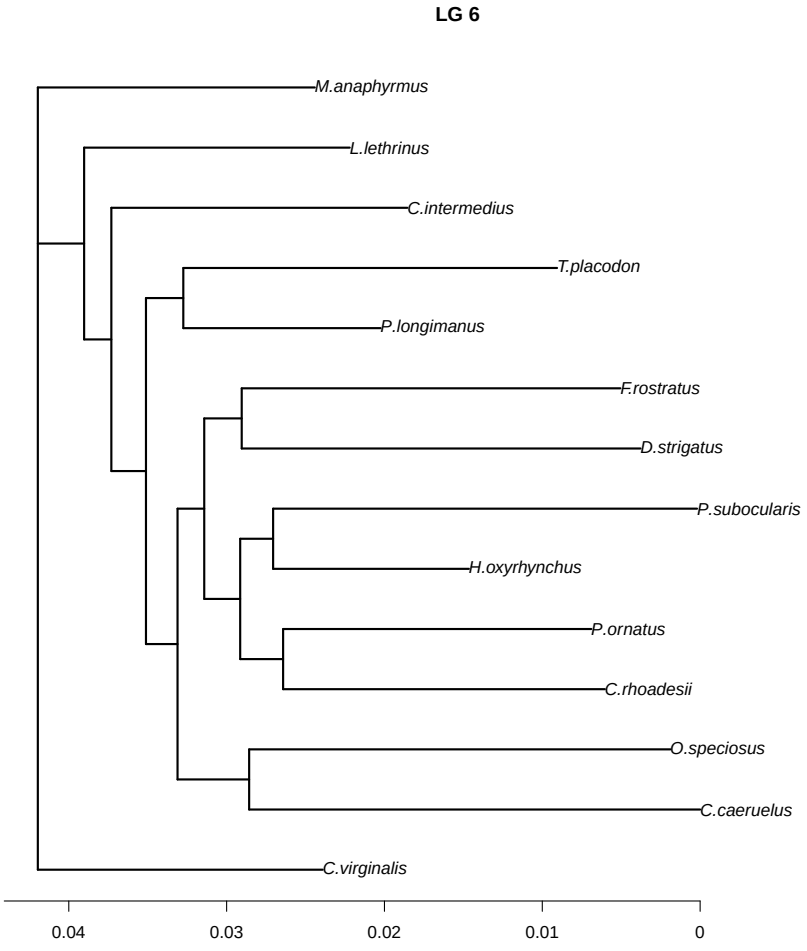

Figure H.6. Phylogenetic tree, linkage group 6

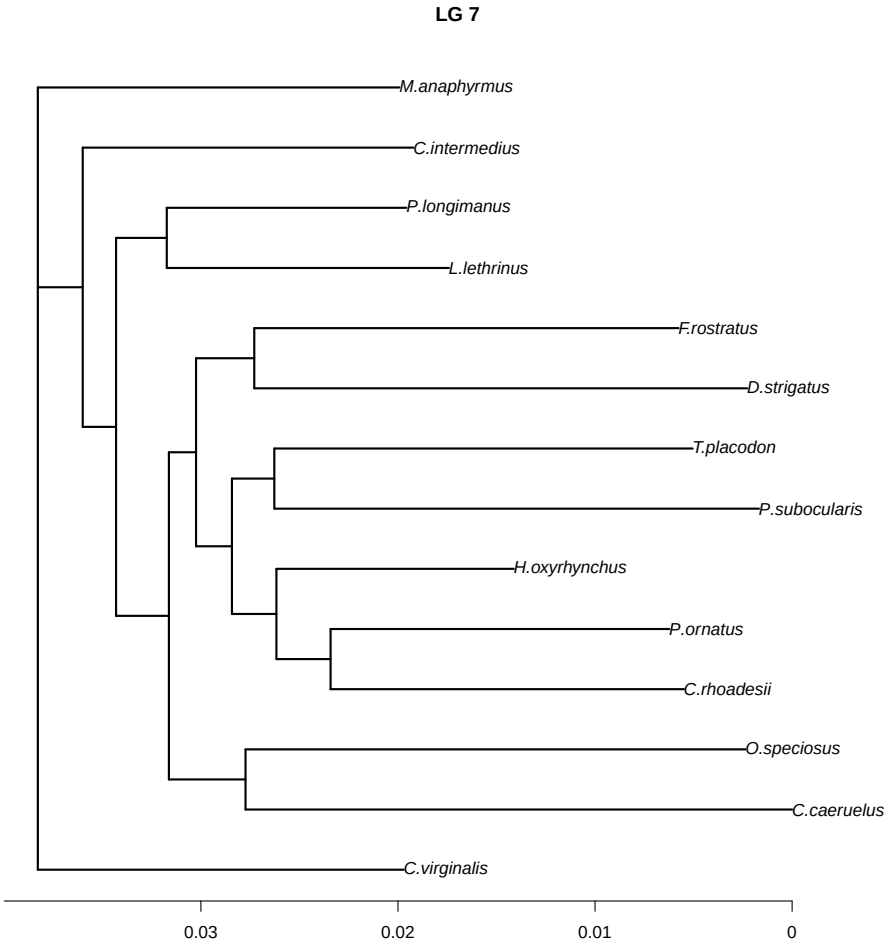

Figure H.7. Phylogenetic tree, linkage group 7

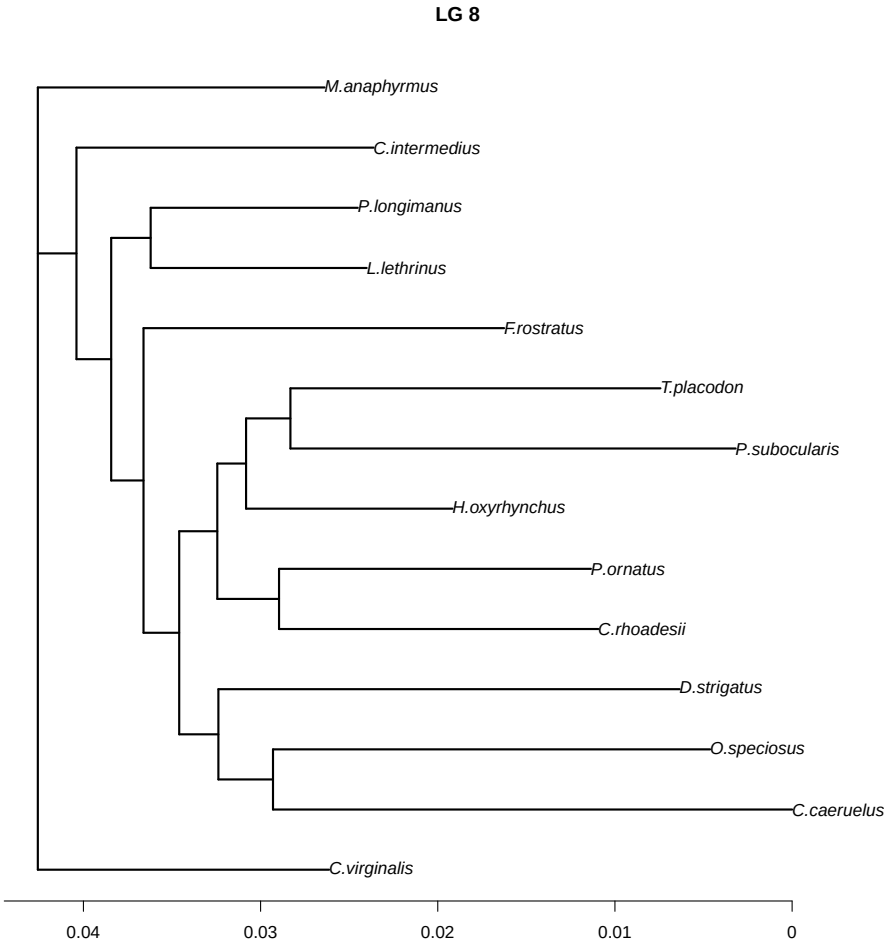

Figure H.8. Phylogenetic tree, linkage group 8

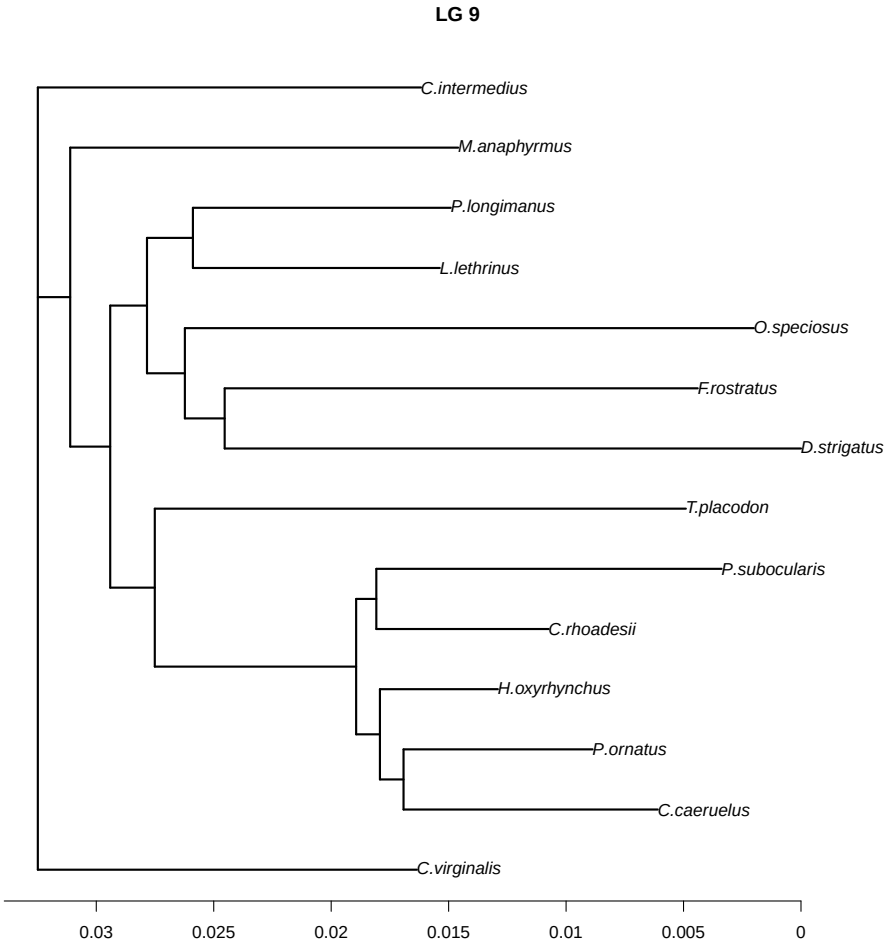

Figure H.9. Phylogenetic tree, linkage group 9

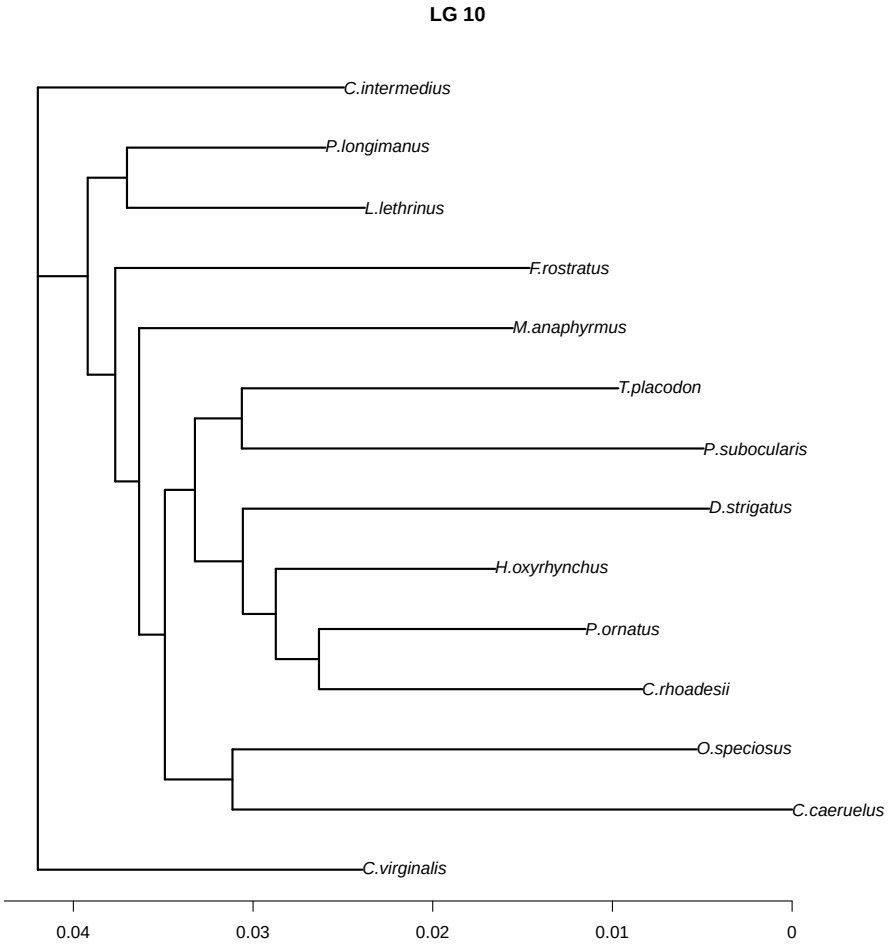

Figure H.10. Phylogenetic tree, linkage group10

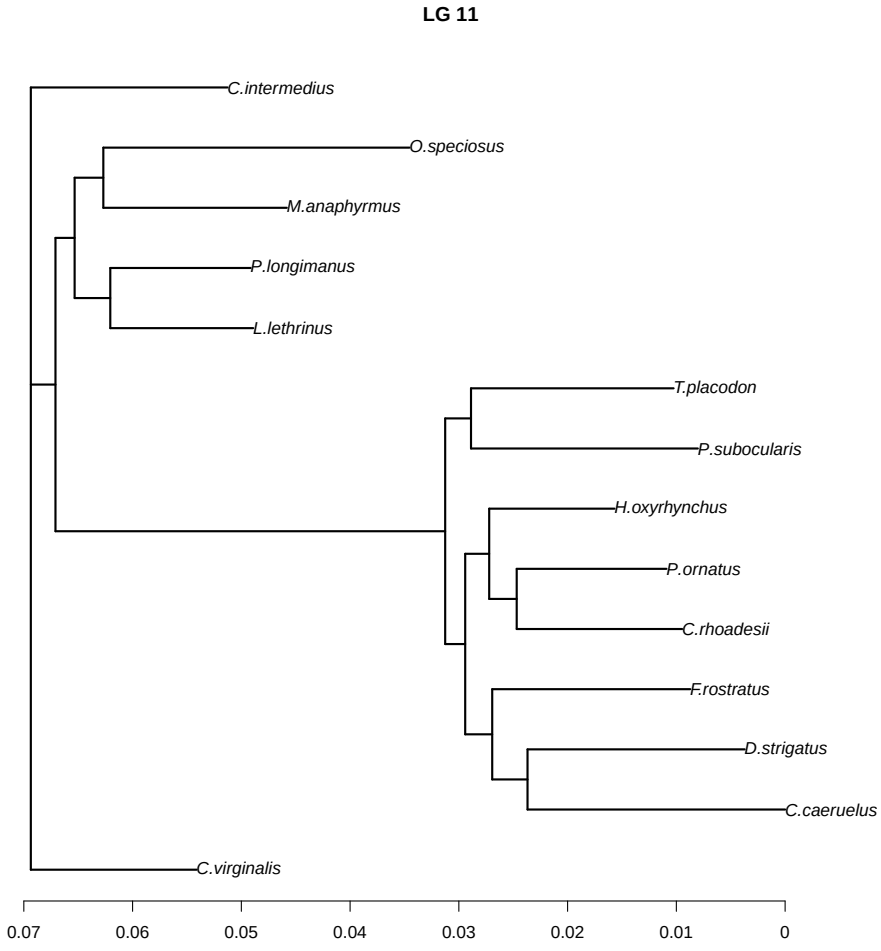

Figure H.11. Phylogenetic tree, linkage group 11

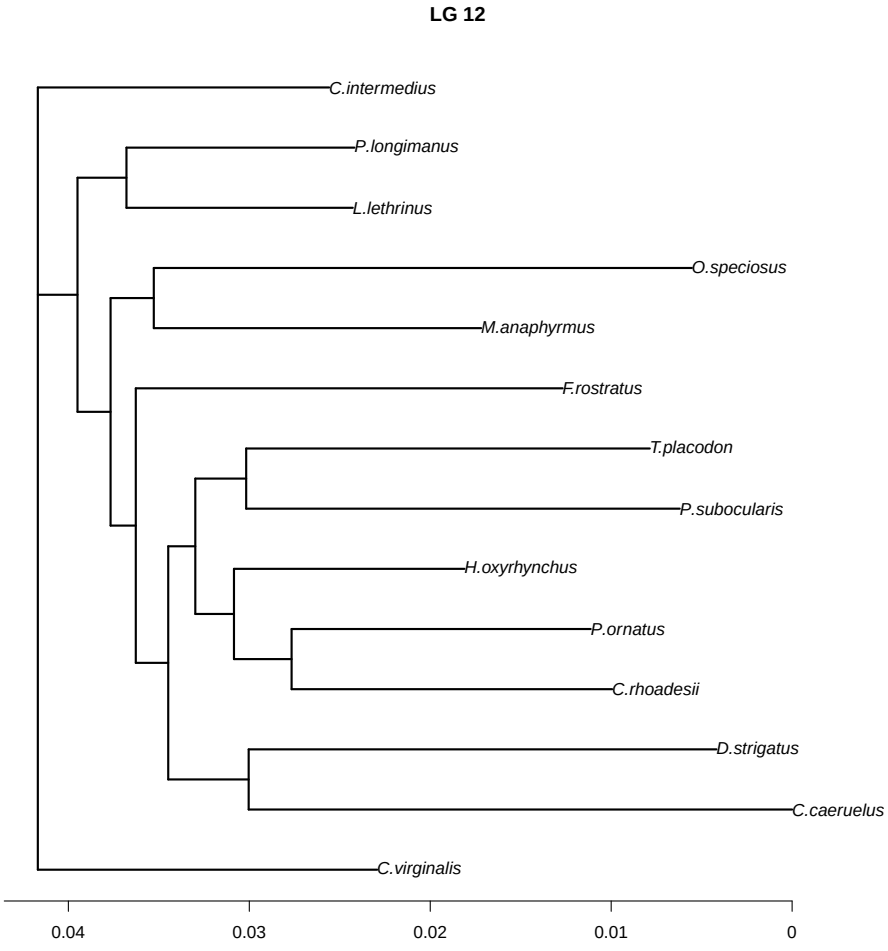

Figure H.12. Phylogenetic tree, linkage group 12

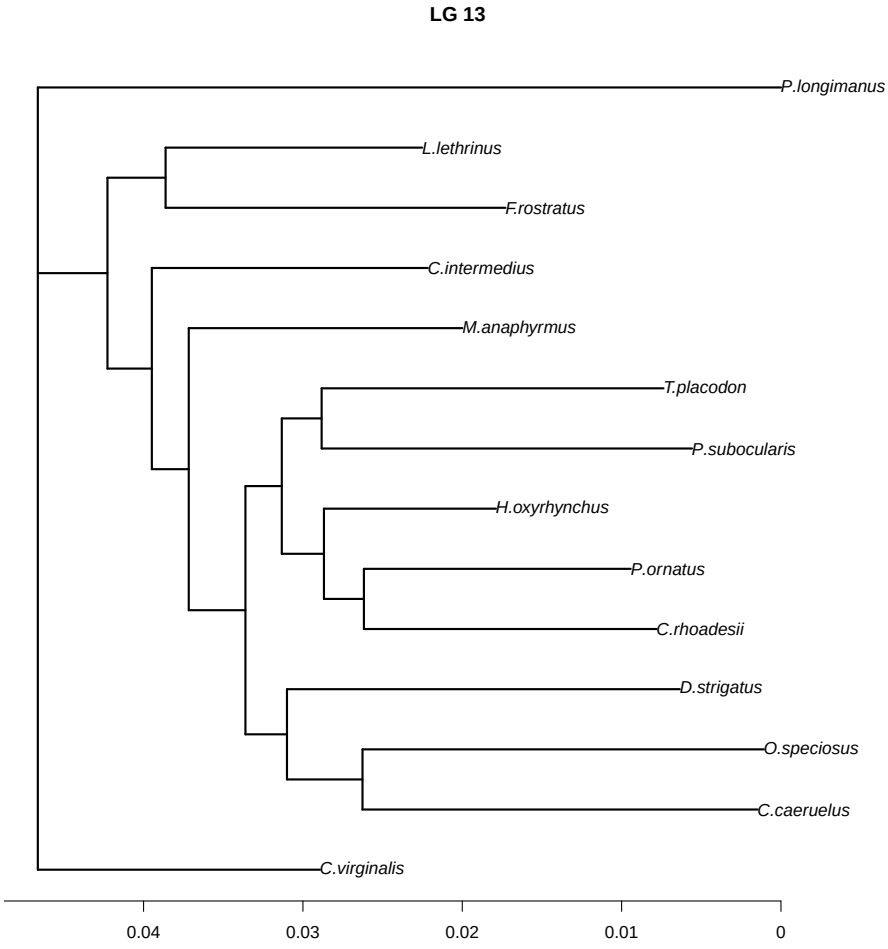

Figure H.13. Phylogenetic tree, linkage group 13

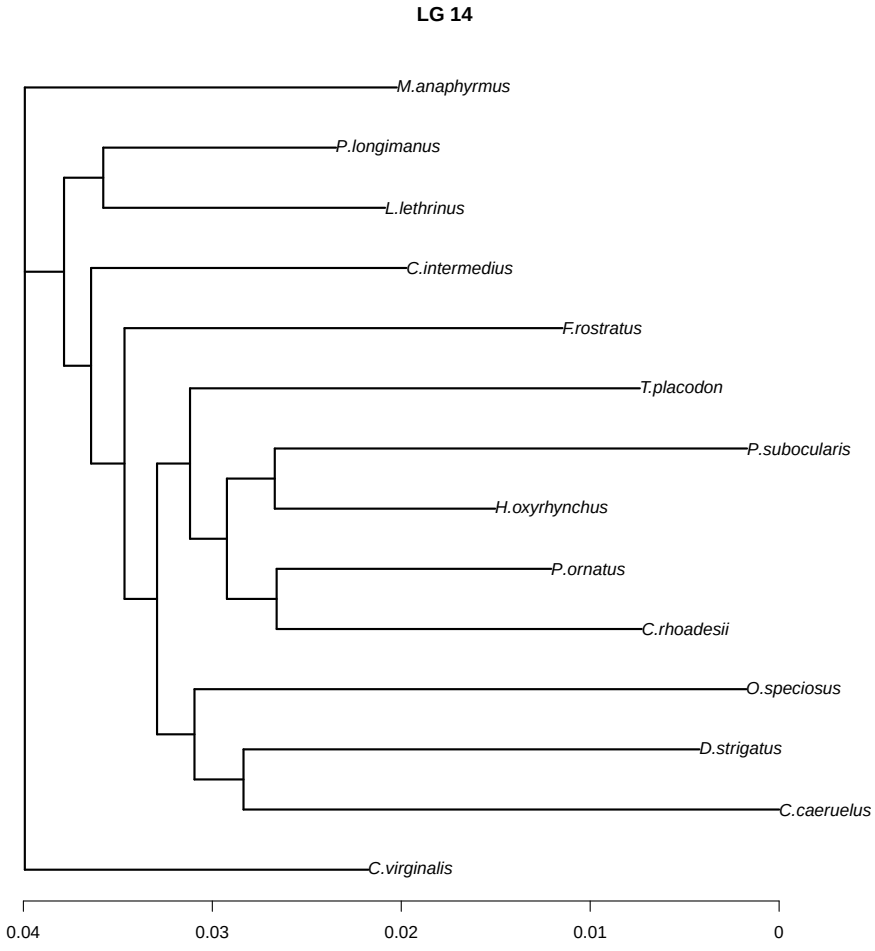

Figure H.14. Phylogenetic tree, linkage group 14

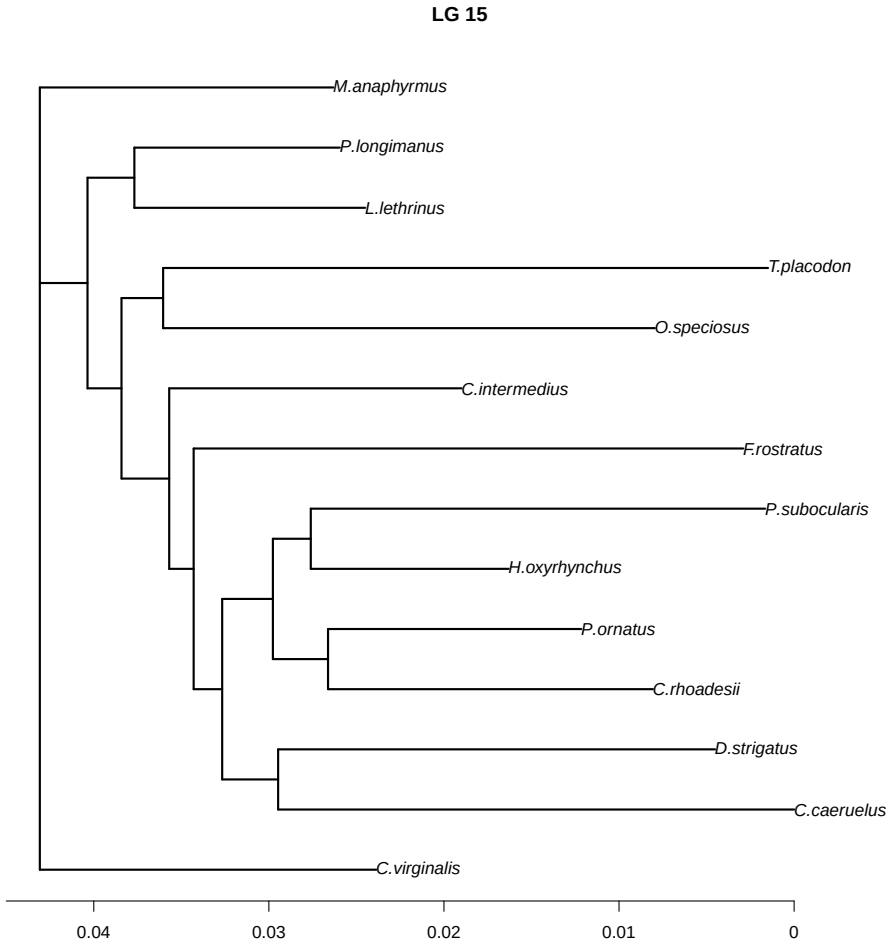

Figure H.15. Phylogenetic tree, linkage group 15

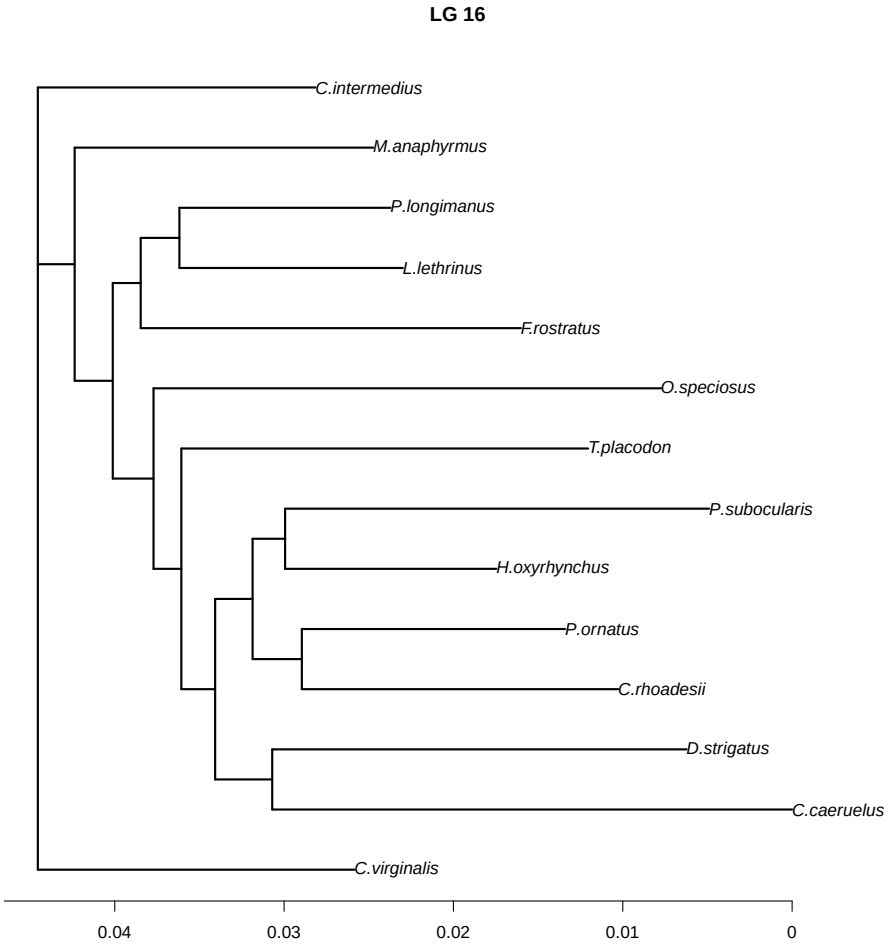

Figure H.16. Phylogenetic tree, linkage group 16

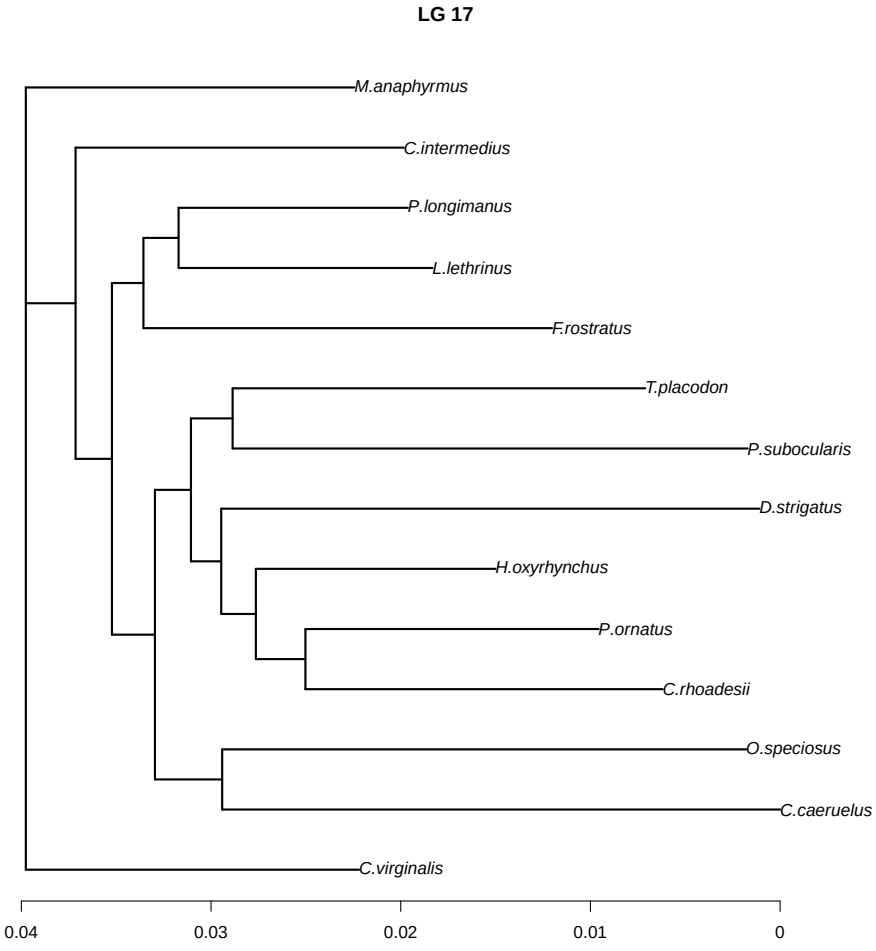

Figure H.17. Phylogenetic tree, linkage group 17

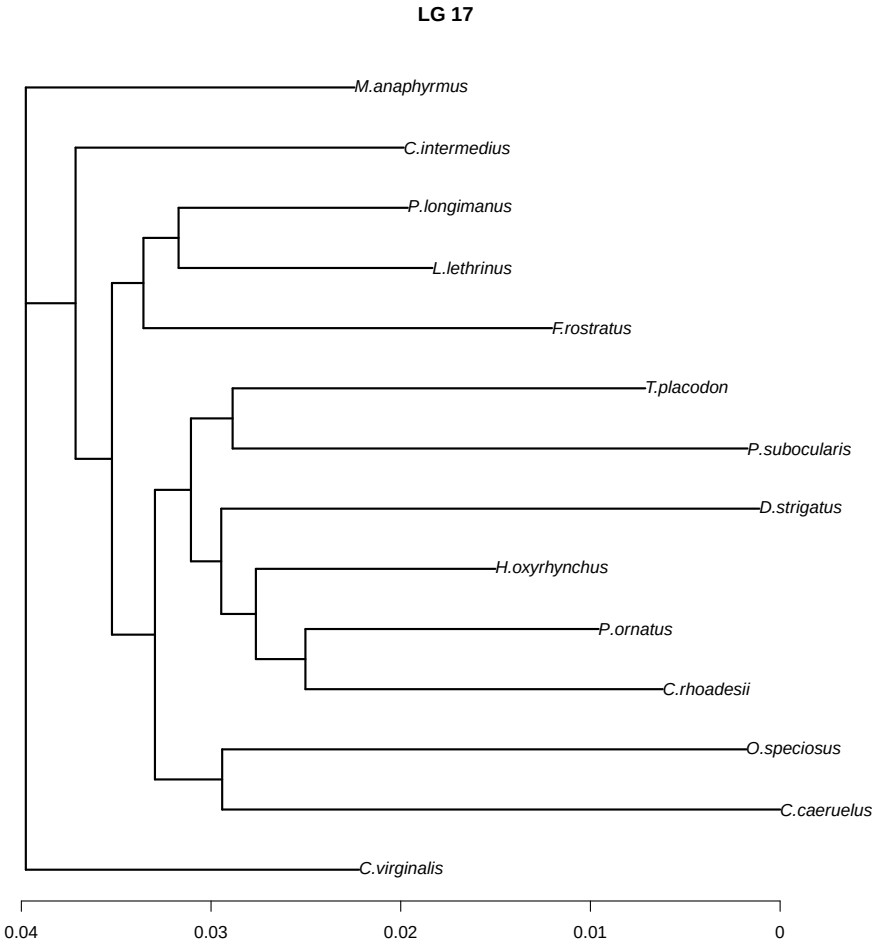

Figure H.18. Phylogenetic tree, linkage group 17

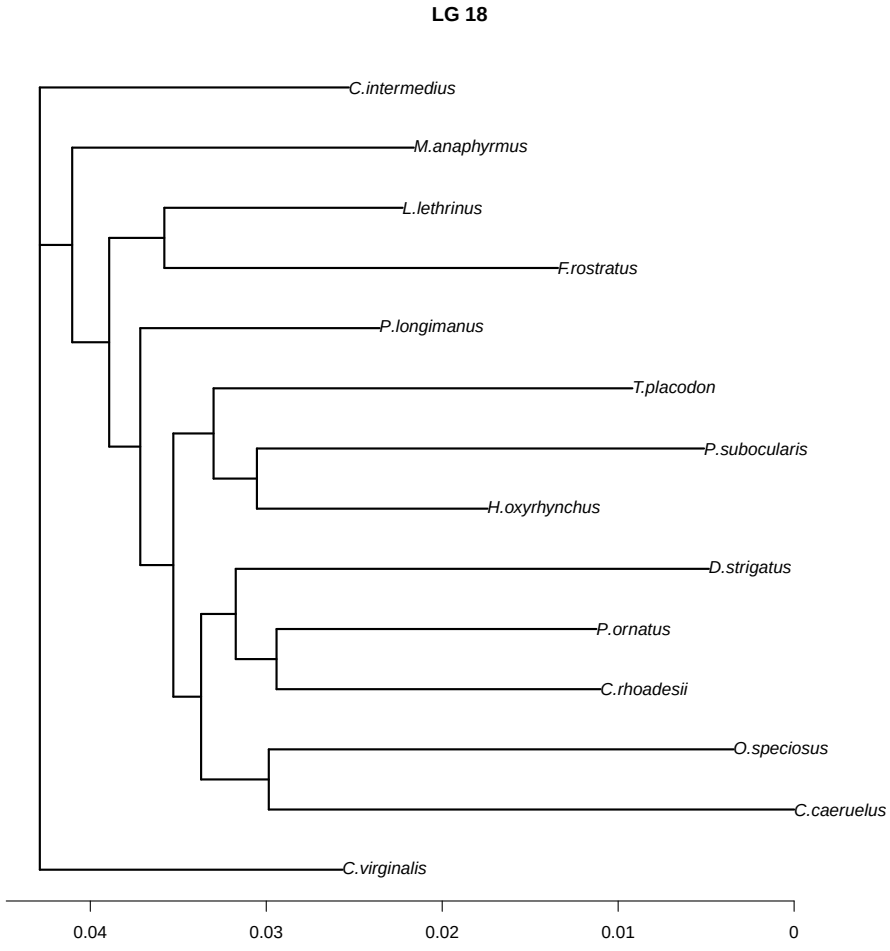

Figure H.19. Phylogenetic tree, linkage group 18

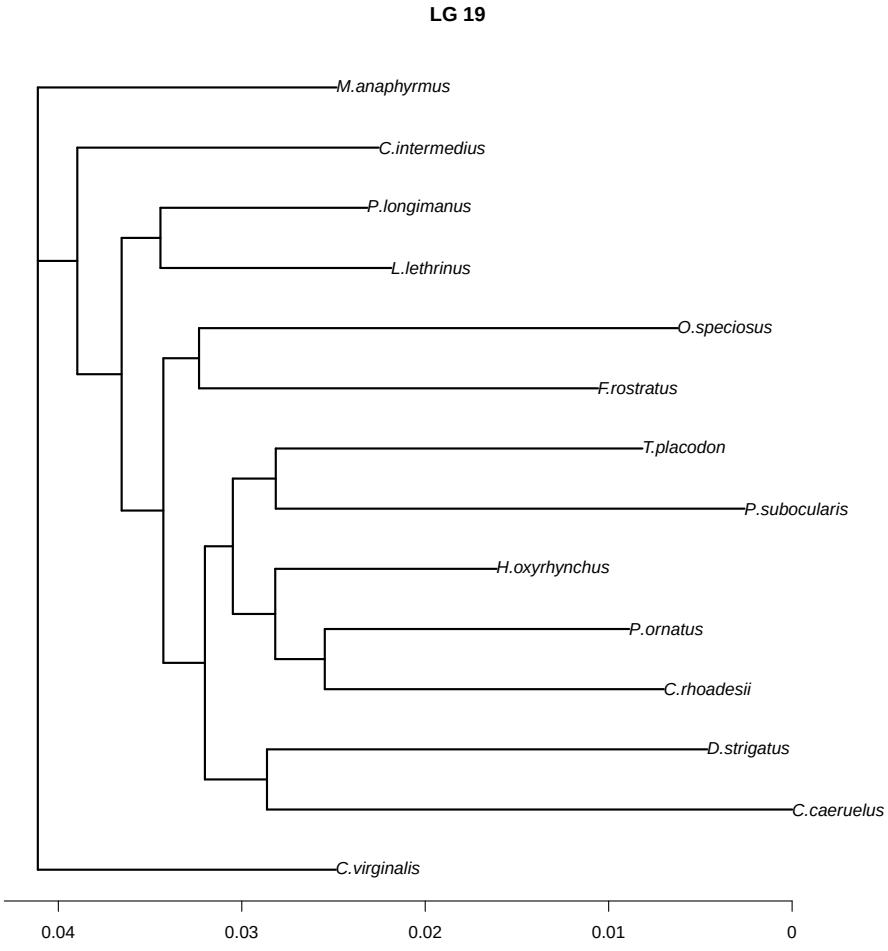

Figure H.20. Phylogenetic tree, linkage group 19

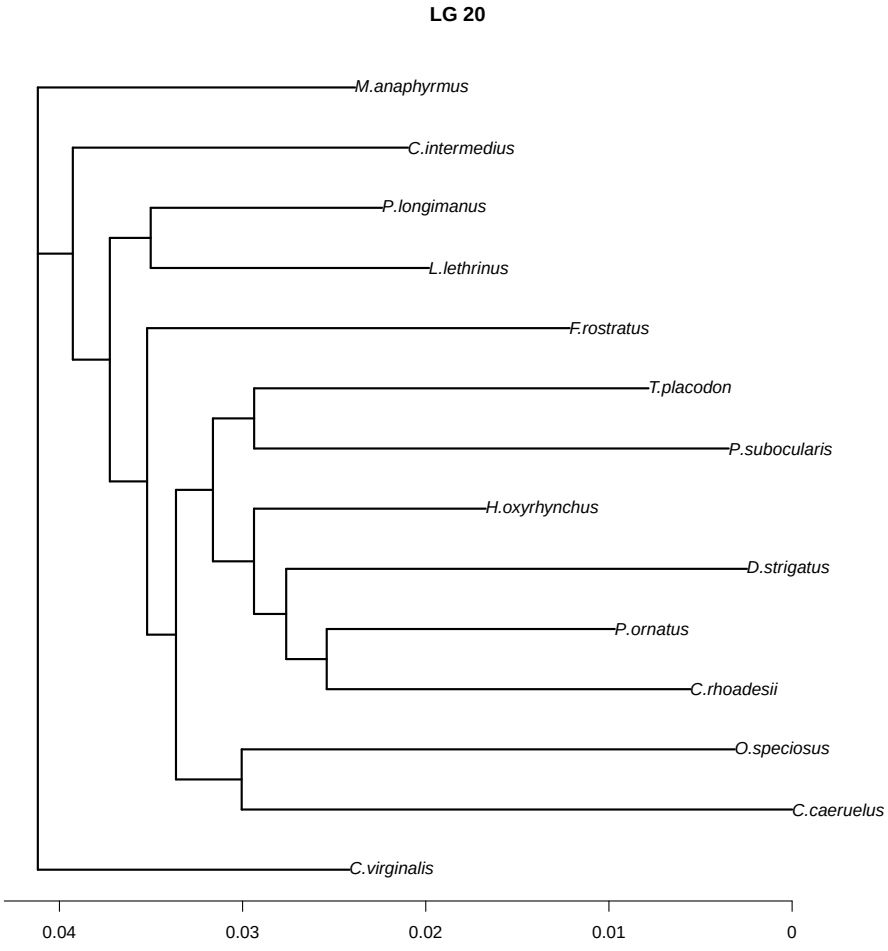

Figure H.21. Phylogenetic tree, linkage group 20

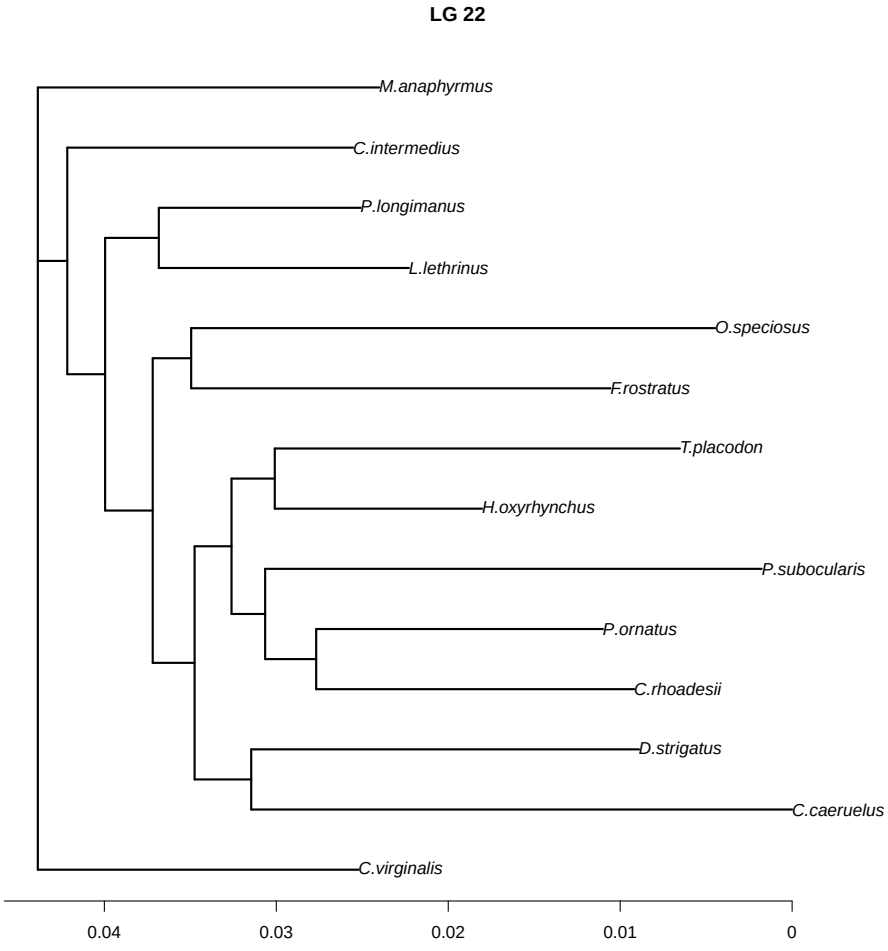

Figure H.22. Phylogenetic tree, linkage group 22

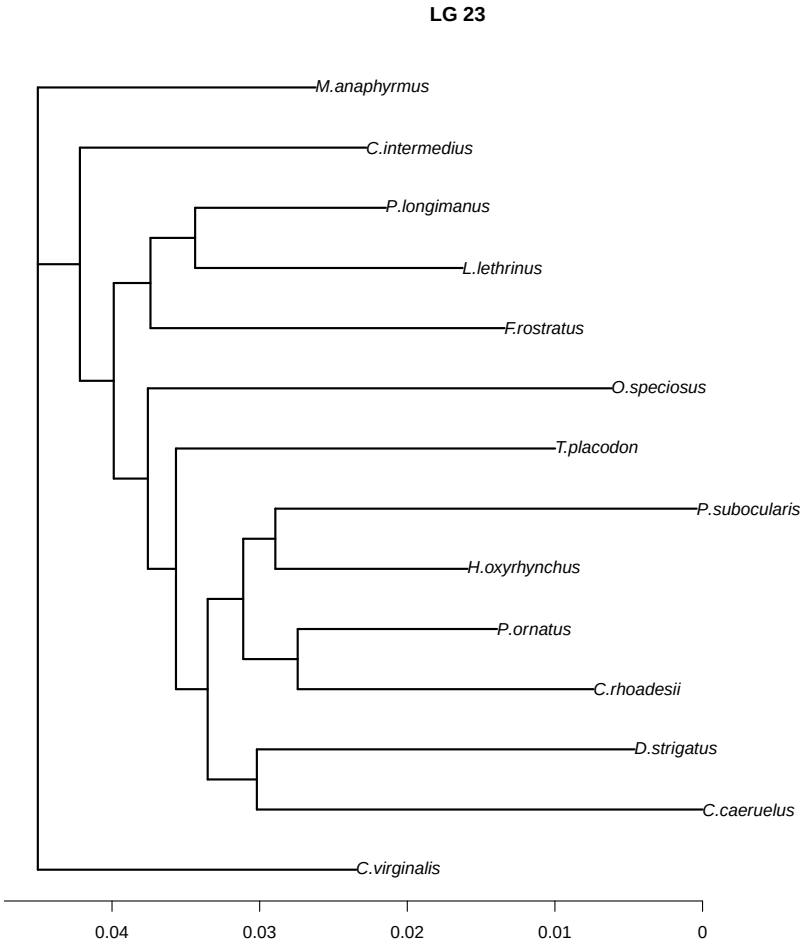

Figure H.23. Phylogenetic tree, linkage group 23

APPENDIX I

FBRANCH

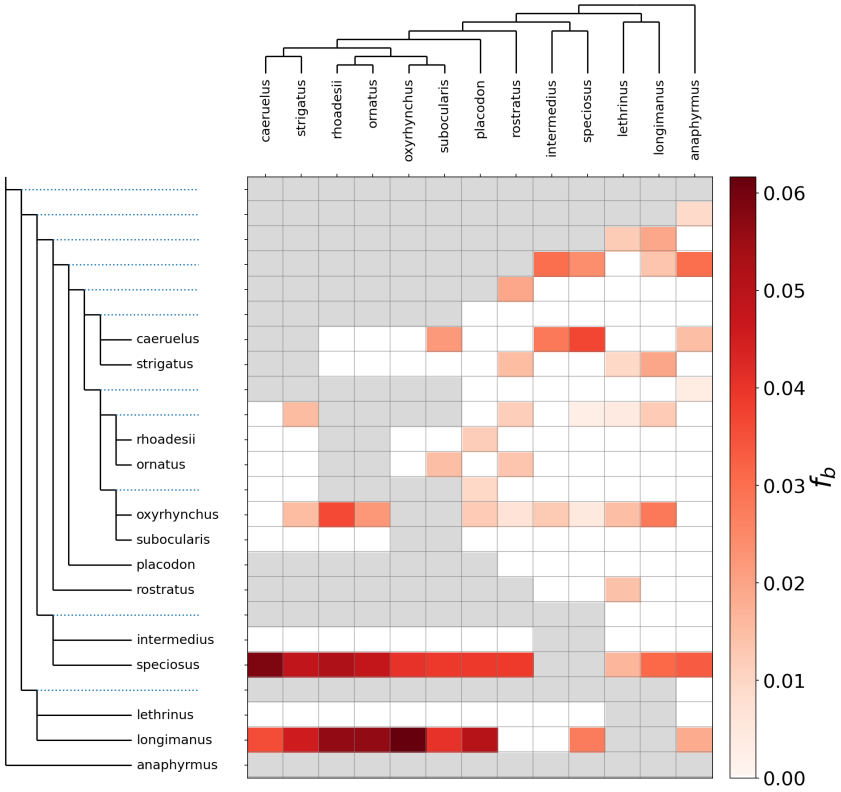

Figure I.1. fbranch result whole genome

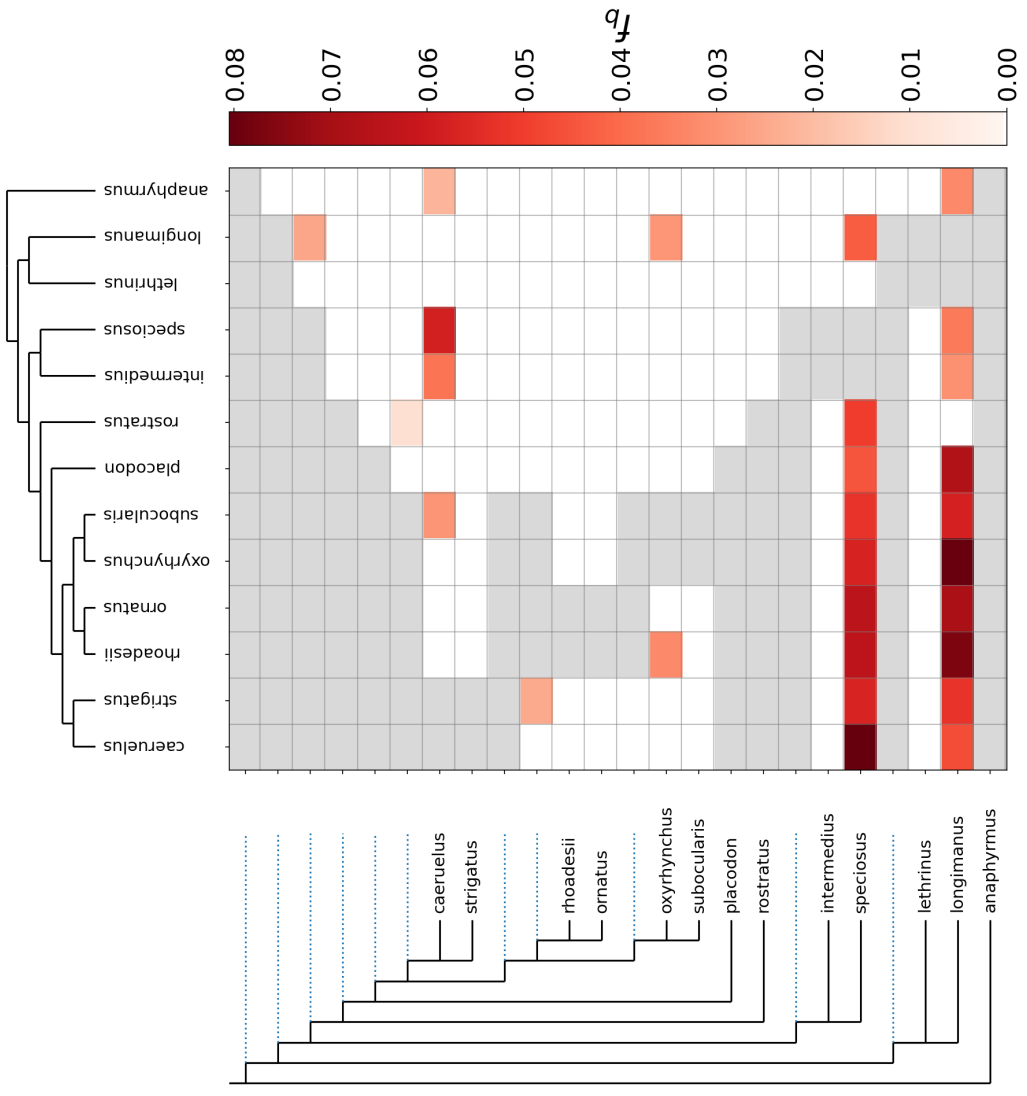

Figure I.2. fbranch result linkage group 1

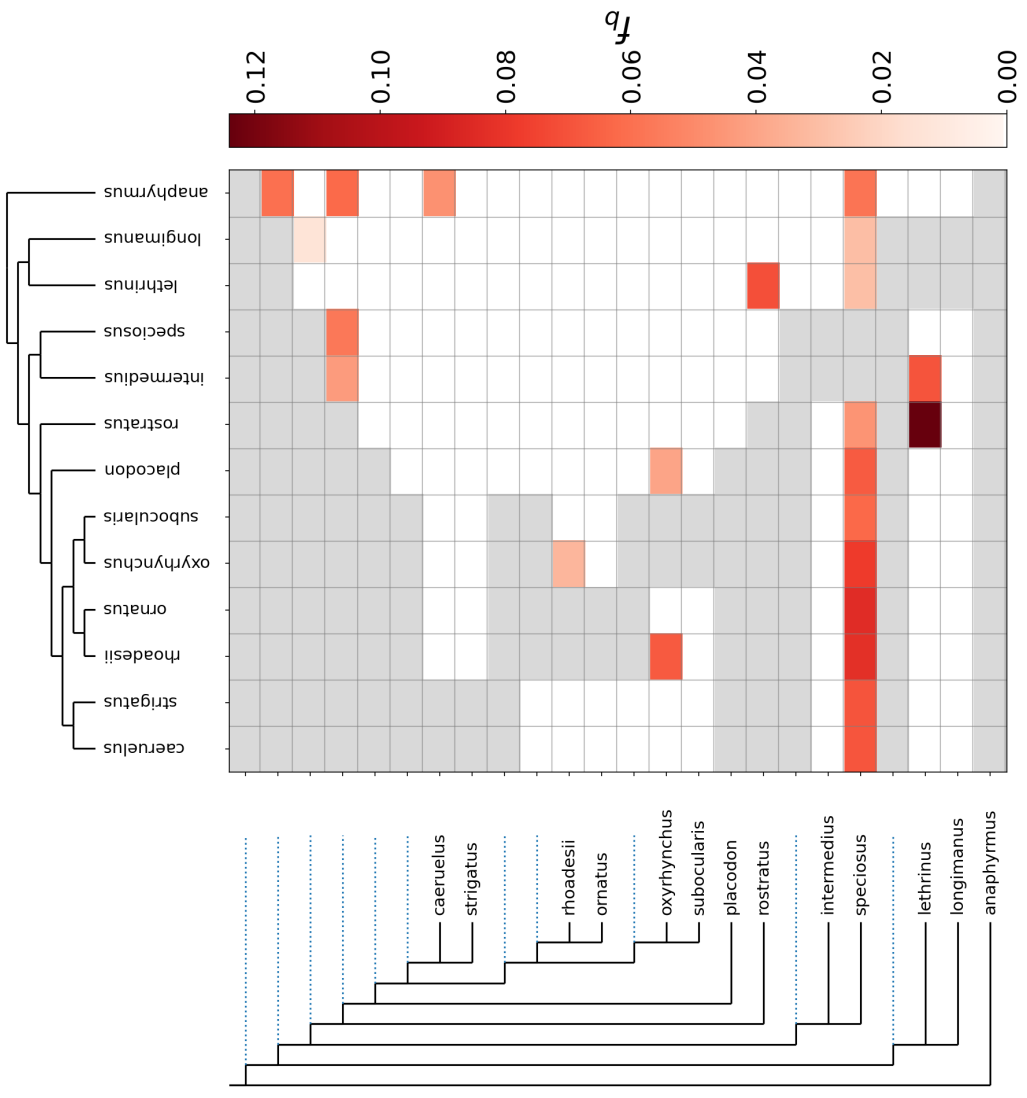

Figure I.3. fbranch result linkage group 2

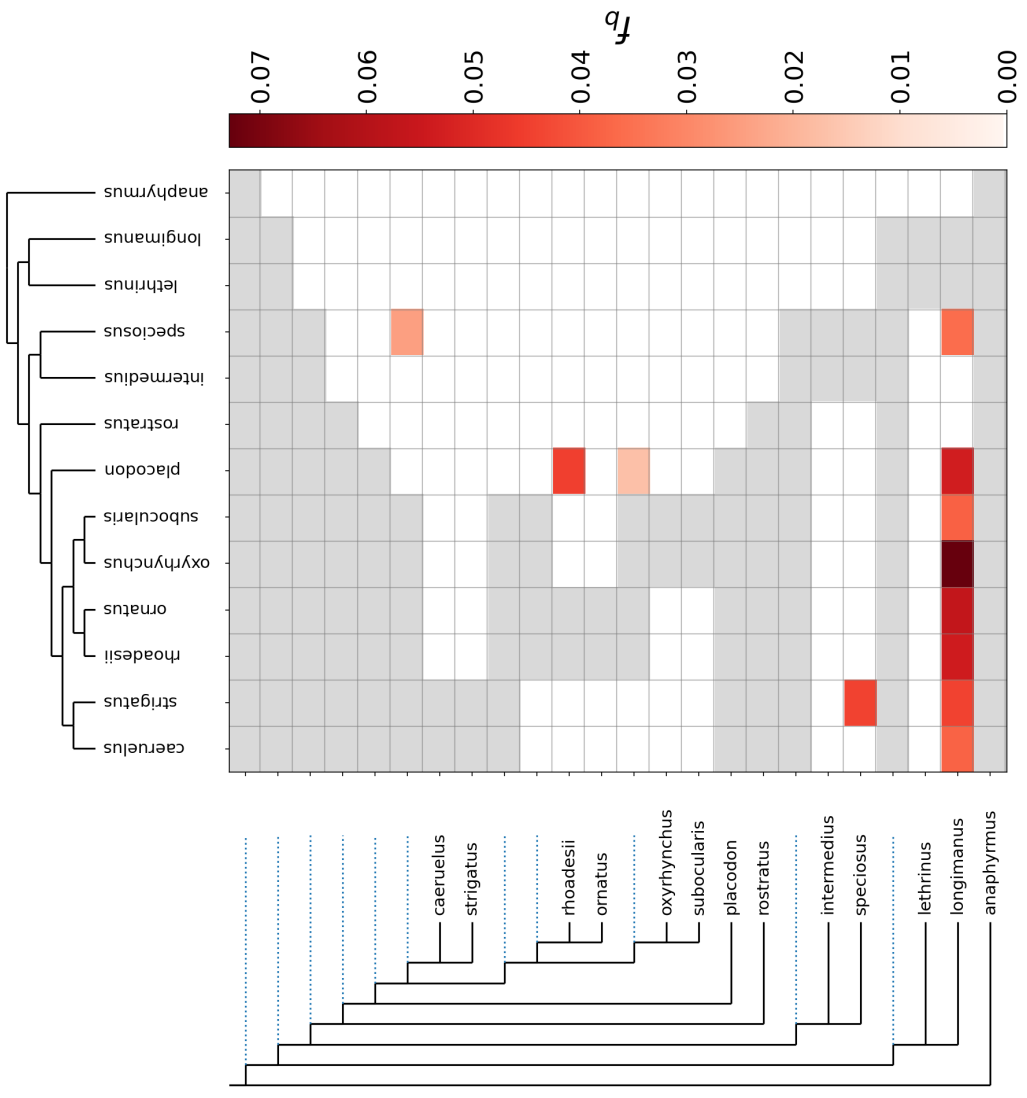

Figure I.4. fbranch result linkage group 3

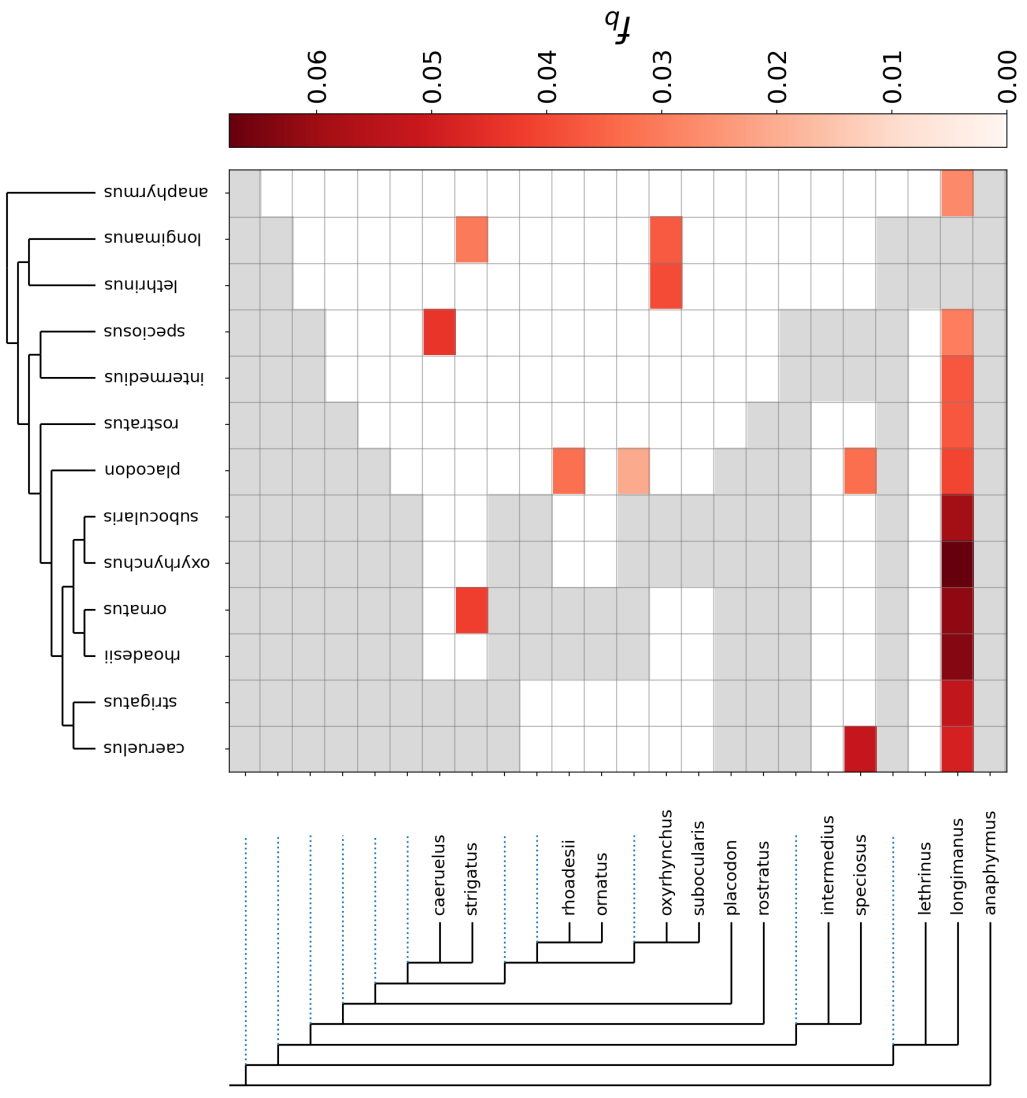

Figure I.5. fbranch result linkage group 4

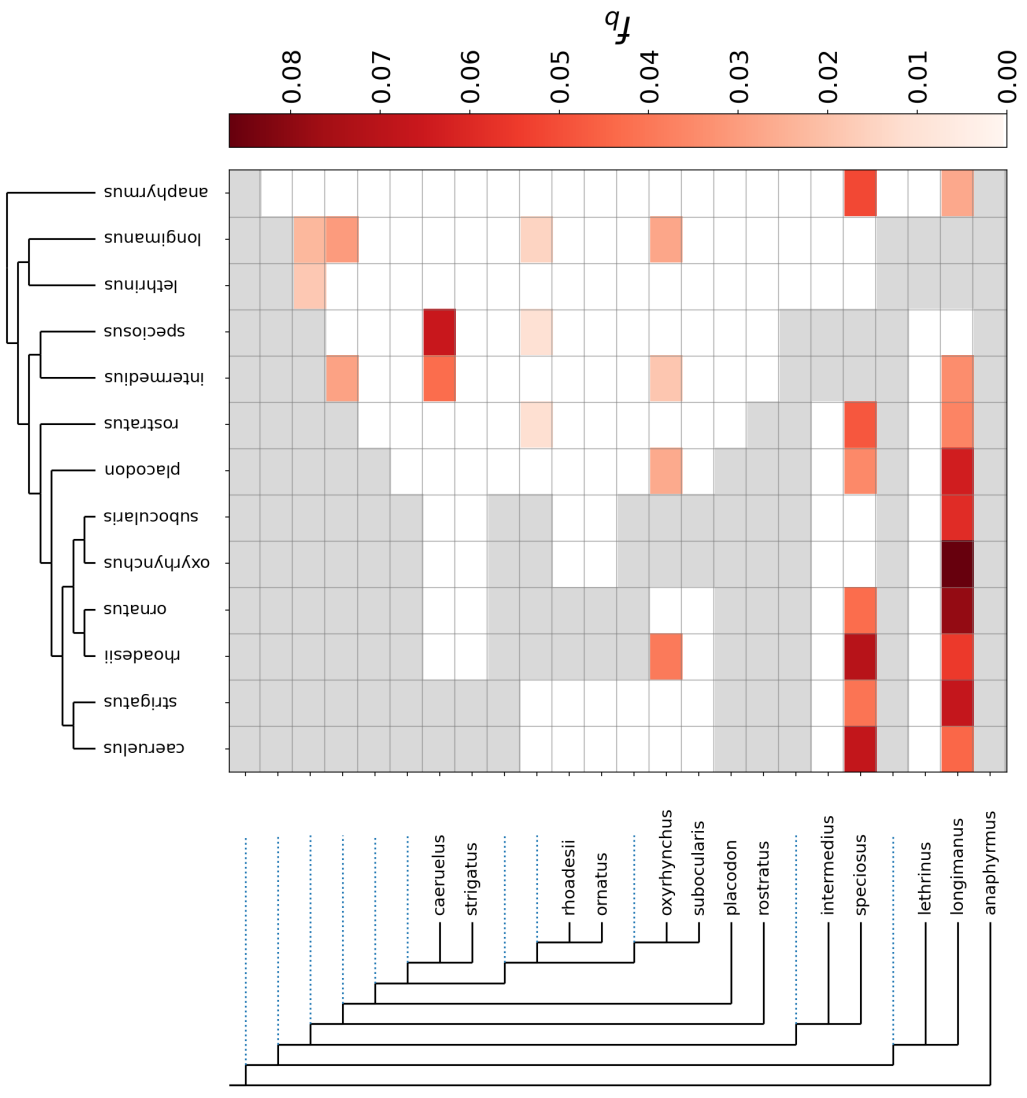

Figure I.6. fbranch result linkage group 5

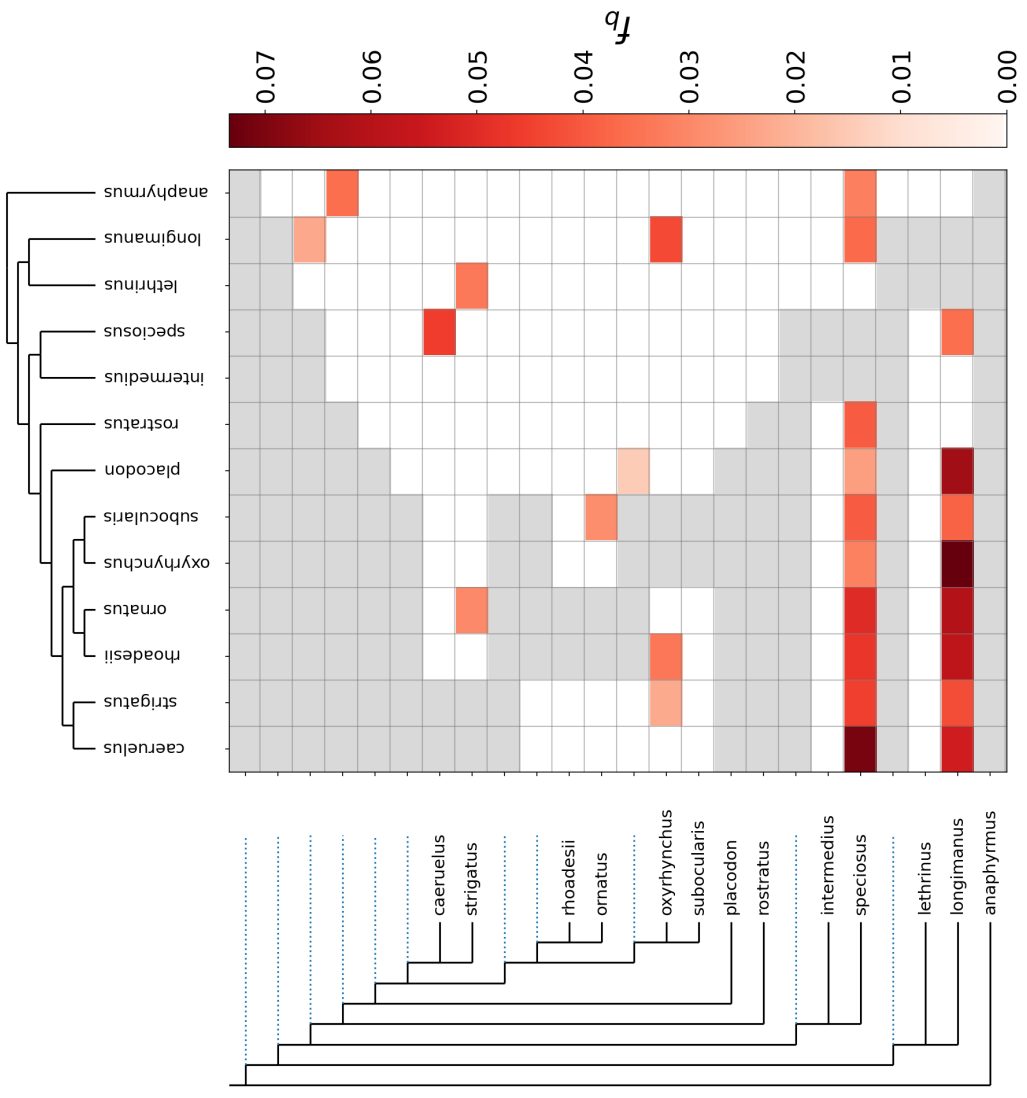

Figure I.7. fbranch result linkage group 6

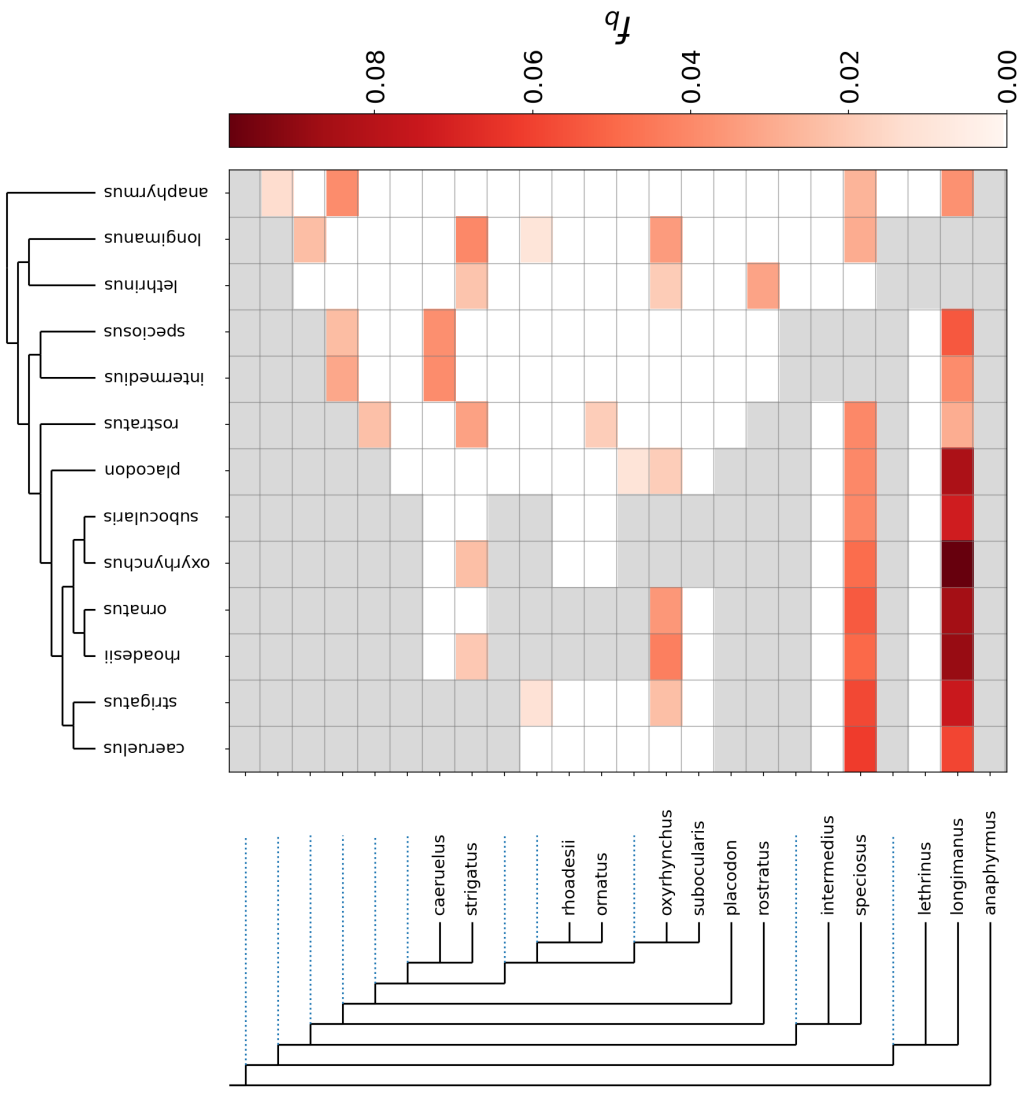

Figure I.8. fbranch result linkage group 7

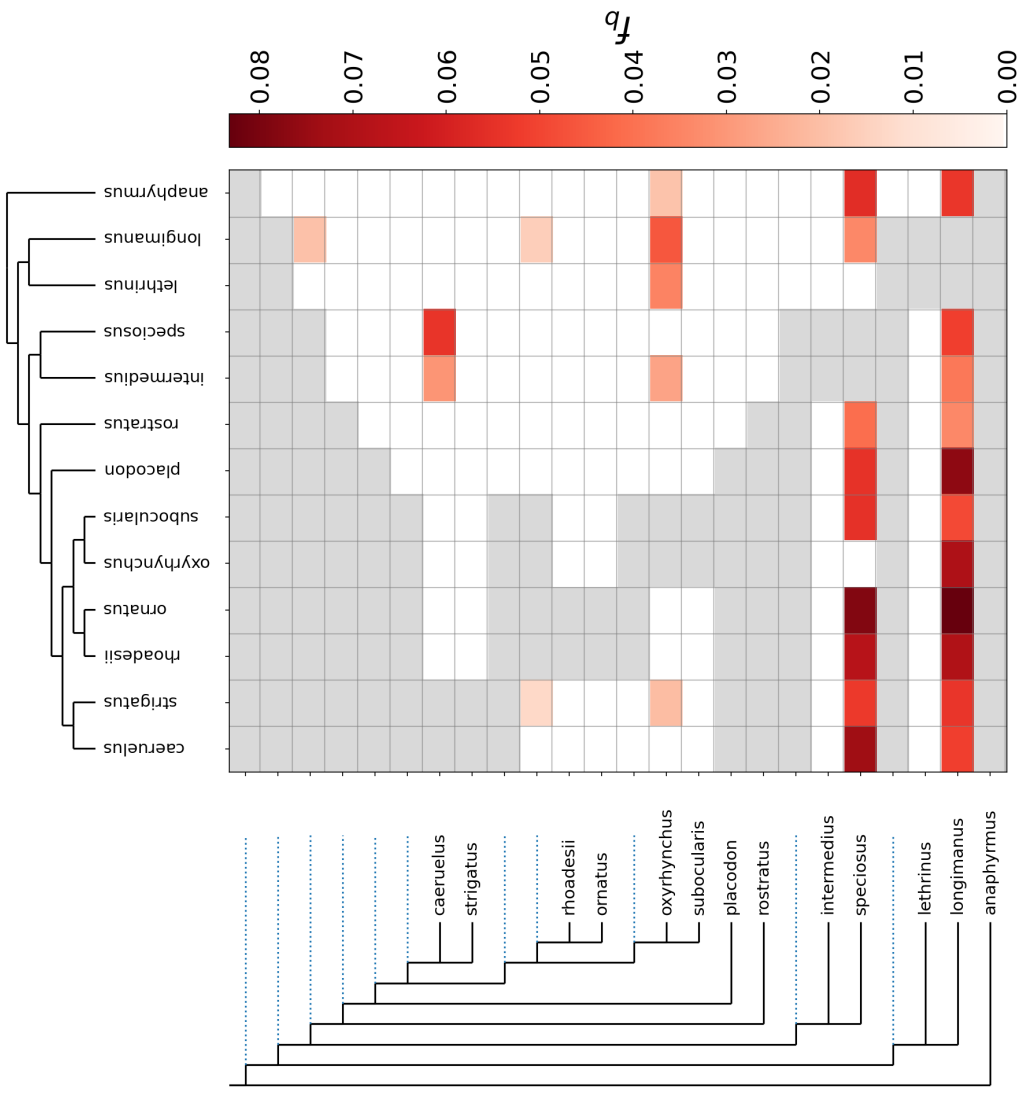

Figure I.9. fbranch result linkage group 8

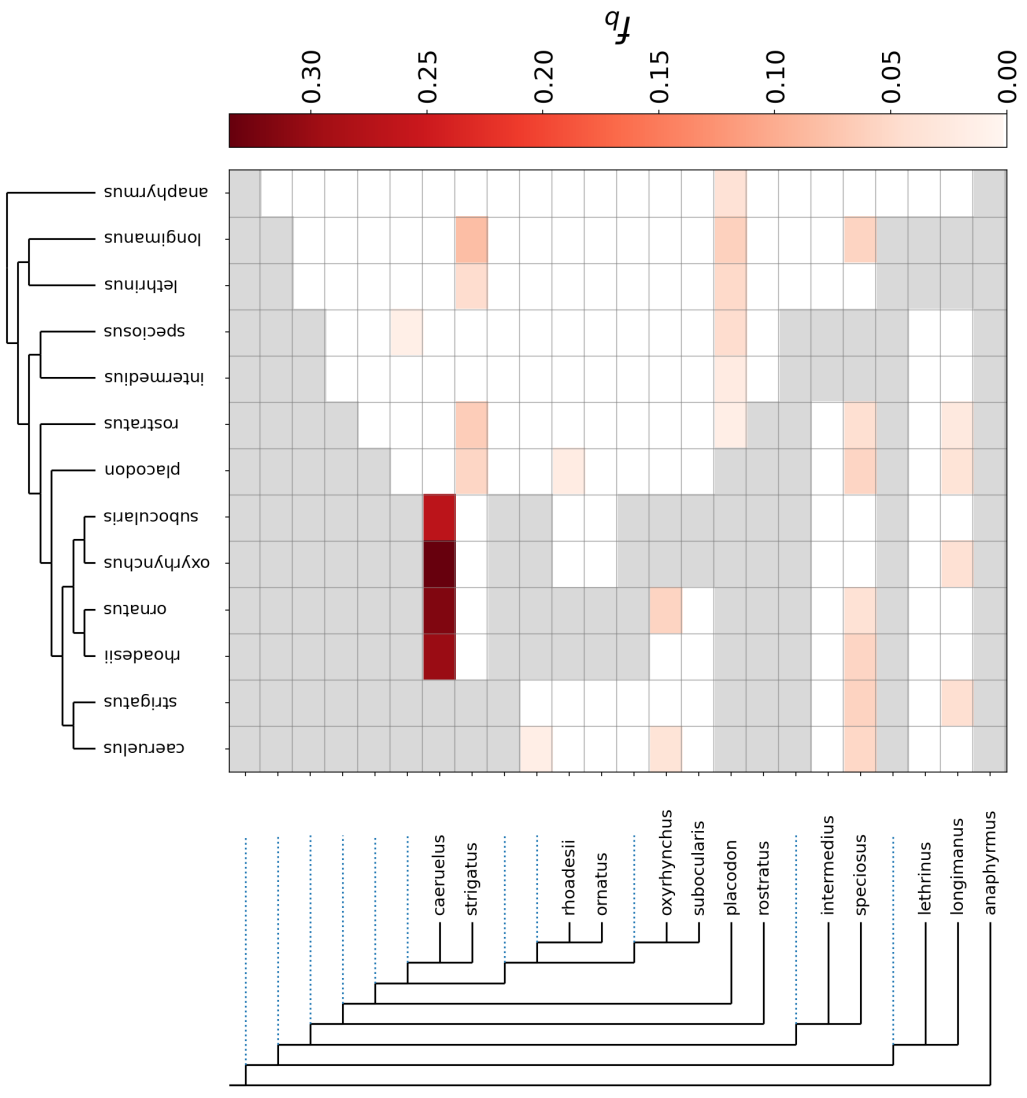

Figure I.10. fbranch result linkage group 9

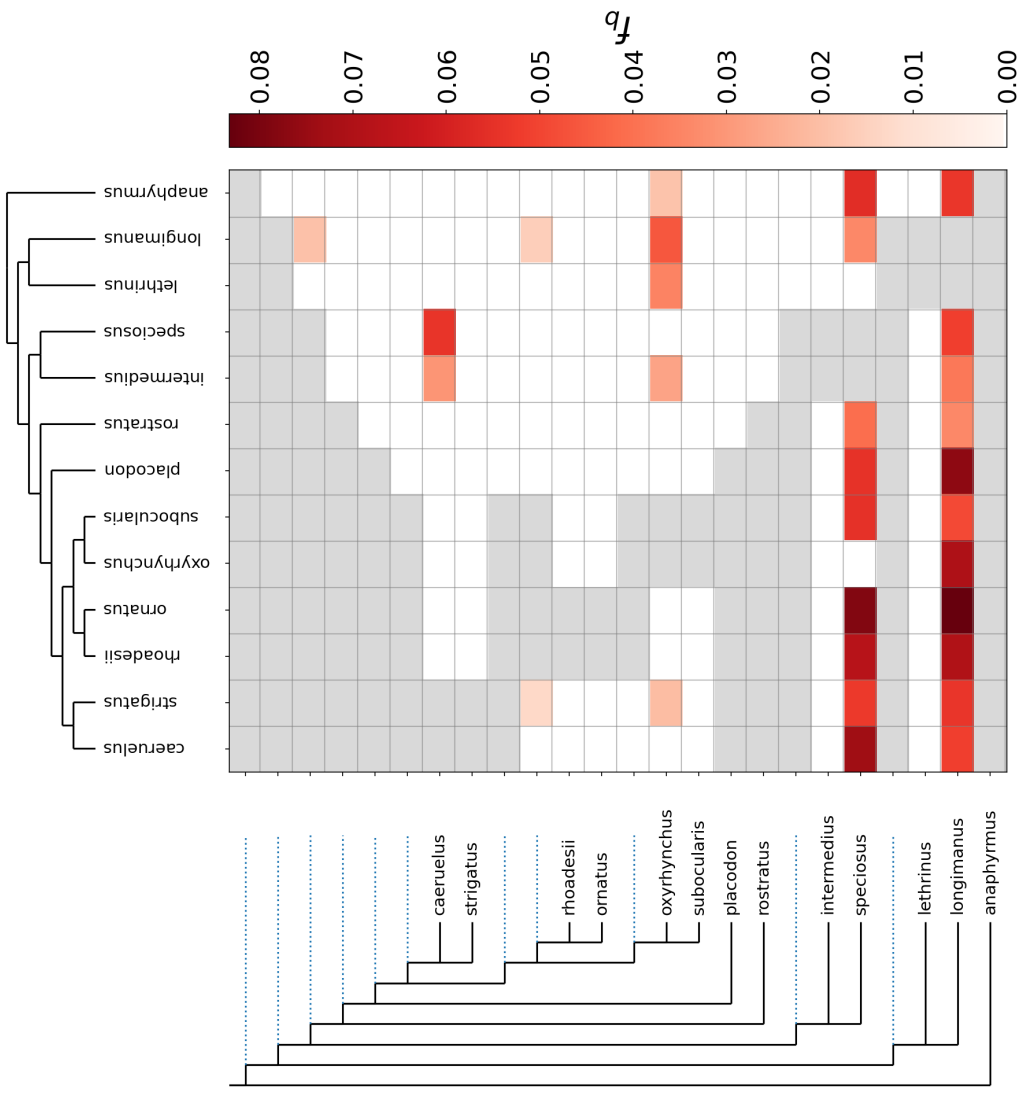

Figure I.11. fbranch result linkage group 10

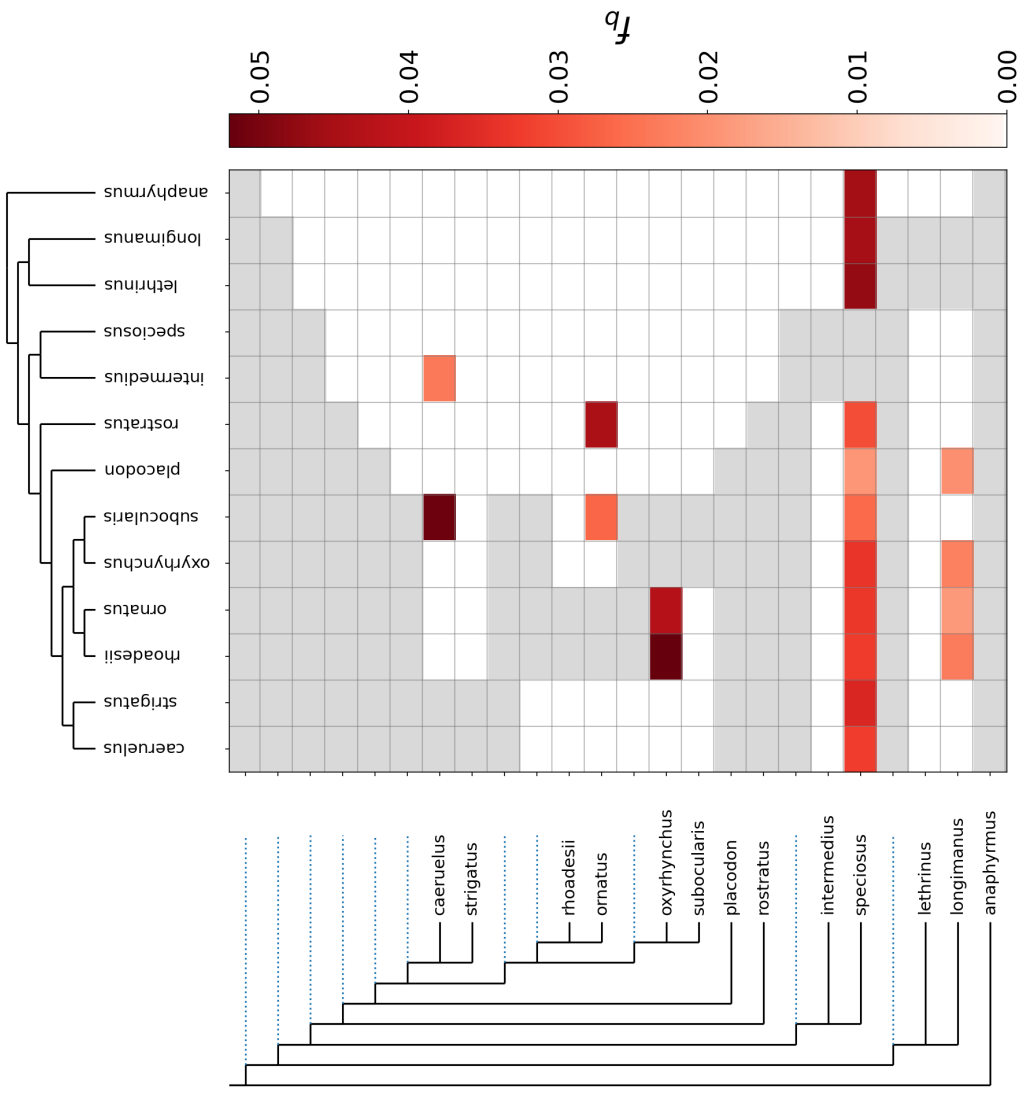

Figure I.12. fbranch result linkage group 11

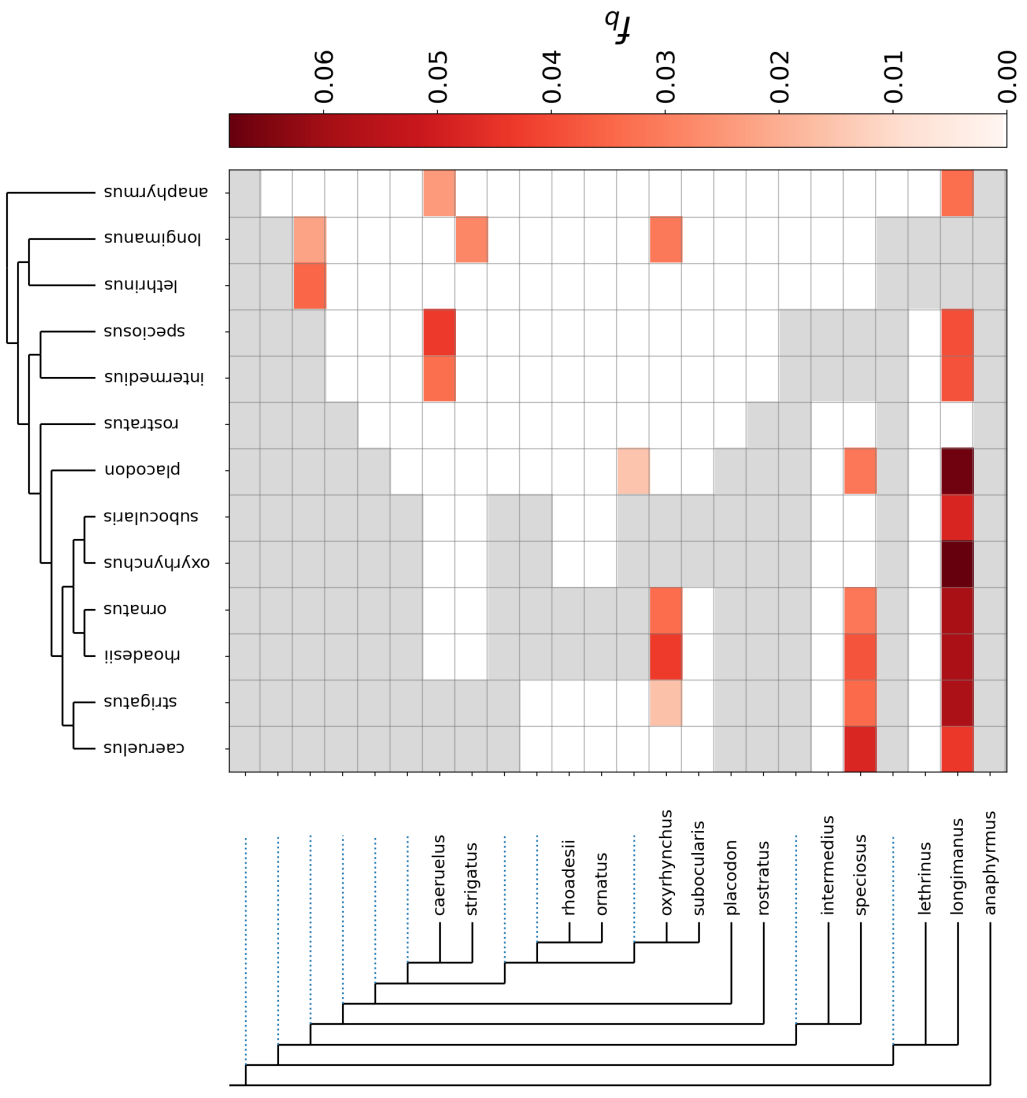

Figure I.13. fbranch result linkage group 12

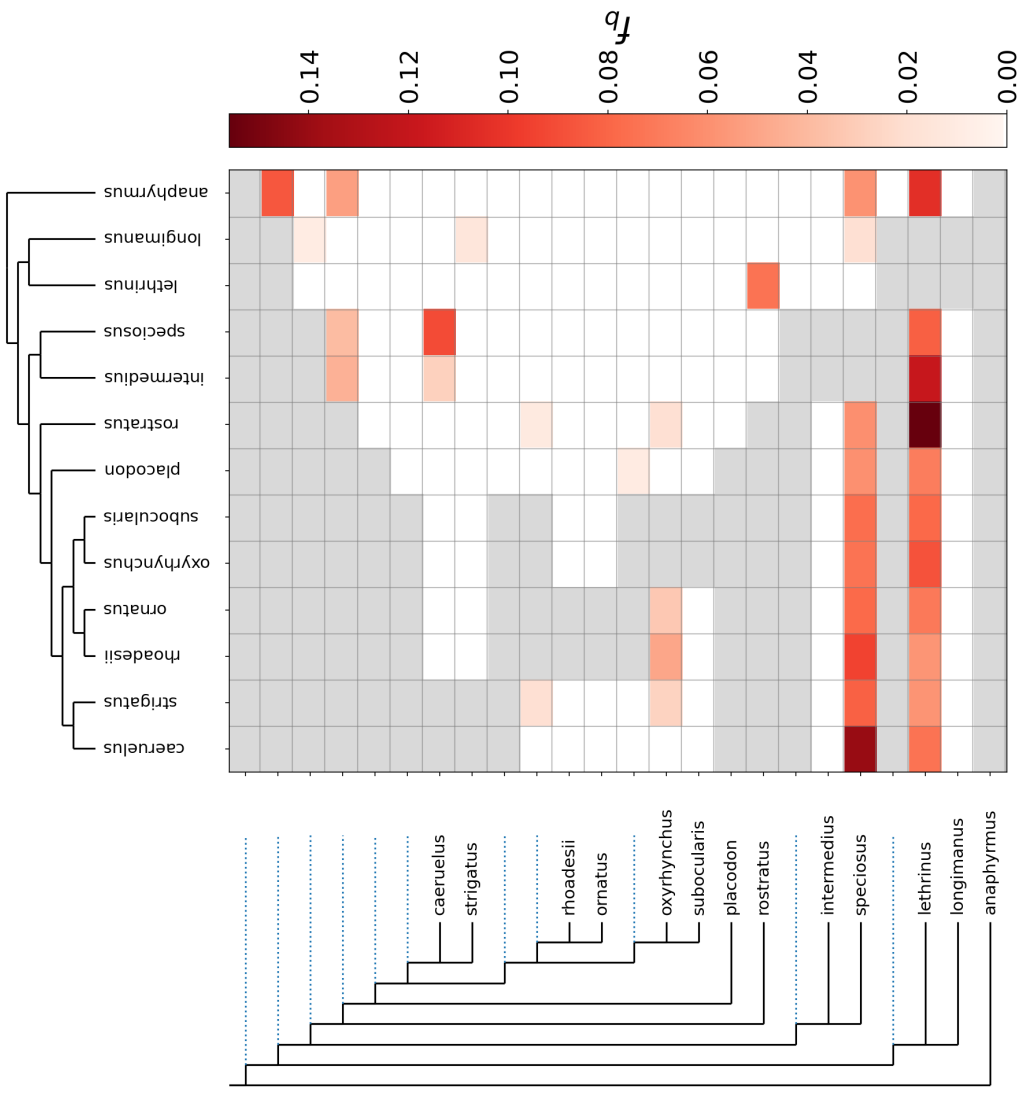

Figure I.14. fbranch result linkage group 13

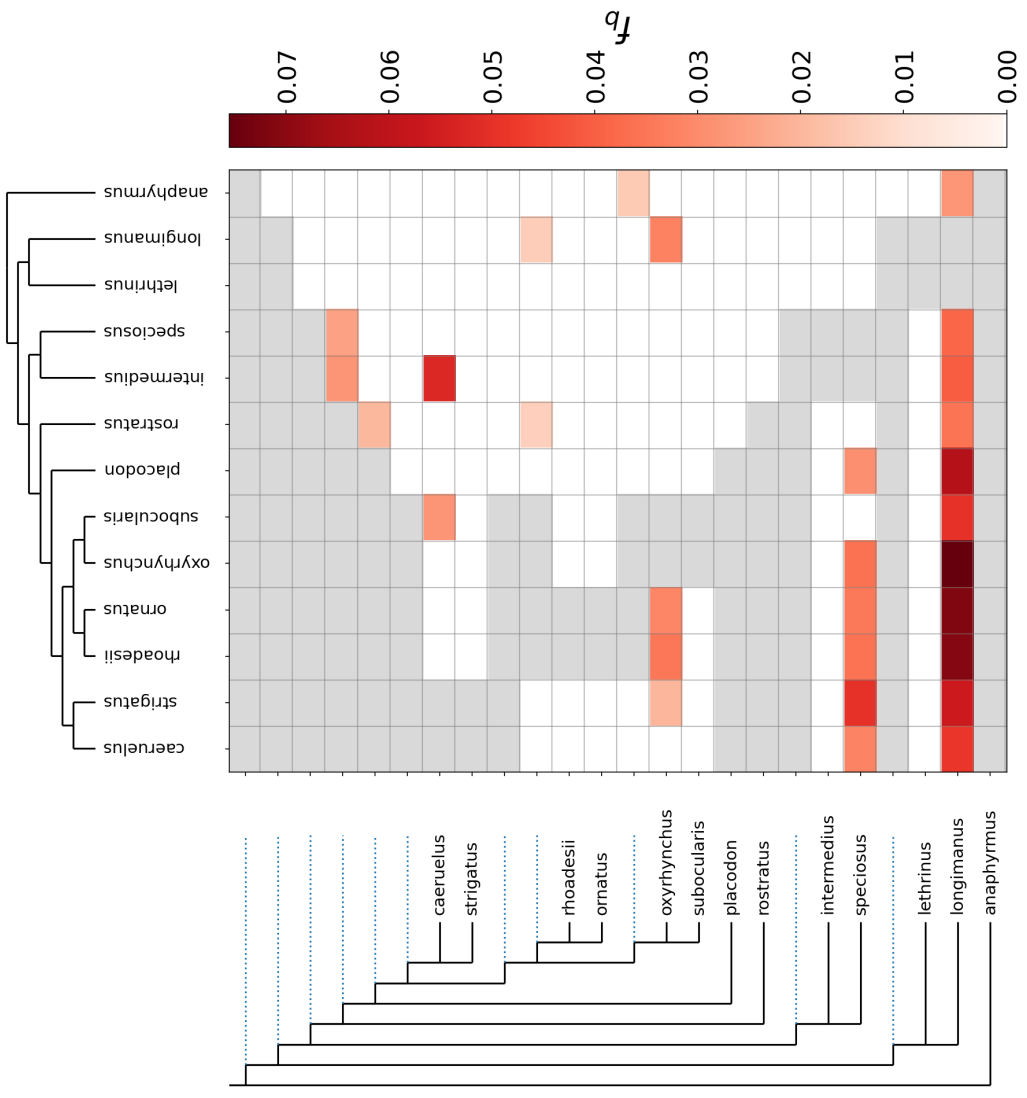

Figure I.15. fbranch result linkage group 14

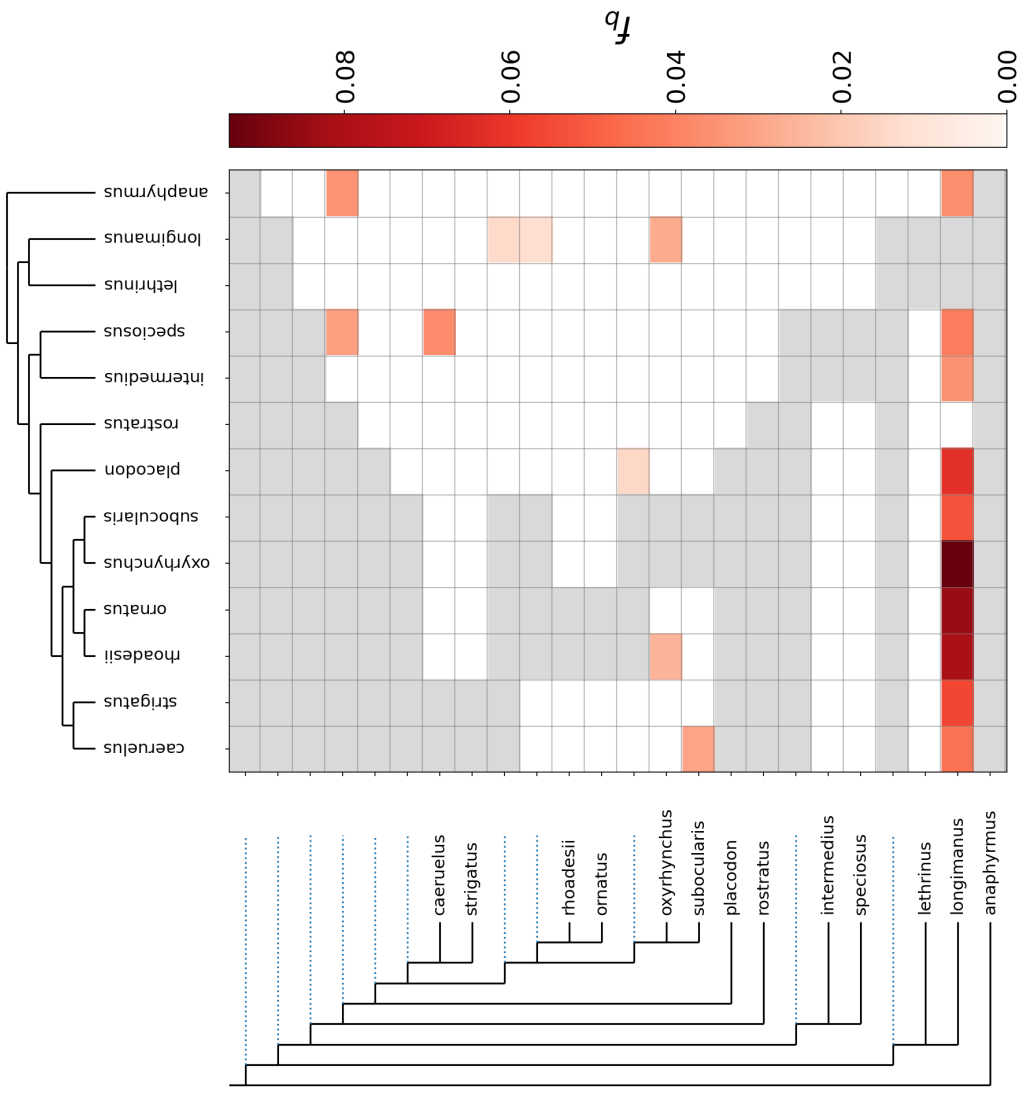

Figure I.16. fbranch result linkage group 15

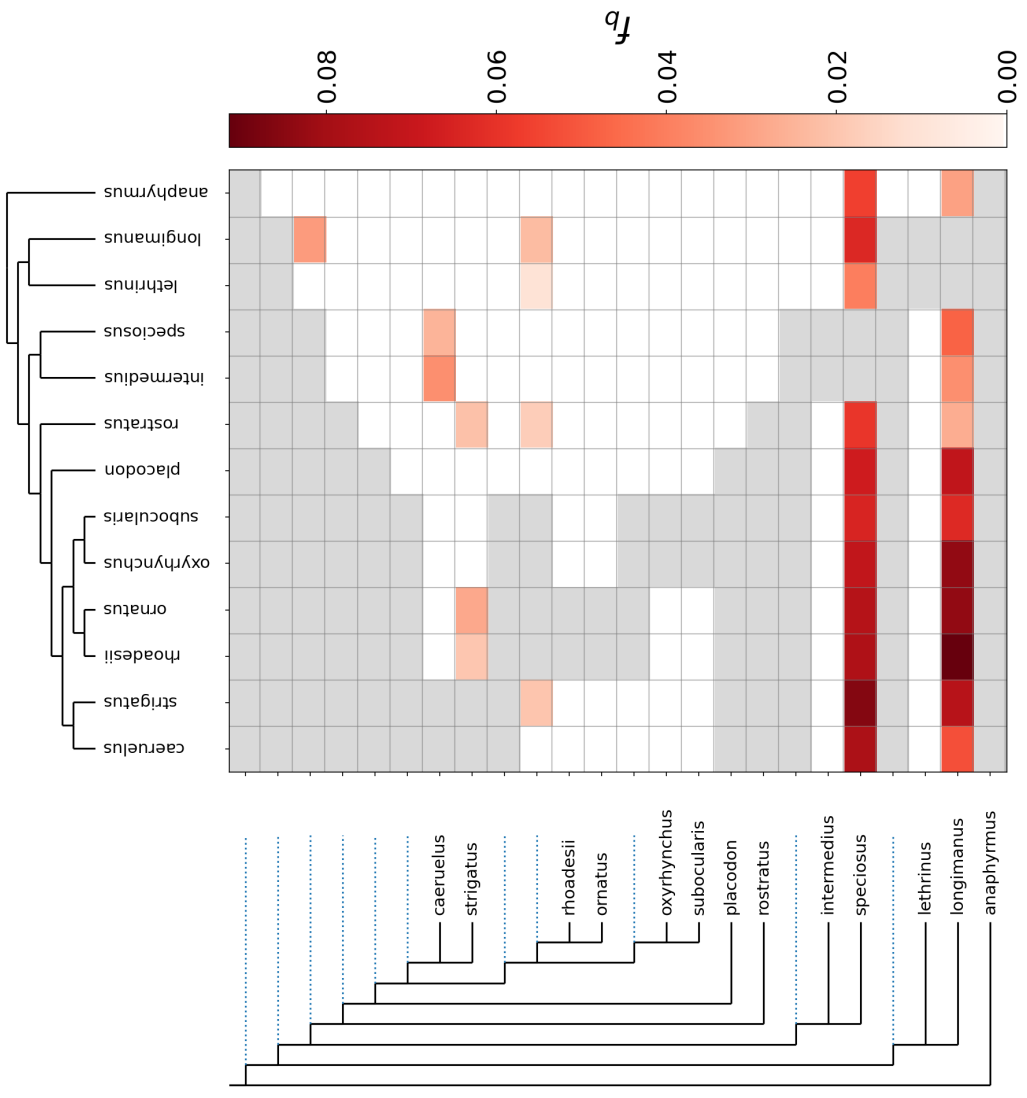

Figure I.17. fbranch result linkage group 16

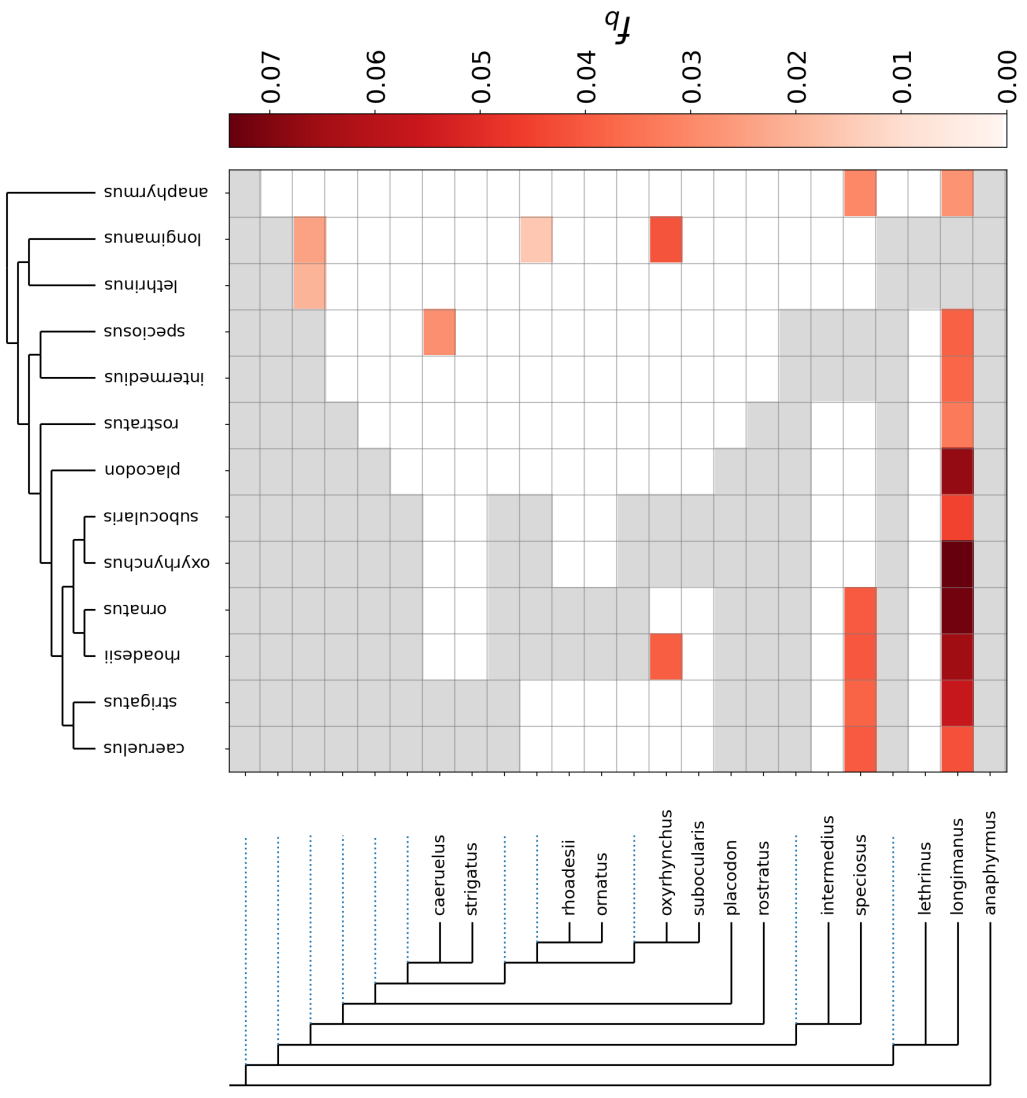

Figure I.18. fbranch result linkage group 17

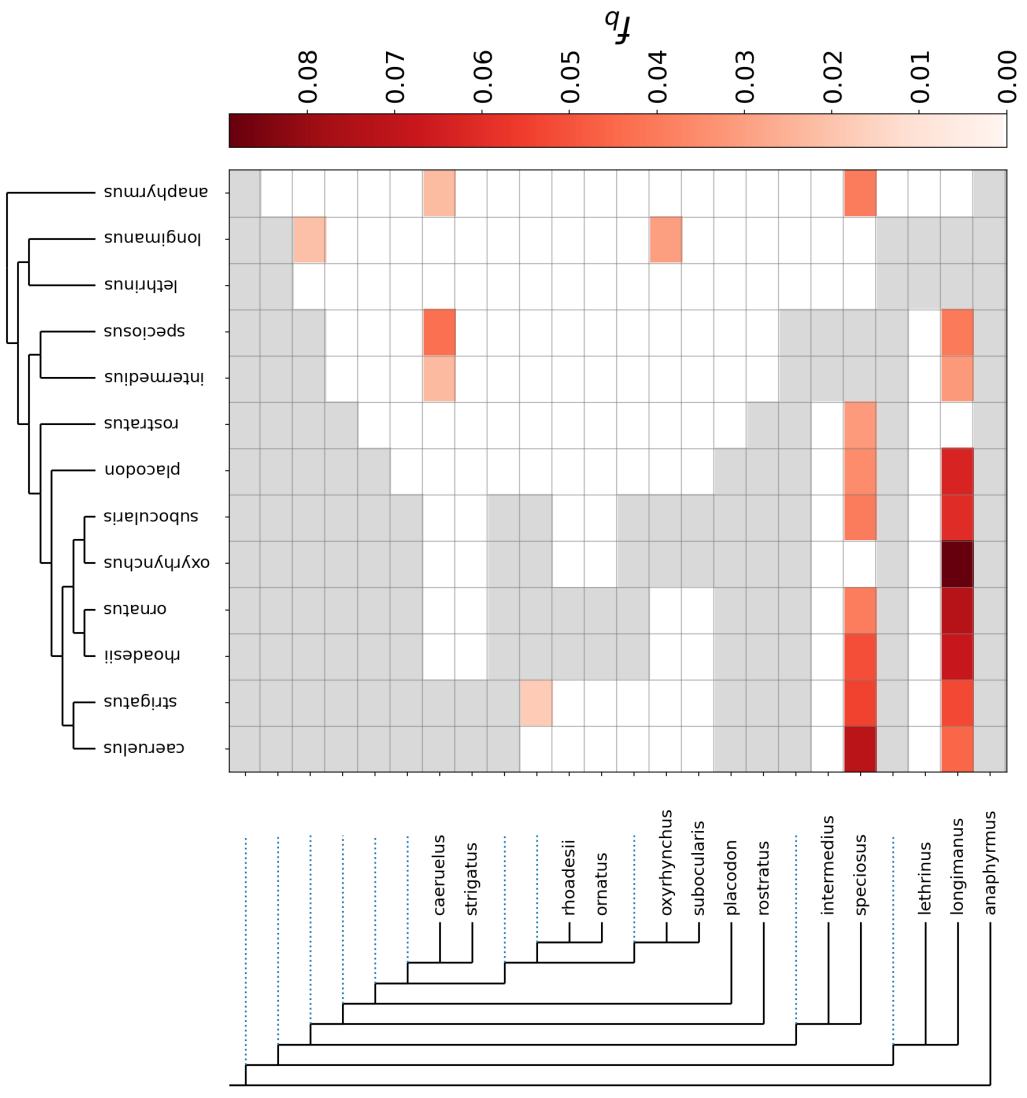

Figure I.19. fbranch result linkage group 18

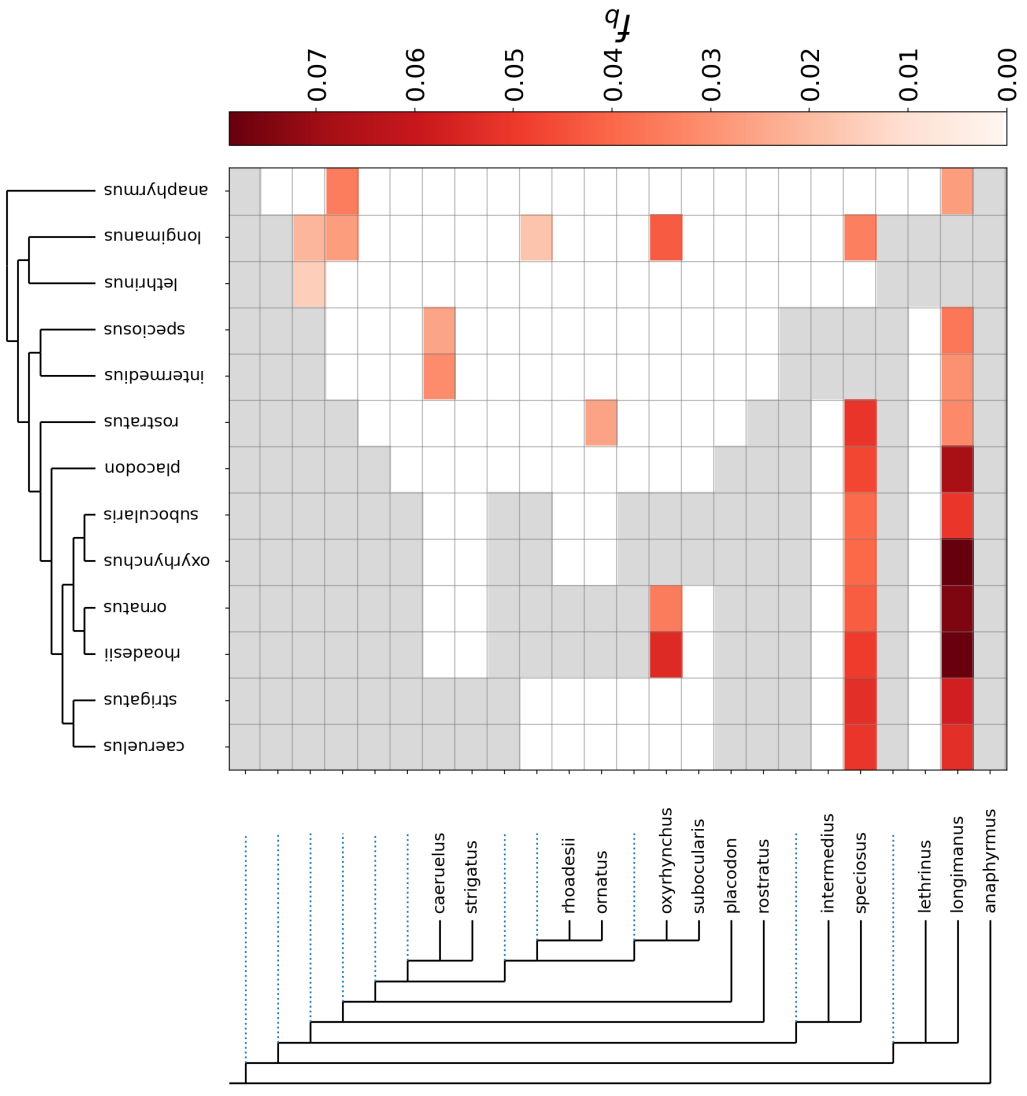

Figure I.20. fbranch result linkage group 19

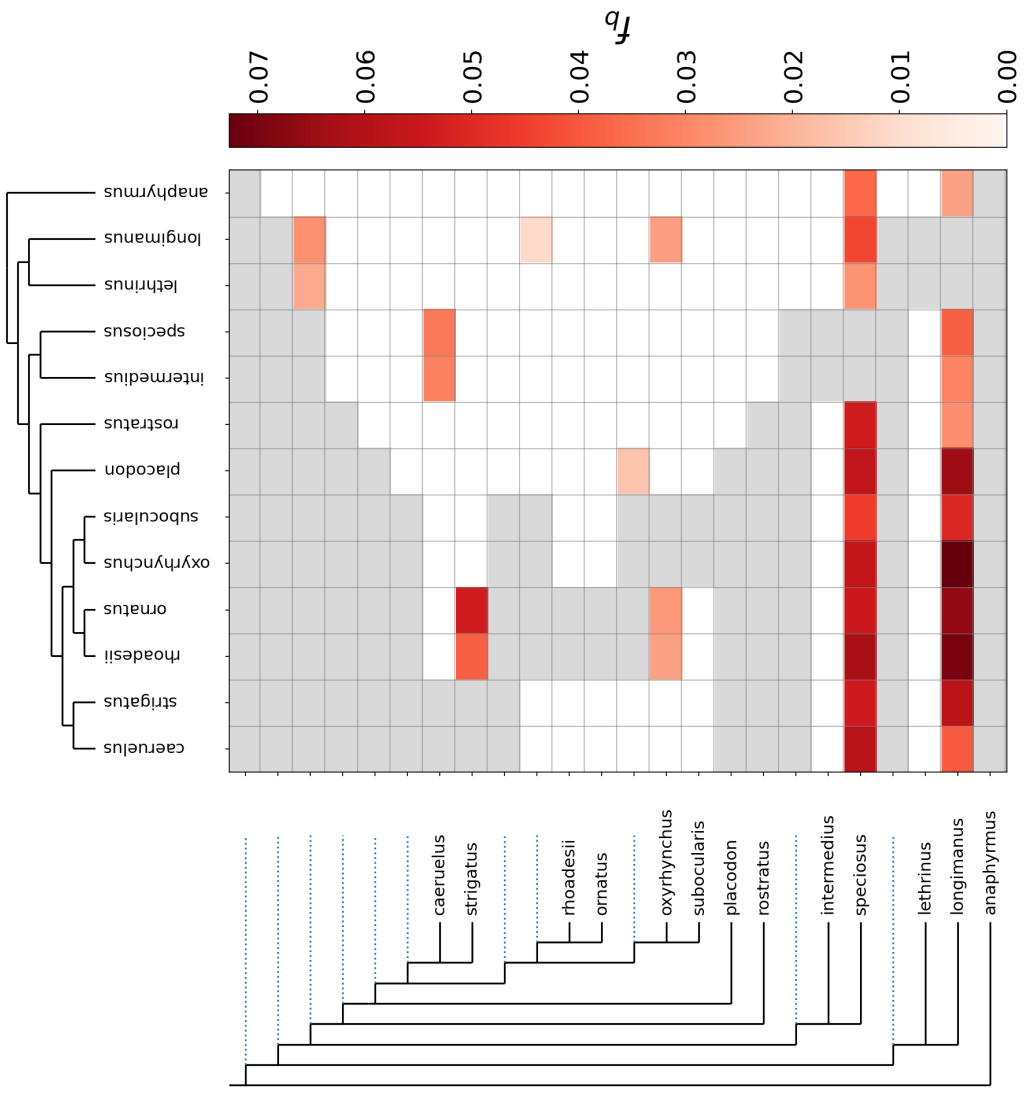

Figure I.21. fbranch result linkage group 20

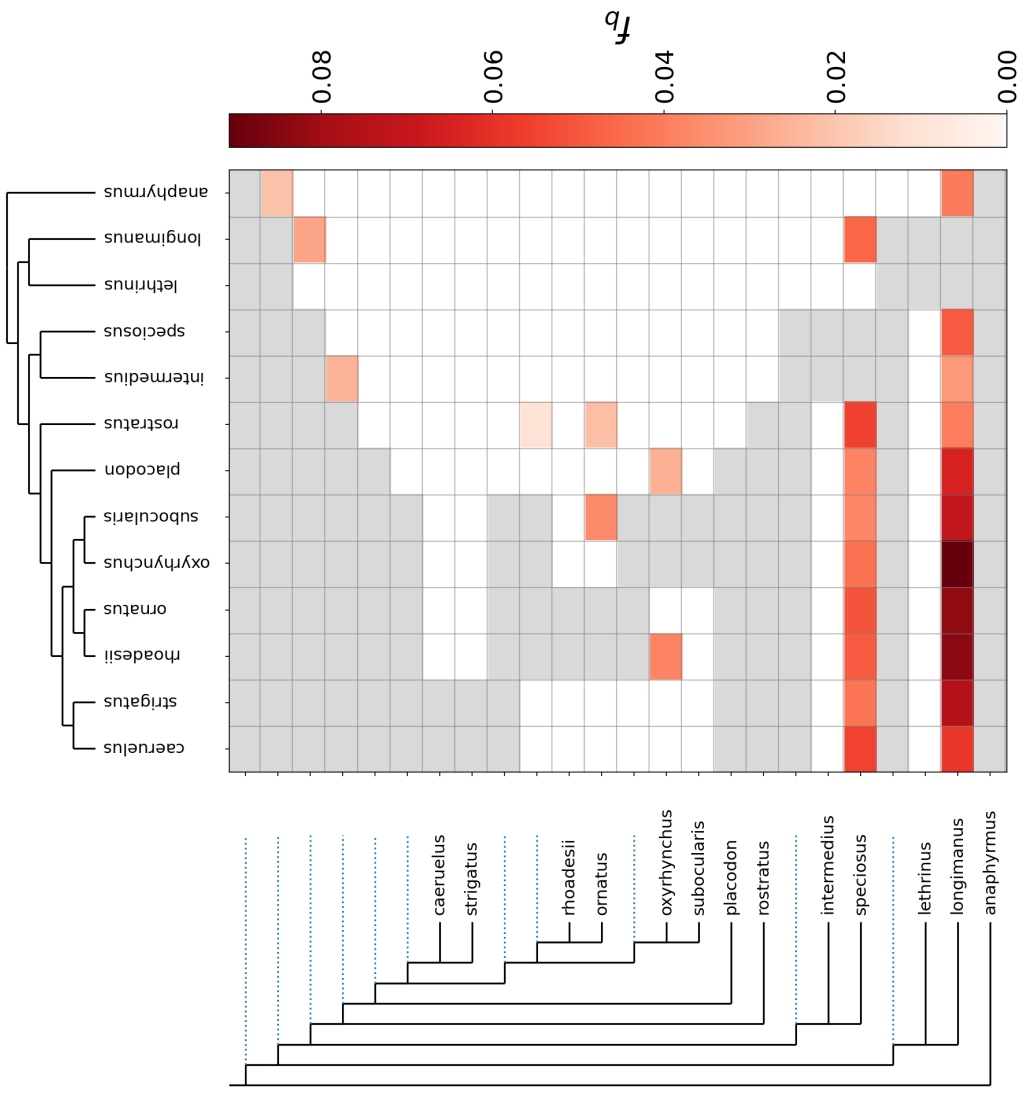

Figure I.22. fbranch result linkage group 22

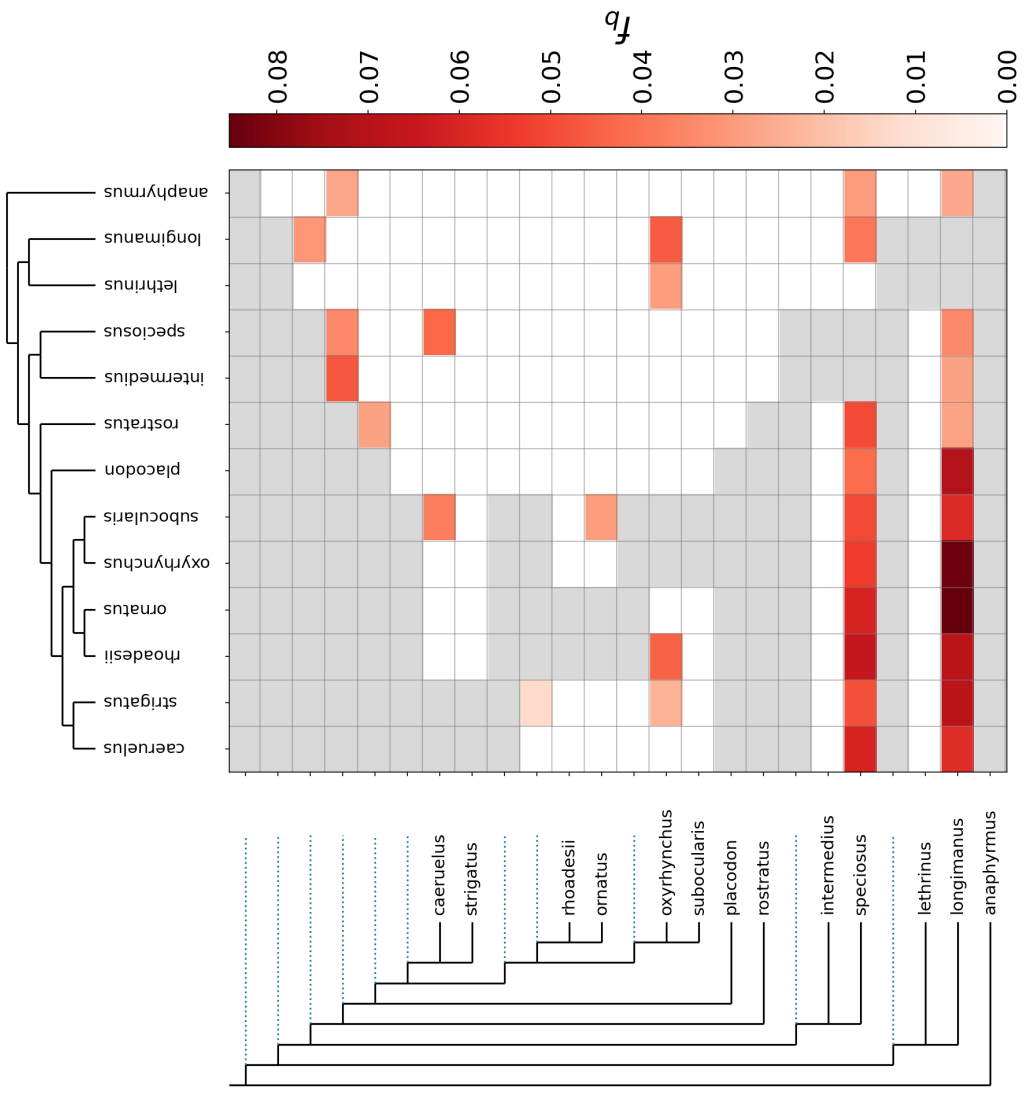

Figure I.23. fbranch result linkage group 23