POPULATION DIVERGENCE AND THE EVOLUTION OF REPRODUCTIVE ISOLATION IN *DROSOPHILA SUBQUINARIA*

WILLIAM JARVIS

Thesis submitted to the University of Ottawa
in partial fulfillment of the requirements for the
Doctorate in Philosophy degree (Ph.D.) in Biology

Department of Biology
Faculty of Science
University of Ottawa

Abstract

To better understand the sources of biological diversity we find in nature, we need insight on the mechanisms underlying the evolution of reproductive isolation and population divergence. Biological systems with patterns of naturally occurring variation in reproductive isolation can provide insight into mechanisms underlying its evolution, the genetic architecture of diverging traits, and the influence of genetic constraints on responses to selection. In this thesis, I used both experimental evolution and quantitative genetic approaches to study these questions in a system of reproductive character displacement in the North American mushroom-feeding fly *Drosophila subquinaria*. This species exhibits a pattern of stronger reproductive isolation against the closely related *Drosophila recens* in sympatry, where both female mating preferences and male chemical signaling traits have diverged from the ancestral allopatric populations. In chapter 2, we recreated this reproductive character displacement in the lab with replicate experimental populations of allopatric *D. subquinaria*. In chapter 3 I found divergence between allopatric and sympatric *D. subquinaria* in the genetic correlations among a set of these chemical signalling traits. In chapter 4, I found evidence indicating that this pattern of genetic correlations has played a key role in shaping the divergence of these traits, and that further adaptation in this system is likely constrained. Future research in this system would benefit from identifying the ancestral pattern of genetic correlations among these traits to better understand their stability over evolutionary time and from tracking the evolution of genetic correlations under controlled experimental conditions where the presence and absence of the sympatric species, *D. recens*, is manipulated.

Résumé

Pour mieux comprendre les origines de la diversité biologique que l'on trouve dans la nature, il faut comprendre les mécanismes qui déterminent l'évolution de l'isolement reproductif et de la divergence des populations. Les systèmes biologiques présentant des modèles de variation naturelle de l'isolement reproductif peuvent nous éclairer sur les mécanismes qui sous-tendent son évolution, sur l'architecture génétique des caractères divergents et sur l'influence des contraintes génétiques sur les réponses à la sélection. Dans cette thèse, j'ai utilisé à la fois l'évolution expérimentale et des approches de génétique quantitative pour étudier ces questions dans un système de déplacement des caractères reproductifs chez la mouche mangeuse de champignons nord-américaine *Drosophila subquinaria*. Cette espèce a évolué vers un isolement reproductif plus fort par rapport à la mouche étroitement apparentée, de manière sympatrique, *Drosophila recens*, où les taux d'accouplement des femelles et les caractéristiques de signalisation chimique des mâles ont divergé par rapport aux populations allopatriques ancestrales. Dans le chapitre 2, nous avons recréé ce déplacement des caractères reproductifs en laboratoire avec des populations expérimentales répliquées de *D. subquinaria* allopatrique. Dans le chapitre 3, j'ai trouvé des divergences entre *D. subquinaria* allopatrique et sympatrique dans les corrélations génétiques entre un ensemble de ces traits de signalisation chimique. Dans le chapitre 4, j'ai trouvé des preuves indiquant que ce modèle de corrélations génétiques a joué un rôle clé dans le façonnement de la divergence de ces traits, et que l'adaptation future dans ce système est probablement limitée. Les recherches futures sur ce système devraient se concentrer sur l'identification du modèle ancestral des corrélations génétiques entre ces traits afin de mieux comprendre leur stabilité au cours de l'évolution ou sur le suivi de l'évolution des corrélations génétiques dans des conditions expérimentales contrôlées où la présence et l'absence de l'espèce sympatrique, *D. recens*, sont manipulées.

Acknowledgements

First and foremost, a huge thank you to my supervisor Dr. Howard Rundle for your constant support, encouragement, and advice. It has been a pleasure and an honour working with one of the finest scientists I've known. Thank you for always letting me pursue my interests, and for making sure that I was financially supported throughout my program, so that I could focus on the research.

Thank you to all the members of the Rundle Lab, past and present, for your help and friendship. Especially, thank you to Kevin Kwok, Mathieu Videlier, Chris Angell, Li Yun, Cameron Kendrick, Tristan Ducharme, Kehinde Osijo, Erica Vong, Imani Badji, Mathew Osbourne, Catherine Sow, and Tia Chen. Thank you to Tim Wheeler and Lisa Sims for their assistance in the field, the work would not have been possible without you.

A huge thank you as well to Dr. Kelly Dyer, Dr. Jacqueline Sztepanacz, Julie Colpitts, and my committee members Vincent Carreau and Julien Martin for being such dependable and encouraging collaborators.

Thank you to my funding sources which made this work possible. In particular, the Natural Sciences and Engineering Research Council of Canada (NSERC Canadian Graduate Scholarship), the Ontario Graduate Scholarship (OGS), and the Society for the Study of Evolution (R.C. Lewontin Early Award).

A massive thank you to my friends and family, I could not have done it without you. Thanks to Eric, Joey, and Matt for getting it. Thank you Mum, Dad, Mac, Jane, Jim, Katie, Alex, Theo, and my buddy Goose.

And finally, thank you so much to my wife Sarah, my best friend.

Table of Contents

List of Figures

List of Tables

Chapter 1: General Introduction

Evolutionary biology has made extraordinary advances over the last 20 years, from the genomics revolution to the explosion in the quantity and quality of biological data. However, there is still so much we don't fully understand about the evolution of our biosphere, past, present, and especially future (Losos et al. 2013). It's likely a feature of the boundless complexity and variation inherent in natural systems, but evolutionary biologists still struggle to find truly general laws in nature (Welch 2017). For instance, while we know that divergent selection among populations can lead to phenotypic diversification (Rundle et al. 2000; McKinnon et al. 2004; Estes and Arnold 2007; Barrett et al. 2008; Schluter 2009), our ability to predict any specific evolutionary trajectory is limited by our understanding of the adaptive landscape and the genetic architecture of the traits under selection (Arnold et al. 2001; Kirkpatrick 2009). Although technological advances in computing have helped us to better model evolutionary processes, there will always be a need for data and experimental work with natural biological systems to test these models. By studying cases of diverging populations in nature, we can gain insight into mechanisms underlying the evolution of reproductive isolation, the genetic architecture of diverging traits, and the influence of genetic constraints on responses to selection. Using both experimental evolution and quantitative genetic approaches, I am studying these questions in a system of reproductive character displacement in the North American mushroom-feeding fly *Drosophila subquinaria*.

Our understating of speciation, a process where populations diverge into groups that no longer exchange alleles, is one area where data from natural systems has proven

invaluable. Speciation has long been a central focus for evolutionary biologists (Coyne and Orr 2004) and studies have examined all aspects of the speciation process, including the rate at which reproductive isolation (RI) accumulates (reviewed in Edmands 2002; Gourbière and Mallet 2010; Coughlan and Matute 2020), the role of ecological divergence in generating RI (Funk et al. 2006), the relative importance of population bottlenecks in speciation (Rundle et al. 1998; Templeton 2008; Matute 2013), and whether selection can strengthen RI in sympatry (Yukilevich 2012; Nosil 2013). The answer to this last question appears to be yes, with multiple examples of reproductive character displacement where RI is stronger in sympatry than allopatry (Gerhardt 1994; Noor 1995; Saetre, G.P. 1997; Rundle and Schluter 1998; Jiggins et al. 2001; Matute 2010b), but how often this occurs and its relative importance for speciation remain unclear (Matute and Cooper 2021a). The typical explanation for this phenomenon, termed reinforcement, is that prezygotic isolation between divergent taxa is strengthened by selection arising from reduced hybrid fitness. There are other processes, however, that can generate patterns of enhanced isolation in sympatry (Noor 1999; Coyne and Orr 2004). For example, ecological character displacement, where competition for limited resources generates selection, can result in increased prezygotic isolation if the traits under selection also impact reproductive interactions (Dodd 1989; Rice and Salt 1990; Rundle and Schluter 1998; Pfennig and Pfennig 2009). Alternatively, if taxa with stronger prezygotic isolation are more likely to persist in sympatry in the face of gene flow while those with weaker isolation collapse into a hybrid swarm, then an apparent pattern of reproductive character displacement can arise without any evolution of prezygotic isolation (Templeton 1981; Noor 1999). However, we still have limited data

on the underlying mechanisms driving the evolution of RI following secondary contact, and the rate at which RI evolves when it does (Matute and Cooper 2021a).

Experimental evolution, where the evolutionary responses of replicate populations to controlled manipulations of biotic and/or abiotic conditions, has proven a powerful approach for gaining insight into the mechanisms underlying the evolution of RI. For example, 'kill the hybrids' experiments, where F1 hybrids between two populations are destroyed every generation, have found that genetic variation for prezygotic RI is often present (reviewed in Fry 2009). However, these studies provide limited insight into speciation itself because experimental populations in this case are already fully isolated (i.e., hybrid fitness is zero and hence speciation is complete). There have been very few evolution experiments testing the consequences of secondary contact in which hybrids were not killed by design, and that also had relevance for cases of reproductive character displacement in nature. Working with *Drosophila serrata*, Higgie et al. (2000) showed that a naturally occurring pattern of reproductive character displacement of sexual display traits—contact pheromones consisting of cuticular hydrocarbons—re-evolved in allopatric populations that experienced experimental sympatry in the lab, although the impact on prezygotic isolation was not tested. Working with *D. yakuba* and *D. santomea*, Matute (2010b) manipulated the degree of hybridization and showed that enhanced prezygotic isolation evolved in the presence of some gene flow (i.e., when hybrid fitness was greater than zero). With so few studies of this kind, key questions remain concerning the processes that generate patterns of reproductive character displacement in nature, in particular the frequency and importance of sympatry in strengthening prezygotic RI (Coughlan and Matute 2020; Matute and Cooper 2021a).

Untangling the extent to which the vast array of biological variation we observe in nature is adaptive is another key focus of evolutionary biology. A narrow reading of the literature might lead one to conclude that, since there is available genetic variance for most traits (Mousseau and Roff 1987), any given trait is likely to respond rapidly to selection. Indeed, this appears to be the case with many long-term artificial selection programs (Moose et al. 2004; Powell and Norman 2006). However, it is likely that selection in nature rarely acts on single traits alone, instead acting on multivariate trait combinations that, due to pleiotropic effects and linkage disequilibrium generating genetic correlations, will necessarily be fewer than the number of traits (Gould 1966; Lande 1979; Lande and Arnold 1983; Blows 2007). This complicates our ability to predict the evolutionary trajectories of diverging populations, as genetic variation may be constrained by the underlying pattern of genetic correlations, which can concentrate variance into phenotypic axes that may not align with the direction of selection (Blows and Walsh 2009; Walsh and Blows 2009; Blows and McGuigan 2015). It seems that in nature evolutionary stasis is the predominant state of most populations (Gould 2002), despite many examples of rapid evolutionary responses (Hairston et al. 2005). Further complicating matters, as multivariate traits evolve, the underlying genetic architecture itself may change, generating potentially complex feedback. Clarifying how genetic (co)variances evolve and influence diversification is critical to our understanding of the fundamental processes of adaptive evolution and speciation.

To study the evolution of genetic constraints and their effects on diversification we must turn to the additive genetic variance-covariance (**G**) matrix, which summarizes the genetic architecture of a suite of traits. Specifically, **G** is a symmetrical matrix of

additive genetic variances for a set of phenotypic traits along the diagonal (i.e., the variance in breeding values, where an individual's breeding value is the sum of the average allelic effects across its loci) and genetic covariances on the off diagonal (i.e., the covariance between a pair of traits due to a combination of pleiotropic effects and linkage disequilibrium). The linear combination of traits with the greatest genetic variance, which is also the leading eigenvector of **G**, is known as $\mathbf{g}_{max}$. This trait combination can bias evolutionary responses by providing a 'genetic line of least resistance' (Schluter 1996). The stability of the **G** matrix, and the orientation of $\mathbf{g}_{max}$, are thus important problems, and ones that are difficult to model without making very specific assumptions (Arnold et al. 2008).

Empirical studies of **G** matrix evolution have varied substantially in their approaches, as new statistical techniques for comparing matrices have emerged. Although several early studies (reviewed in Steppan et al. 2002; Arnold et al. 2008), including experimental manipulations (Blows and Higgie 2003; Hine et al. 2011), suggested that **G** can indeed evolve over short time scales, there were some limitations to the statistical frameworks employed (Aguirre et al. 2014). Specifically, in the case of element-by-element comparisons or correlations of principal components, there is not always a clear evolutionary relevance as the link between correlated matrix elements and responses to selection is not directly evident (Hansen and Houle 2008). Additionally, because of the logistical challenges of obtaining robust quantitative genetic datasets, studies have often been restricted to comparisons between a single pair of populations and in the absence of any estimates of selection for the traits in question. Lastly, comparisons of **G** matrices have tended to lack any measure of uncertainty around estimates of **G,** as well as around

the comparisons themselves. Aguirre et al. (2014) tested several statistical methods that addressed one or more of these concerns and demonstrated that a tensor approach (i.e. employing a 4th order genetic covariance tensor) was a robust and generally applicable method of **G** comparison (see Hine et al. 2009; Morrissey et al. 2019 as well). The tensor approach has been used to identify divergence in genetic (co)variances among ecotypes of an Australian native wildflower (Walter et al. 2018), as well as among-population variation in **G** across the range of *Parus major* (Chantepie et al. 2024). In contrast, Henry and Stinchcombe (2023) found **G** to be mostly stable across a cline in an annual vine *I. hederacea*, as did Delahaie et al. (2017) when assessing among-population variation in **G** in blue tits (*Cyanistes caeruleus*).

While the stability of **G** over short to moderate timescales is still uncertain, it is clear that variation in **G** among populations needs to be accounted for when assessing the effect of genetic constraints on evolutionary trajectories. To do this, any quantitative assessment of the influence of genetic constraints on an observed pattern of divergence should incorporate estimates of selection and genetic (co)variances that allow these parameters to vary among populations. Building upon a model proposed by Zeng (1988), Chenoweth et al. (2010) developed such an approach by making use of among-population datasets on phenotypic divergence, selection, and genetic (co)variance (**G**). These can be used to calculate a set of symmetric among-population variance-covariance matrices that capture variation in phenotypic trait means (**D**), in sexual selection gradients (**B**), and in predicted responses to selection while accounting for genetic (co)variances (**R**). By comparing **D** to each of these other matrices, one can assess whether the observed pattern of phenotypic divergence is best explained by selection alone or by filtering selection

through the genetic (co)variance of those traits. The difficulty of creating robust datasets for each of these parameters has meant that, to my knowledge, this approach has not been fully implemented beyond the original application in Chenoweth et al. (2010).

Biological systems with patterns of naturally occurring variation in reproductive isolation can provide insight into mechanisms underlying its evolution, the genetic architecture of diverging traits, and the influence of genetic constraints on responses to selection. Using both experimental evolution and quantitative genetic approaches, I studied these questions in a system that exhibits a pattern of reproductive character displacement. The species is the North American mycophagous fly, *Drosophila subquinaria*. Populations of *D. subquinaria* that are either allopatric or sympatric to the related species *Drosophila recens* exhibit a pattern where females from sympatric populations are more likely to reject mating with heterospecific males than are allopatric females (Jaenike et al. 2006; Dyer et al. 2018). This reproductive character displacement is consistent with the strengthening of RI by selection in sympatry as, in addition to genetic incompatibilities causing F1 hybrid male sterility, females that mate with male *D. recens* suffer roughly 90% mortality of their offspring due a cytoplasmic incompatibility caused by a Wolbachia infection carried by around 98% of *D. recens* (Jaenike et al. 2006). Some introgression must occur via *D. recens* female × *D. subquinaria* male matings as roughly 2.5% of *D. subquinaria* harbor a *D. recens* mtDNA haplotype (Jaenike et al. 2006; Bewick and Dyer 2014). These species diverged relatively recently, but have come into secondary contact within the last 12,000 years since the end of the Wisconsin glaciation and are often found feeding and mating on the same mushrooms in the sympatry zone just the east of the Rocky Mountains (Shoemaker et al. 1999; Jaenike

et al. 2006). Selection against heterospecific matings appears to have fixed rejection by *D. subquinaria* females in sympatry, but genetic variation for this behavior remains in allopatric populations.

There are limited morphological differences between *D. subquinaria* and *D. recens* apart from internal male genitalia, but the species can be distinguished by the mix of long-chain hydrocarbons found on their cuticles (termed CHCs). The relative concentration of each compound is a quantitative trait, and collectively they form a multivariate CHC phenotype that differs between sexes, populations, and species (Curtis et al. 2013). CHCs are a means of moisture control, regulating the permeability of the cuticle, and also act as contact pheromones that play a key role in the primarily chemosensory mating interactions in this species (Dyer et al. 2013; Giglio and Dyer 2013). Male CHCs, and female mate preferences for them, have diverged among *D. subquinaria* populations in a pattern of reproductive character displacement that matches the observed variation in sexual isolation between sympatry and allopatry (Dyer et al. 2013; Rundle and Dyer 2015). Although it seems likely that divergent sexual selection between regions has led to the observed pattern of phenotypic divergence in CHCs, there has not been a comprehensive examination of the mechanism underlying the evolution of reproductive character displacement, for both sexual isolation and CHC phenotypes, as well as the relative role of genetic architecture for these traits in structuring divergence, and the stability of genetic architecture during this process.

In chapter 2 of my thesis, I use data from an experimental evolution study to illuminate the mechanisms driving the evolution of reproductive isolation and CHC divergence in *D. subquinaria* during secondary contact. Allopatric *D. subquinaria* were

brought into experimental sympatry with *D. recens* in both a “kill the hybrids” treatment following previous studies and a treatment that allowed the hybrids to live. The heterospecific mating rate and multivariate CHC phenotype in each treatment were tracked over 12 generations to observe how these traits respond to sympatric conditions. In chapter 3, I explore the genetic architecture of these CHC traits, its stability during the reproductive character displacement observed in nature, and its alignment with multivariate phenotypic divergence. In chapter 4, I assess how genetic architecture and sexual selection have interacted to create the observed pattern of phenotypic divergence, and the potential influence of genetic constraints on future adaptation in this system.

Chapter 2: An experimental test of the evolutionary consequences of sympatry in *Drosophila subquinaria*

The following thesis chapter is a modified version of a manuscript that has been published in *Evolution*. The article was modified to meet thesis formatting requirements. This work was done in collaboration with Kelly Dyer, Nicholas Arthur, and my supervisor, Howard Rundle. I undertook all cuticular hydrocarbon (CHC) sample processing and the analysis of all data (mating rates and CHCs), and I took the lead role in interpreting the results and writing the manuscript.

Article Citation:
Will M C Jarvis, Nicholas J Arthur, Howard D Rundle, Kelly A Dyer. 2024. An experimental test of the evolutionary consequences of sympatry in *Drosophila subquinaria*. Evolution. 78(3): 555–565 https://doi-org.proxy.bib.uottawa.ca/10.1093/evolut/qpad236

2.1 Introduction

Patterns of naturally occurring variation in reproductive isolation can provide insight into mechanisms underlying its evolution, and hence our understanding of speciation. For example, prezygotic isolation is often stronger between sympatric as opposed to allopatric taxa (Coyne and Orr 1989, 1997; Yukilevich 2012; Nosil 2013; Matute and Cooper 2021a), and pairs of taxa with partially overlapping geographic ranges frequently show enhanced isolation in sympatry, a pattern termed 'reproductive character displacement' (Gerhardt 1994; Noor 1995; Sætre et al. 1997; Rundle and Schluter 1998; Jiggins et al. 2001; Matute 2010b). Enhanced prezygotic isolation in sympatry is often interpreted as the product of selection generated by costly heterospecific reproductive interactions. Attention has focused on reinforcement, where prezygotic isolation between divergent taxa is strengthened during secondary contact due to selection arising from reduced hybrid fitness. There are other processes, however, that can generate patterns of enhanced isolation in sympatry (Noor 1999; Coyne and Orr 2004). For example, competition for shared, limiting resources can cause ecological character displacement and, if the traits under selection also impact reproductive interactions, increased prezygotic isolation could evolve as a side-effect (Dodd 1989; Rice and Salt 1990; Rundle and Schluter 1998; Pfennig and Pfennig 2009). Alternatively, if taxa with stronger prezygotic isolation are more likely to persist in sympatry than are weakly isolated ones (which either fuse or go extinct), then an apparent pattern of reproductive character displacement can arise without any evolution of prezygotic isolation following secondary contact (Templeton 1981; Noor 1999).

While elegant approaches have been used in attempts to distinguish among these processes (e.g., Rundle and Schluter, 1998; van der Niet et al., 2006; Hopkins & Rausher, 2012; Yukilevich, 2012; Nosil, 2013), this can nevertheless be challenging via comparative studies alone. A powerful approach for gaining insight into the mechanisms underlying the evolution of reproductive isolation is to track the evolutionary responses of replicate populations to controlled manipulations of biotic and/or abiotic conditions. Such 'experimental evolution' studies have made fundamental contributions to our understanding of speciation, for example demonstrating that adaptation to different environments tends to result in stronger reproductive isolation than independent adaptation to similar environments, that population bottlenecks on their own seldom generate reproductive isolation, and that disruptive selection on a single population is not conducive to speciation unless the traits involved are directly linked to mating behavior (Rice and Hostert 1993; Coyne and Orr 2004; Fry 2009; White et al. 2020).

Experimental evolution has also been used to study the consequences of secondary contact between divergent taxa. For example, a long history of 'kill the hybrids' experiments, in which all F1 hybrids between two populations are destroyed each generation, has revealed that genetic variation for prezygotic isolation is often present, and that this can evolve in sympatry when gene flow is prevented entirely (reviewed in Fry 2009). However, many of these studies are poorly replicated, lack controls, and/or do not address hypotheses about actual patterns of reproductive isolation in nature. They also provide limited insight into speciation because experimental populations are already fully reproductively isolated (i.e., hybrid fitness is zero and hence speciation is complete). There have been very few evolution experiments testing the

consequences of secondary contact in which hybrids were not killed by design and that had relevance for patterns of reproductive isolation in nature. Working with *Drosophila serrata*, Higgie et al. (2000) showed that a naturally occurring pattern of reproductive character displacement of sexual display traits—contact pheromones consisting of cuticular hydrocarbons—re-evolved in allopatric populations that experienced experimental sympatry in the lab, although the impact on prezygotic isolation was not tested. And working with *D. yakuba* and *D. santomea*, Matute (2010b) manipulated the degree of hybridization and showed that enhanced prezygotic isolation evolved in the presence of some gene flow (i.e., when hybrid fitness was greater than zero). With few such studies, key questions remain concerning the processes that generated existing patterns of reproductive character displacement in nature (Coughlan and Matute 2020; Matute and Cooper 2021a).

Here I use experimental evolution to test the consequences of secondary contact in the North American mushroom-feeding fly *Drosophila subquinaria*. This is an ideal system for doing so; *D. subquinaria* occurs in western N. America and, in an area just east of the Rocky Mountains, its range partially overlaps that of the closely related eastern species, *Drosophila recens*. Evidence suggests that this region of sympatry is a secondary contact zone within the last 12,000 years since the end of the Wisconsin glaciation (Jaenike et al. 2006). *D. subquinaria* exhibits a pattern of reproductive character displacement in which prezygotic sexual (i.e., behavioral) isolation with *D. recens* is stronger in sympatry than in allopatry (Jaenike et al. 2006; Bewick and Dyer 2014; Dyer et al. 2018). In laboratory assays in which *D. subquinaria* females are confined with *D. recens* males, females from sympatric populations never mate whereas

females from outside this region (i.e., allopatric) mate with *D. recens* males on average 35% of the time.

Strong but incomplete postzygotic isolation also exists between *D. subquinaria* and *D. recens*. Genetic incompatibilities cause F1 hybrid male sterility in both directions of crosses between these species. In addition, cytoplasmic incompatibility from a *Wolbachia* infection in *D. recens* causes a further ~90% offspring mortality when *D. subquinaria* females mate with *D. recens* males (Shoemaker et al. 1999). Nevertheless, genetic evidence indicates that occasional hybridization, and some gene flow, occurs in the wild. For example, about 3% of wild-caught *D. subquinaria* harbor a *D. recens* mtDNA, indicating mating of *D. recens* females with *D. subquinaria* males and then backcrossing into *D. subquinaria* (Jaenike et al. 2006). At nuclear loci, phylogenetic analyses and demographic modelling suggest that ongoing gene flow, primarily from *D. recens* into *D. subquinaria*, creates patterns of incomplete lineage sorting and phylogenetic incongruence (Dyer et al. 2018; Ginsberg et al. 2019).

Female mate choice and species discrimination in *D. subquinaria* are based, at least in part, on chemosensory cues, with females assessing a suite of male contact pheromones consisting of (epi)cuticular hydrocarbons or CHCs (Curtis et al. 2013; Dyer et al. 2013; Giglio and Dyer 2013). Male (and female) CHCs, and female mate preferences for them, have diverged among *D. subquinaria* populations in a pattern of reproductive character displacement that echoes the displacement of sexual isolation (Dyer et al. 2013; Rundle and Dyer 2015). To examine whether these patterns of reproductive character displacement in *D. subquinaria* could be an evolutionary response

to secondary contact with *D. recens*, I ran a multi-generation evolution experiment in which replicate *D. subquinaria* populations, originally from allopatry, evolved either in the presence or absence of *D. recens* individuals during mating each generation, creating control vs. 'experimental sympatry' treatments. To gain further insight, I also attempted to vary the strength of postzygotic isolation in the experimental sympatry treatment (i.e., F1 hybrid fitness and hence the opportunity for gene flow), either allowing hybrids that survived to adulthood to live (*hybrids live* treatment) or killing them each generation (*hybrids killed* treatment). The *hybrids live* treatment reflects the current state of prezygotic isolation between these species in natural sympatric populations, and I expected some hybridization to occur given evidence of gene flow in nature and results of previous laboratory mating assays. In both treatments, I expect the actual number of hybrid adult offspring produced to be low, as ~90% of hybrids die as eggs due to cytoplasmic incompatibility. I periodically assayed CHCs and the rate at which offspring were formed in heterospecific mating trials as an index of sexual isolation (i.e., female *D. subquinaria* mating rates with male *D. recens*) over 12 generations of experimental evolution in 16 replicate populations under these three treatments. If selection arising from the presence of *D. recens* is the underlying cause of the pattern of reproductive character displacement, one would predict that female *D. subquinaria* mating rates with *D. recens* would decrease at each sampling generation in both experimental treatments, and that male CHC profiles would evolve to differ from the allopatric control.

Studies have shown that when sympatry between two taxa leads to the evolution of enhanced prezygotic isolation between them, the effects of this can cascade within a taxon, generating prezygotic isolation between the sympatric and allopatric populations

as a by-product (Howard 1993; Hoskin et al. 2005; Ortiz-Barrientos et al. 2009; Hoskin and Higgie 2010). Consistent with this, sympatric *D. subquinaria* females discriminate against conspecific males from populations allopatric to *D. recens* (Jaenike et al. 2006; Bewick and Dyer 2014). In an attempt to gain insight into this potential side-effect of evolution in sympatry, I also tracked sexual isolation of our experimental populations against their ancestral (allopatric) *D. subquinaria* population.

2.2 Methods

Stocks and fly maintenance

A stock population of *D. subquinaria* was created through round-robin crosses over six generations using 15 isofemale lines previously collected from allopatry (Missoula, Montana; 2009-2011). This population was previously characterized as harboring genetic variation for discrimination against *D. recens* (Bewick and Dyer 2014). A *D. recens* stock population was generated in the same way from 11 isofemale lines collected from sympatry (Hastings Lake, Alberta; 2009-2011). Separate marked mutant populations of both the *D. subquinaria* and *D. recens* stocks were subsequently generated via six rounds of backcrossing of a recessive autosomal *dark-eye* mutation into each, maintaining >200 flies each generation to prevent inbreeding. The *dark-eye* phenotype involves a different locus in each species such that F1 hybrids between these 'mutant' populations have wild-type eye color. Stocks were subsequently maintained by mass transfer of 500+ individuals each generation on Instant *Drosophila* Medium (Carolina Biological Formula 4-24), supplemented with commercial mushroom *Agaricus bisporus*. Flies were maintained at 20 °C under a 12- hour light/dark cycle and

60% relative humidity. Virgins were collected using light CO_2 anesthesia and held at a density of <20 flies per vial.

Experimental evolution

Sixteen experimental populations were derived from the two *D. subquinaria* stocks (the wild-type or dark-eye) and were assigned to one of three experimental treatments, yielding four allopatric control populations (i.e., *D. recens* absent), six experimental sympatry populations with *D. recens* present during mating and in which hybrids were allowed to live (*hybrids live*), and six experimental sympatry populations with *D. recens* present during mating and in which all hybrid offspring were killed each generation (*hybrids killed*; Fig. S2.1). Half the populations within each treatment were maintained on a schedule offset by two weeks, generating two experimental blocks that were balanced by treatment. Allopatric control and sympatric *hybrids live* populations were derived from the wild-type *D. subquinaria* stock, while those in the sympatric *hybrids killed* treatment were derived from the *dark-eye* mutant *D. subquinaria* stock. Likewise, sympatry in the *hybrids live* treatment used flies from the wild-type *D. recens* stock, while the *hybrids killed* treatment used flies from the *dark-eye* mutant *D. recens* stock. This ensured that males of the two species had the same eye color within a given treatment and avoided segregating eye mutations in the event of gene flow in the *hybrids live* treatment, preventing eye color from impacting mating patterns while allowing hybrid offspring to be identified and removed in the *hybrids killed* treatment. For simplicity, I refer to experimental populations as *D. subquinaria*, recognizing that introgression into them from *D. recens* is possible in the *hybrids live* treatment.

(Introgression into *D. recens* is not possible because the *D. recens* flies used during experimental sympatry were drawn anew from the appropriate *D. recens* stock each generation and then discarded after use.)

Experimental populations were maintained via non-overlapping generations by allowing 10 females to oviposit for seven days in each of 10 standard *Drosophila* bottes, after which they were discarded. Virgin offspring were collected within 24 hours of emergence, at which point hybrid individuals, *i.e*., adult flies with wild-type eye color, were discarded in the *hybrids killed* treatment after recording the proportion of hybrid offspring. After rearing virgins to sexual maturity (7-11 days old), a 4-day mating phase began in which females were placed in their respective treatments, either with or without male and female *D. recens* from the appropriate stock (Fig. S2.1). Four replicate cages were created for each population; each cage in the experimental sympatry treatments contained 40 experimental males, 40 experimental females, 40 *D. recens* males, and 40 *D. recens* females. For allopatric control populations, 80 male *D. subquinaria* and 80 female *D. subquinaria* were used in each cage to maintain the same total density of males and females. Cages were plastic 32-ounce 'deli' containers (10 cm diameter × 13 cm high), with Instant *Drosophila* medium (Carolina Biological Formula 4-24) and a mushroom cap as a mating substrate. Flies in both experimental sympatry treatments were lightly coated with either yellow or pink fluorescent dust (Day Glo Color Corp., pigments A-11 and A-17-N), alternating colors by species between generations. Preliminary assays indicated that dust color did not affect mating rate or preference (data not shown). Following the mating phase, all males and *D. recens* females were identified using UV light and discarded, and 100 experimental '*D. subquinaria*' females

from each population were then transferred to fresh bottles (10 females/bottle) for oviposition to create the next generation, as described above.

Phenotypic Assays

As an index of sexual isolation, I performed no-choice mating trials in which experimental females were presented with heterospecific *D. recens* males or conspecific ancestral (i.e., allopatric) *D. subquinaria* males in separate assays (Fig. S2.2). Trials were conducted at generations 0 (i.e., prior to initiating the treatments), 6, and 12, as previously described (Bewick and Dyer 2014). For trials involving heterospecific *D. recens* males, single virgin females from each experimental population were placed with single *Wolbachia*-free, wild-type virgin *D. recens* males in standard food vials, and after 24 h the males were discarded. This *Wolbachia*-free *D. recens* stock was derived from the original wild-type stock and subsequently treated with tetracycline to remove the endemic *Wolbachia* infection for use in the mating trails. In crosses with *Wolbachia*-free males, there is no cytoplasmic incompatibility and larvae develop normally (Shoemaker et al. 1999). I scored for the presence of larvae 7 days after the male was removed. This was done both by visual inspection and by pouring a 20% sucrose solution into each vial; when larvae are present, they float to the top (Nichols et al. 2012). The mating rate between heterospecific flies is generally low, and a 24 h mating period allows a higher rate and thus provides increased power to detect any change. In theory, post-mating prezygotic barriers (e.g., differential sperm usage including cryptic female choice) could contribute to the presence of larvae in such trials, in addition to sexual isolation arising from pre-copulatory female choice, although I

view that an unlikely in the current case (see Chapter 2 Discussion). An average of 88.2 ± 15.9 SD trials were performed for each population at each sampling generation.

To assess changes in the mating rate of experimental females with conspecific ancestral *D. subquinaria* males, additional no-choice mating trials were conducted at generations 0, 6 and 12. In this case, individual females from the allopatric control and *hybrids live* treatments were paired with single males from the 'ancestral', wild-type allopatric *D. subquinaria* stock from which they were originally derived, while females from the *hybrids killed* treatment were individually paired with single males from the 'ancestral' *dark-eye* allopatric *D. subquinaria* stock from which they were originally derived. Trials took place in small vials (4.5 cm long × 1 cm diameter) that contained a blended mushroom-agar medium. Vials were observed and the number that mated within 2 h was recorded. The mating rates of conspecific flies is much higher than heterospecific pairings, thus a shorter mating period was used and mating itself could be observed. An average of 91.2 ± 14.1 SD trials were performed for each population for each sampling generation.

CHC profiles from individual 7-11 day old virgin flies were also sampled from each population at generations 0, 6 and 12, as described in Dyer et al (2014). Each population sample included an average of 15.3 ± 4.2 SD males and 16.5 ± 3.6 SD females. In brief, CHCs were extracted within 3 h of incubator lights turning on by immersing individual flies in 100 μl of hexane for ~ 4 min, after which they were discarded. Samples were stored at -20 °C and were subsequently shipped from Athens, GA, USA, to Ottawa, ON, CA. Samples of females from all generations for a single allopatric control population, and of males from sampling generation 0 of a different

allopatric control population, were lost (cause unknown) and hence were not included in the analyses. Gas chromatography was performed on an Agilent 6890N dual-channel 'fast' (220 V oven) gas chromatograph (Agilent Technologies, Wilmington, DE), employing flame ionization detection, following Curtis et al. (2013). Individual CHC profiles were determined by integrating the area under 11 peaks for females and 17 peaks for males (Fig. S2.3), representing a subset of the compounds identified by Curtis et al. (2013) that were consistently present in all current samples. To correct for technical error associated with estimating absolute concentrations, CHC abundances were expressed as relative proportions by dividing the area under each peak by total area under all peaks for a given individual. Relative concentrations are a form of compositional data to which standard statistical analysis should not be applied (Aitchison 1982). To address this, values were transformed to centered log ratios (CLRs) following Bonduriansky et al. (2015). Seven individual flies with multivariate outlier CHC profiles were removed prior to analysis based on extreme Mahalanobis distances, likely representing sample contamination or integration errors.

Statistical Analyses

Variation among treatments in the mating rate of experimental *D. subquinaria* females with heterospecific *D. recens* males, and with conspecific, ancestral *D. subquinaria* males, was analyzed using separate generalized linear mixed effects models with a binomial distribution and a logit link function given the nature of the response variable (i.e., the number of successful vs. unsuccessful matings). Fixed effects included treatment, sampling generation (a continuous variable), and their interaction. Random

effects included a factor with unique levels for each combination of experimental block and sampling generation (because mating assays were performed separately by block each generation), and population (to account for repeated sampling of the same population across generations), with all populations having unique IDs (i.e., 16 levels). Models were fit in R v. 3.6.1 (R Development Core Team 2020) using the *lme4* package (Bates et al. 2015), with significance of fixed effects determined via partial (i.e., type 3) Wald χ^2 tests using the *car* package (Fox et al. 2019). Given a significant treatment × sampling generation interaction for the heterospecific mating rate (see Results), a simplified generalized linear model was subsequently fit, separately by sampling generation, that included fixed effects of treatment and experimental block (2 levels). Significant treatment effects, when present, were further characterized via post hoc pairwise comparisons among treatment levels. This was done by calculating the odds ratios from the marginal means, with p-values adjusted for multiple comparisons via the Tukey procedure, as implemented in the *emmeans* package (Lenth 2021). Using the odds ratio allows us to account for unknown sources of environmental variation that causes among-assay (i.e., block and generation) variation in mating rate (see Fig. S3). As these factors affect the control and treatments similarly for any given assay in which these populations are all tested at the same time, this variation can be accounted for by expressing treatment effects relative to the paired allopatric control populations.

CHCs were analyzed separately in males and females because some compounds are sex-specific. A principal component analysis (PCA) of the covariance matrix of population mean values of CLR-transformed traits was performed using the generation 12 samples, thereby focusing on the among-population variation that had arisen by the

end of the evolution experiment. Results were qualitatively unchanged if the PCA was instead performed on population means calculated separately by sampling generation, indicating that the dominant axes of among-population variation were those present at generation 12 (results not shown). Working with principal components (PCs) has the advantage of reducing dimensionality, including removing the unit-sum constraint inherent in CLR-transformed data, and it avoids statistical issues arising from strong multicollinearity among the traits. The first three PCs accounted for 85.8/89.1% of the generation 12 among-population variation in male/female CHCs respectively (Tables S2.1, S2.2). Males and females from all three sampling generations (i.e., 0, 6, and 12) were separately scored using the eigenvectors (i.e., unit-scaled loadings) of these PCs to calculate their values of these three composite traits.

I tested for the effect of treatment, *i.e.* whether the control, *hybrids killed*, and *hybrids live* differ, on each of the three PCs of among-population CHC variation in males and in females via separate general linear mixed effects models of the individual-level data. The independent variables mirrored those from the analysis of the mating data, except the random effect of population had unique levels for each combination of generation and population (to account for non-independence among individuals both within and across generations). The models were fit using *lme4* (Bates et al. 2015), with significance of fixed effects determined via partial (i.e., type 3) *F*-tests employing Satterthwaite-approximated degrees of freedom via the *lmerTest* package (Kuznetsova et al. 2017). When a significant treatment × sampling generation interaction was detected, a simplified model was subsequently fit, separately by sampling generation, that included fixed effects of treatment and experimental block (2 levels), and a random

effect of population. Significant treatment effects, when present, were further characterized via post hoc pairwise comparisons of marginal means among treatment levels, with p-values adjusted for multiple comparisons via the Tukey procedure implemented in the *emmeans* package (Lenth 2021).

Finally, I wanted to compare the evolutionary responses to our treatments with existing patterns of CHC divergence in nature, both between species (*D. subquinaria* and *D. recens*) and geographic contexts (i.e., *D. subquinaria* populations that are sympatric vs. allopatric with *D. recens*). To do this, I used CHC data from Dyer *et al.* (2014) that characterized multiple populations of both species collected from allopatry and sympatry. I used only the 'inland' allopatric populations of *D. subquinaria* ('coastal' allopatric populations form a third, distinct group that differs from both inland allopatric and sympatric populations). Trait means were calculated for each of the four combinations of species × geographic context: sympatric *D. subquinaria* (four populations, 141 total individuals), allopatric *D. subquinaria* (three populations, 157 individuals), sympatric *D. recens* (four populations, 198 individuals), and allopatric *D. recens* (three populations, 162 individuals). A PCA, separate from the earlier one above, was then performed on these data to quantify the primary axes of CHC divergence among these natural populations. This was done separately by sex and used the same CHCs as the current study; for natural populations, an average of 26.08 ± 3.6 SD males and 24.54 ± 2.6 SD females were used in the PCA. The first two PCs accounted for ~99.0% of the variation among these four groups in both males and females (Tables S2.3, S2.4). Males and females from generation 12 of the experimental populations were then projected into this trait space of naturally occurring among-population CHC

variation. This was done by scoring experimental individuals using the two PC eigenvectors to calculate their values for these two composite traits. Treatment effects on these two traits in each sex were tested via separate linear mixed effect models that included fixed effects of treatment and experimental block (2 levels), and a random effect of population, mirroring within-generation CHC analyses above.

2.3 Results

Evolution of sexual isolation

Initially there was no difference between treatments in the sexual isolation (i.e., heterospecific mating rate) of *D. subquinaria* females with heterospecific *D. recens* males, but differences in the strength of sexual isolation arose across generations in both experimental sympatry treatments relative to the allopatric control (Figs. 2.1A). This generated a significant treatment × sampling generation interaction for heterospecific mating rate (Wald χ^2_2 = 11.03, p = 0.004; Table S2.5). Analyzing separately by assay generation, treatment effects were absent at the start of the experiment prior to any opportunity for an evolutionary response (likelihood ratio χ^2_2 = 1.47, p = 0.480). Treatment effect was significant by generation 6 (likelihood ratio χ^2_2 = 19.10, p < 0.0001), with the odds of a heterospecific mating being significantly lower relative to the allopatric control in both experimental sympatry treatments (*hybrids live*: odds ratio (OR) = 0.710, p = 0.0251; *hybrids killed*: OR = 0.564, p < 0.0001). The two experimental sympatry treatments did not differ significantly from each other (Table S6). The pattern was similar in generation 12, with treatment effect again significant (likelihood ratio χ^2_2 = 30.92, p < 0.0001). The odds of a heterospecific mating were

significantly lower relative to the allopatric control in both experimental sympatry treatments (*hybrids live*: OR = 0.60, $p = 0.0004$; *hybrids killed*: OR = 0.48, $p < 0.0001$), and both point estimates were lower than in generation 6 (Fig. 2.1A). Again, the two experimental sympatry treatments did not differ, although the point estimate suggested greater discrimination in the *hybrids killed* compared to the *hybrids live* treatments (Table S2.6).

I also present the results as absolute mating probabilities (Fig. 2.1B). These should be interpreted with caution given demonstrated among-assay variation in heterospecific mating rates in the allopatric controls (i.e., Fig. S2.4) which are not accounted for in this approach. These results echo those above in that treatment effects are initially absent in generation 0, and then arise as a pattern of a lower average proportion of heterospecific matings in the experimental relative to the control treatments in subsequent generations (Fig. 2.1B). However, these results also suggest a decrease in the absolute level of sexual isolation across all treatments from generation 0 to 6 (i.e., the average proportion of heterospecific mating rises in all treatments, albeit less in the experimental sympatry treatments compare to the controls).

With respect to mating rates of experimental females with conspecific, ancestral (i.e., allopatric) *D. subquinaria* males, the treatment × generation interaction was non-significant (Table S2.5), although there was a significant main effect of treatment (Fig. 2.1C; $\chi^2_2 = 17.59$, $p = 0.0002$). While the odds of females mating with ancestral conspecific males in the hybrids live treatment did not differ from that in the allopatric control, females from the *hybrids killed* treatment had a significantly reduced odds of mating with ancestral *D. subquinaria* males compared to both the *hybrids live* (OR =

0.627, $p < 0.0001$) and control treatments (OR = 0.516, $p < 0.0001$; Table S2.7). Patterns are qualitatively unchanged when assessing absolute conspecific mating rates instead (Fig. 2.1D).

Evolution of CHCs

In males, differences among populations in the primary axis of among-population CHC variation (i.e., PC1) were initially absent, but differences evolved across generations between the *hybrids killed* treatment and both the *hybrids live* and allopatric control treatments (Fig. 2.2A), generating a significant treatment × sampling generation interaction ($F_{2,36.8} = 5.00$, $p = 0.012$; Table S2.8). Analyzing separately by generation, at the start of the experiment the three treatments did not differ significantly ($F_{2,9.4} = 0.05$, $p = 0.954$; Table S2.9). By generation 6, the *hybrids killed* treatment had diverged from the *hybrids live* and allopatric controls, although not yet significantly ($F_{2,11.8} = 1.08$, $p = 0.370$). By generation 12, however, there was a significant treatment effect overall ($F_{2,12.0} = 11.07$, $p = 0.002$) with the *hybrids killed* differing from both the other two, and the latter two did not differ significantly from each other (Fig. 2.2A; Table S2.9). For PC2 in males, the treatment × generation interaction was non-significant (Table S2.8), but there was a significant main effect of treatment ($F_{2,45.9} = 4.39$, $p = 0.018$) with the two experimental sympatry treatments differing from one another but neither differing significantly from the control (Fig. 2.2B; Table S2.9). Neither treatment nor its interaction with generation were significant for PC3 in males (Table S2.8).

In females, neither treatment nor its interaction with generation were significant for PC1 of among-population CHC variation (Table S2.8), although the pattern of trait

means was similar, but somewhat weaker, to that of PC1 in males (Fig. 2.2C). For PC2, the treatment × generation interaction was non-significant (Table S2.8), but there was a significant main effect of treatment ($F_{2, 39.0} = 15.30$, $p < 0.0001$) with the *hybrids killed* differing from both the *hybrids live* and control, which did not differ from one another (Fig. 2.2D; Table S2.9). Neither treatment nor its interaction with generation were significant for PC3 in females (Table S2.8).

To place the evolutionary responses of the experimental populations within the context of natural patterns of CHC variation, I performed sex-specific PCAs of previously published CHC data from multiple natural sympatric and allopatric populations of *D. subquinaria* and *D. recens* (Dyer et al. 2013). This identified two trait combinations in each sex that accounted for ~99% of the variation in CHC means among the four combinations of species and geographic context (i.e., sympatric *D. subquinaria*, allopatric *D. subquinaria*, sympatric *D. recens*, allopatric *D. recens*; Tables S2.3, S2.4). In line with previous analyses of these data, values of these two composite CHC traits distinguished sympatric from allopatric *D. subquinaria* in both sexes, and both of these from *D. recens*, while sympatric and allopatric *D. recens* differed little (Figs. 2.3 and S2.5, and see Dyer et al. 2014). Projecting generation 12 experimental flies into this trait space for these two trait combinations revealed no significant treatment effect on PC1 in either sex (males: $F_{2, 12.0} = 0.39$, $p = 0.69$; females: $F_{2, 10.5} = 1.99$, $p = 0.19$). However, there was a significant evolutionary response to the treatments in PC2 in males ($F_{2, 11.9} = 13.14$, $p < 0.001$; Fig. 2.3), and a similar, but marginally non-significant, effect on PC2 in females ($F_{2, 10.6} = 3.08$, $p = 0.088$; Fig. S2.5). In both sexes, it was the *hybrids killed* treatment that tended to differ from both the allopatric control

and *hybrids live* treatments, with the latter two showing little difference. The response to the *hybrids killed* treatment involved a decrease in PC2 and no change in PC1. This does not align closely with the divergence of sympatric from allopatric *D. subquinaria* populations in nature, which involved an increase in PCs 1 and 2.

2.4 Discussion

Patterns of enhanced prezygotic isolation in sympatry are well established in nature, but there are few evolutionary tests of the underlying causes (e.g., see Rundle and Schluter 1998; Hopkins and Rausher 2012; Yukilevich 2012; Nosil 2013). Manipulating a putative selective agent and then tracking the evolution of reproductive isolation in replicate populations can provide one of the strongest tests possible of the mechanism responsible for a speciation event (Schluter 2000). Here I used such an evolution experiment in which I exposed *D. subquinaria* populations to *D. recens* during mating. I found evolved differences in which sexual isolation was stronger in the experimental sympatry compared to allopatric control populations, replicating a pattern of reproductive character displacement in mate discrimination in nature (Fig. 2.1A). Specifically, after 12 generations the probability of heterospecific matings in the experimental *hybrids killed* and *hybrids live* treatments were, respectively, 0.48 and 0.60 times the probability of heterospecific matings in the allopatric control (Table S2.6; using absolute mating rates: 0.69 and 0.79 times, respectively, that of the control).

Our analyses used the allopatric populations as internal controls for each assay in a given block and generation, represented as the odds ratio of experimental to control mating rates. This is appropriate considering the paired experimental design (i.e., mating

rates were estimated for a set of allopatric controls that were tested at the same time as a set of experimental populations) and given the presence of substantial among-assay variation in mating rates caused by unknown, and hence uncontrolled, factors (Fig. S2.4). A similar interpretation arises from considering absolute mating rates in that treatment differences are initially absent but build across generations, with lower probabilities of mating (i.e., stronger sexual isolation) in experimental treatments relative to the controls. However, changes in absolute mating rates also suggest an initial decrease in sexual isolation across all populations (Fig. 2.1B), with isolation weakening less in the experimental relative to control treatments. That this reflects uncontrolled environmental variation, and not a general weakening of sexual isolation in lab populations, is suggested by additional, longer-term data on heterospecific mating rates in some of the control populations (Fig. S2.4). These data demonstrate that, after 22 generations under the same conditions, heterospecific mating rates were the same as they were in generation zero.

Based on the experimental design, I can rule out several potential explanations for the observed treatment effects on sexual isolation. A differential extinction/fusion process, in which only strongly isolated taxa persist in sympatry, cannot explain it, as no populations were lost during the experiment. Ecological character displacement, which can produce enhanced isolation under sympatry as a by-product, is also unlikely here. Our experimental *D. subquinaria* populations were reared for the majority of their life cycle in the absence of *D. recens*, with experimental sympatry only occurring for the short mating phase each generation. The ecology of *Drosophila* lab bottles is also extremely different, and highly simplified, compared to that of natural populations,

further suggesting that the ecological character displacement is unlikely (an argument also made by Matute, 2010a). On balance, the evolution of reproductive character displacement in this experiment was most likely a response to selection generated by costly heterospecific reproductive interactions.

I assayed sexual isolation against *D. recens* via the presence/absence of larvae in vials where virgin experimental females were held for 24 h with *Wolbachia*-free *D. recens* males. In theory, treatment effects on the presence/absence of larvae could have arisen not from changes in sexual isolation, as I suggest, but from changes in the survival of 'hybrid' larvae. Differences in larval survival among treatments seem unlikely, however, because larval rearing conditions were the same in all of them, meaning selection during this life stage should not have differed. If gene flow from *D. recens* to *D. subquinaria* occurred in the *hybrids live* treatment, this could increase the survival of subsequent hybrid larvae by reducing genetic incompatibilities between these populations. However, this would result in an apparent decrease in sexual isolation in this treatment (i.e., more larvae surviving) relative to the allopatric controls, while I observed the opposite. Census *D. subquinaria* population sizes were also twice as large in the allopatric controls compared to experimental sympatry treatments, generating a potential confound of N_e with experimental treatments. This was done because I felt it more important to control for density during mating interactions, and it is hard to conceive how this could explain the consistent divergence of experimental sympatry populations from the controls.

In studies of reproductive character displacement, costs of heterospecific reproductive interactions are typically assumed to be due to the low fitness of resulting

hybrid offspring, driving a process of reinforcement. However, heterospecific interactions can also be costly to the interacting individuals themselves, i.e., prior to the formation of a hybrid zygote (Gröning and Hochkirch 2008; Hoskin and Higgie 2010). For example, physical harm may occur to the individuals involved, it may increase their risk of predation, and it may expose them to different parasites or diseases. Heterospecific courtship signals may also interfere with conspecific signals, impeding mate acquisition, and if heterospecific individuals are less likely to respond, courtship and attempted matings may also constitute a waste of time, energy, and nutrients, all irrespective of the fitness of any resulting hybrid offspring. These latter processes have been termed reproductive interference (Gröning and Hochkirch 2008).

Finally, it is also possible that hybrid matings do occur but post-mating prezygotic barriers, such as not using stored sperm or conspecific sperm precedence, may reduce or eliminate the production of hybrid offspring (Marshall et al. 2002). The heterospecific mating assays scored the presence/absence of larvae, which can be affected both by mating rate and sperm usage. However, when the *Wolbachia* infection is absent (as in these assays), hybrid matings produce many offspring (Shoemaker et al. 1999) and it thus seems unlikely that this would decline to zero (i.e. absence of any offspring) in only 12 generations of this experiment. In addition, I know there is substantial segregating variation for heterospecific discrimination in the *D. subquinaria* population I used. On balance, while changes in postmating prezygotic traits may contribute to these results, I suspect the observed patterns are more likely due to changes in premating traits. Nevertheless, more work on the potential role of postmating prezygotic barriers would be valuable in this system.

It can be difficult empirically to separate costs of heterospecific interactions themselves versus those arising via the reduced fitness of any resulting hybrid offspring (e.g., Cooley et al. 2006). The distinction between parent and offspring fitness is not always straightforward, and these costs may often co-occur. Insight can nevertheless be gained in some situations. The response to experimental sympatry in Higgie *et al.* (2000), for example, appears to have been driven by selection for efficient mate recognition by *D. serrata* in the presence of *D. birchii*, as hybrids were rarely produced during their experiment, meaning little opportunity for reinforcing selection. In our current experiment, up to ~5% of the adults produced each generation in the *hybrids killed* treatment were F1 hybrids (~1.5% on average; Fig. S2.6). Given that *Wolbachia*-induced cytoplastic incompatibility via an endemic infection in *D. recens* causes ~90% mortality of hybrid embryos, this suggests a substantially higher rate of hybrid zygote production (i.e., on average 15% of eggs), and hence the opportunity for reinforcing selection.

Hybrids were removed in the *hybrids killed* treatment but were allowed to survive and potentially reproduce in the *hybrids live* treatment. This latter treatment was designed to mimic the current situation in the wild, where genetic evidence of gene flow confirms that hybrids are formed and, at least occasionally, reproduce. Not killing hybrids would be expected to weaken reinforcing selection and possibly permit gene flow, so it is perhaps surprising that the response to experimental sympatry was not weaker in this compared to the *hybrids killed* treatment. This could be because selection arose largely from heterospecific costs prior to the formation of hybrid zygotes (i.e., reproductive interference), and hence was similar in both treatments. Alternatively,

reinforcing selection via low hybrid fitness may have contributed, yet have been of a similar strength in both treatments. F1 hybrids produced in the *hybrids killed* treatment had a fitness of zero because they were removed before the next generation. In the *hybrids live* treatment, hybrids were likely produced, at least initially, at a similar rate to that in the *hybrids killed* treatment. While these were not killed as part of the experimental protocol, their fitness is low due to both mortality from cytoplasmic incompatibility and F1 male sterility. The CHC data here provided no indication of substantial introgression into experimental *D. subquinaria* populations, as their CHC profiles (a set of quantitative traits) did not become more similar to *D. recens* (Fig. 2.3 & S2.5). This suggests that the reproductive success of any surviving F1 females was low, and hence their fitness may have been zero, or close to it, in this treatment as well. Additional work will be needed to determine whether selection operated prior to and/or after the formation of hybrid zygotes, and simply genotyping to assay for hybrids would not necessarily have been able to disentangle these possibilities.

While enhanced prezygotic isolation evolved in response to experimental sympatry, I found no evidence that the effects of this cascaded to generate sexual isolation between the experimental *D. subquinaria* females and conspecific ancestral males from allopatry (Fig. 2.1B). This is, perhaps, not surprising as such an effect is a by-product of selection in sympatry and is thus likely to be weaker. In nature, sexual isolation between allopatric and sympatric populations of *D. subquinaria* is much weaker than isolation between *D. subquinaria* and *D. recens* in sympatry (Bewick and Dyer 2014; Dyer et al. 2018). However, other traits not captured in our mating assays that increase sexual isolation, for instance postmating prezygotic barriers, may have evolved.

I did find that females from the *hybrids killed* treatment mated with ancestral *D. subquinaria* males less on average than did females from either of the other two treatments, but this effect was present in all generations, including generation zero prior to any opportunity for an evolutionary response to our treatments (Fig. 2.1C, 2.1D). This may be an artifact of our experimental design that used a *dark-eye* mutation in the *hybrids killed* treatment to identify and remove hybrids. To control for eye color, our mating trials used males of matching color; wild-type males were therefore used when testing females from the control and *hybrids live* treatments, while *dark-eye* males were used in the *hybrids killed* treatment. An overall lower mating success of *dark*-eye males compared to wild-type males could explain this result.

CHC profiles for male *D. subquinaria* also responded to experimental sympatry, but only in the *hybrids killed* treatment. Males from this treatment diverged over the course of the experiment from the allopatric control in their scores for the main axis of among-population variation in CHC profiles (PC1), suggesting that CHCs contributed to the enhanced sexual isolation that evolved. A similar, although weaker and non-significant pattern was observed for PC1 in females, potentially representing a correlated response to selection on male CHCs. (Direct statistical tests comparing treatment responses between the sexes are not possible because the suite of CHCs differ somewhat between them.) I have not performed functional confirmation that CHCs were the target of selection in the *hybrids killed* treatment, though previous studies, including a manipulative perfuming experiment, provided direct evidence that male CHCs, and female preference for them, differ between sympatry and allopatry in nature and contribute to sexual isolation (Dyer et al. 2013; Rundle and Dyer 2015).

In contrast, males in the *hybrids live* treatment showed no significant change in CHC profile, despite this treatment also evolving enhanced sexual isolation. This implies a different basis to the altered mate discrimination that evolved in these two treatments, suggesting selection was targeting different phenotypes. There are multiple sensory modalities, and hence traits, involved in mate choice in *D. subquinaria* (and *D. recens*), including visual, gustatory, and chemosensory signals (Giglio and Dyer 2013), and changes in these or other isolating phenotypes such as postmating prezygotic traits must underlie the enhanced isolation in the *hybrids live* treatment. In both treatments, PC2 accounted for a much smaller proportion of the among-population CHC variation and showed no evidence of an evolved response to experimental sympatry in either sex. Similar to conspecific mating rates, there was a main effect in PC2 that distinguished the *hybrids killed* treatment from the other two treatments in males and females (i.e., that was present even at generation 0). This could again be a pleiotropic effect of the eye color mutation used only in the *hybrids killed* treatment.

I also projected the final (i.e., generation 12) CHC profiles of our experimental *D. subquinaria* populations into the trait space of existing CHC variation among populations of both species in nature (Fig. 2.3). This revealed that the *hybrids killed* experimental sympatry treatment evolved CHCs that were less like *D. recens* than were those of the allopatric control, as expected if these changes contributed to the enhanced sexual isolation that evolved in this treatment. This supports earlier comparative evidence that the reproductive character displacement of CHCs in *D. subquinaria* in nature underlies, at least in part, the reproductive character displacement of sexual isolation (Rundle and Dyer 2015). Interestingly, the changes in CHCs that occurred in response to experimental

sympatry were nearly orthogonal to the difference in CHCs between sympatric and allopatric *D. subquinaria* populations in nature. In particular, the response to the *hybrids killed* treatment involved a decrease in PC2 and no change in PC1 of among-population CHC variation, while in nature sympatric *D. subquinaria* populations have a higher value of PC1 (and PC2) compared to allopatric populations.

That the evolved response to experimental sympatry in the lab does not mirror the evolutionary response to sympatry in nature is perhaps not surprising. First, we cannot recreate the exact genetic variation that was present when sympatry initially occurred in nature. Second, the response we observed in 12 generations may differ from a much longer-term response to selection. Third, the laboratory environment differs from nature and CHCs are not only a means of chemical communication (Chenoweth and Blows 2005; Curtis et al. 2013), but they also serve to waterproof the cuticle, playing a key role in desiccation resistance (Gibbs et al. 1997; Kwan and Rundle 2009; Stinziano et al. 2015). Evolutionary constraints arising from desiccation selection may be much reduced in the lab where humidity is constantly high, allowing CHC expression to be optimized for mate discrimination. Along similar lines, Matute (2010a) found that both sexual and postmating-prezygotic isolation evolved in response to experimental sympatry in the lab, whereas in nature divergence is only found in the latter. As with Matute's experiment, our results highlight differences in the response to selection between lab and field populations. Understanding how ecological constraints vary among populations, and how these impact evolutionary divergence in both ecological and reproductive traits, is an important topic for future work.

2.5 Figures

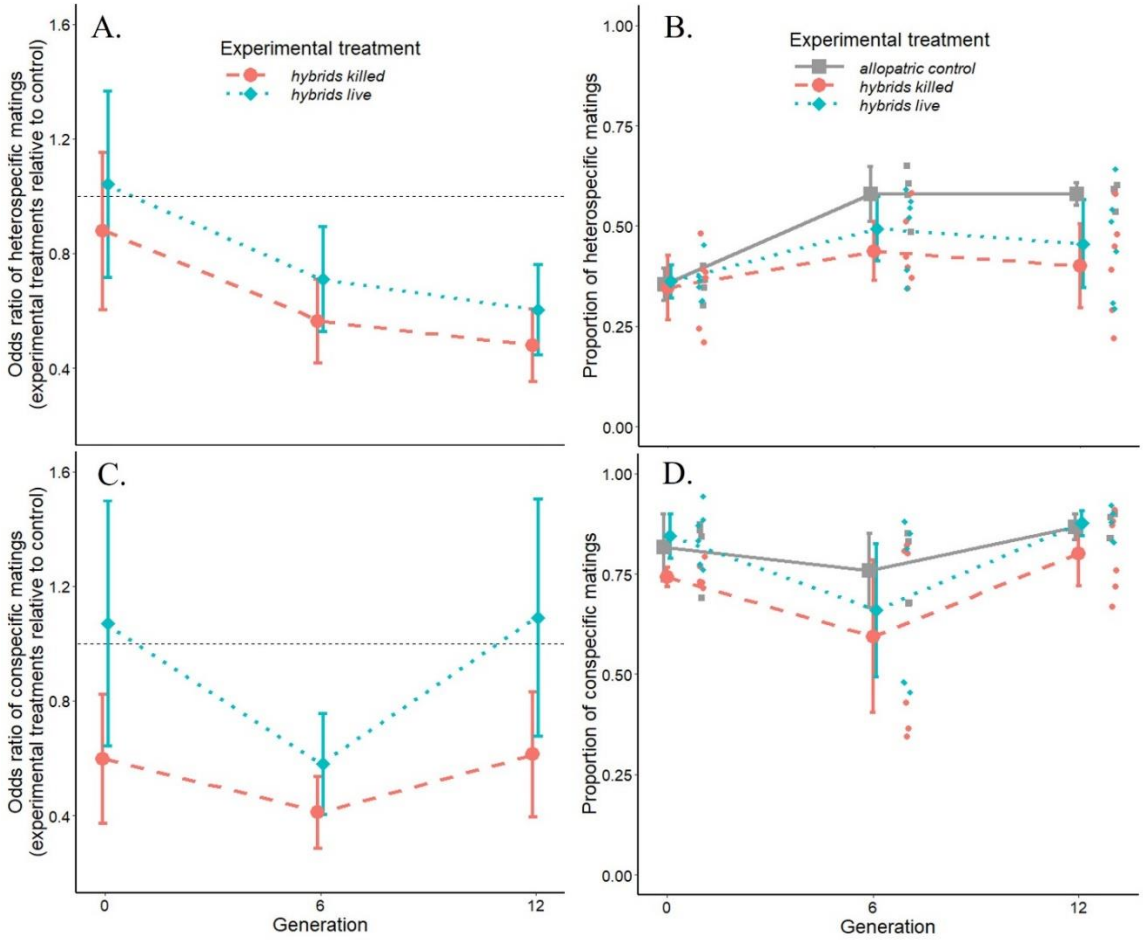

Figure 2.1 Odds ratios of marginal mean mating rates (±95% confidence intervals) of experimental *D. subquinaria* females with **(A)** heterospecific *D. recens* and **(C)** conspecific ancestral *D. subquinaria* males for both experimental sympatry treatments relative to the allopatric control. The horizontal dotted line at 1.0 indicates no difference from the allopatric control. Mean absolute mating rates (±95% confidence intervals) of experimental *D. subquinaria* females with **(B)** heterospecific *D. recens* and **(D)** conspecific ancestral *D. subquinaria* males. Individual populations for each treatment are plotted at each generation as smaller points, offset to the right each generation for clarity.

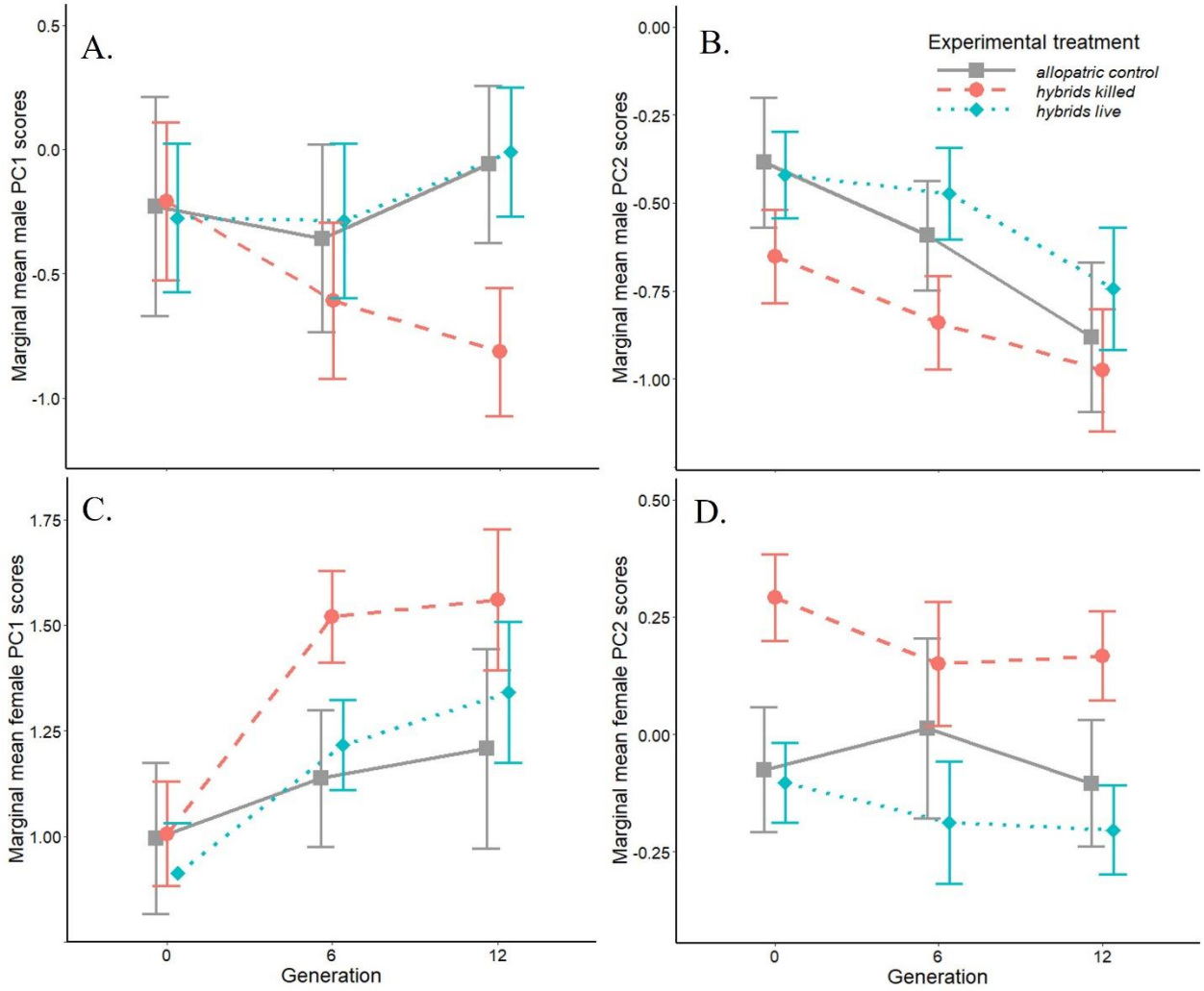

Figure 2.2 Changes in among-population CHC profiles for each experimental treatment over the course of 12 generations, after accounting for fixed differences between blocks. Marginal means (±95% confidence intervals) are for **(A)** male PC1 scores, **(B)** male PC2 scores, **(C)** female PC1 scores, **(D)** and female PC2 scores.

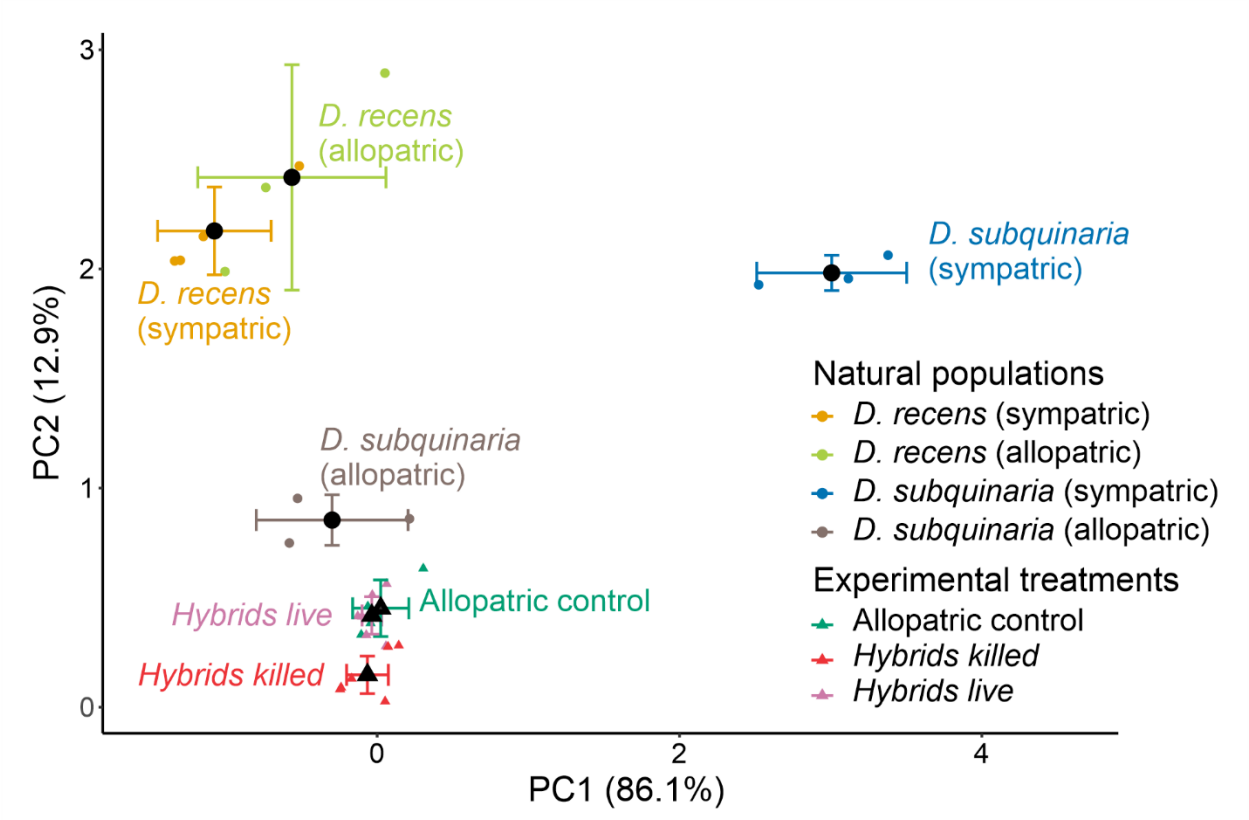

Figure 2.3 Evolutionary response of males from the experimental populations within the trait space that distinguishes sympatric and allopatric populations of *D. recens* and *D. subquinaria* in nature. Principal component analyses of CHC profiles were performed separately by sex on data from Dyer et al. (2014) and then the CHC profiles of experimental populations were projected into this space (see Methods). Colored points are means for natural (circles) and experimental (triangles) populations; black points are group means (± 95% confidence intervals), treating populations as replicates. See Fig. S2.5 for females.

Chapter 3: Divergence in genetic (co)variances and the alignment of g_{max} with phenotypic divergence

The following thesis chapter is a modified version of a manuscript that is currently in preparation for submission to *Evolution*. The manuscript was modified to meet thesis formatting requirements. This work was done in collaboration with Vincent Careau and my supervisor, Howard Rundle. I took the lead role in all aspects of the work including study design, data collection and processing, analyses, interpretation, and writing.

3.1 Introduction

Population divergence is a key component of biological diversity and is intimately linked to the origin of new species (Schluter 2001, 2009; Rundle and Nosil 2005). Divergence is generally multivariate, involving suites of traits that evolve in response to differences in local selection (Lande and Arnold 1983). It has long been appreciated that the evolution of multivariate phenotypes is impacted by the genetic architecture of the traits in question; the patterns of genetic variation and covariation among a suite of traits can either constrain or promote divergence (Lande 1979; Arnold et al. 2001; Walsh and Blows 2009; Blows and McGuigan 2015). However, we also know that genetic architecture is not static: genetic variances and covariance will themselves evolve as divergence proceeds (Agrawal et al. 2001; Eroukhmanoff 2009), potentially altering evolutionary trajectories and complicating our ability to make inferences concerning evolutionary processes. Understanding how multivariate genetic architecture impacts divergence, and how often that architecture changes during divergence, are fundamental questions in evolutionary genetics for which there remains limited empirical data from natural systems.

Multivariate genetic architecture can be characterized with **G**, the matrix of additive genetic variances and covariances for a set of traits. **G** describes patterns of standing genetic variation and covariation within a population, with the latter arising from pleiotropy and/or linkage. **G** shapes evolutionary divergence, at least over the short term, by concentrating genetic variation in certain trait combinations (Lande 1979; Hansen and Houle 2008; Kirkpatrick 2009; Chenoweth et al. 2010). $\mathbf{g}_{max}$, the leading eigenvector of **G**, is the multivariate trait combination with the greatest genetic variance. Depending on

whether selection aligns with trait combinations with high or low genetic variance, adaptive divergence can be enhanced or hampered. Much attention has been paid to whether the primary axis of phenotypic divergence tends to align with $\mathbf{g}_{max}$, as there is no expectation that this should be the case for a novel vector of selection during the early stages of adaptative divergence. If it does not, evolutionary responses will be biased away from the direction of selection and toward the 'genetic line of least resistance' (Schluter 1996). Case studies from natural populations provide mixed results. For example, Chenoweth et al. (2010) show that, for sexually selected male traits in *D. serrata*, divergence between natural populations is strongly biased by genetic (co)variances away from the direction of sexual selection. Similarly, genetic correlations among life-history traits constrain evolutionary trajectories across *Brassica rapa* populations (Mitchell-Olds 1996) and across multiple species of *Gryllus* crickets (Bégin and Roff 2004), suggesting that genetic constraints can influence long-term evolution. In contrast, McGuigan et al. (2005) found that, although neutral divergence due to genetic drift was constrained by genetic (co)variances, adaptive divergence was not. It is thus unclear how often, and to what extent, genetic (co)variances facilitate or constrain population divergence.

While **G** may structure divergence, there is also good reason to expect that **G** itself will change during, and even as a result of, divergence. Genetic variances and covariances are the product of the frequencies and effect sizes of segregating alleles affecting the traits (Barton and Turelli 1987; Reeve 2000; Blows et al. 2004). Contrary to assumptions under an infinitesimal model, distributions may not be Gaussian, and allele frequencies may change appreciably as populations diverge, especially under strong selection (Jones et al. 2003, 2004, 2007; Guillaume and Whitlock 2007; Revell 2007).

Selection-induced changes in allele frequencies can increase or decrease genetic variance (Turelli 1988; Shaw et al. 1995; Blows and Higgie 2003; Conner et al. 2011; Uesugi et al. 2017). For example, if rare alleles are brought under positive selection in a novel environment, genetic variance in the direction of selection may increase, potentially accelerating adaptive diversification (Agrawal et al. 2001). Multivariate selection can also alter trait covariances through shifts in linkage disequilibrium (Samuk et al. 2017), as can genetic drift (Doroszuk et al. 2008), bottlenecks (Roff 2000), and inbreeding (Phillips et al. 2001). Quantitative genetic modeling of changes in **G** is difficult because information concerning the distribution of underlying alleles is lacking. However, simulation studies have explored the evolution of **G** under various conditions (Jones et al. 2003, 2004; Arnold et al. 2008), and experimental manipulations have shown changes in response to both drift (Phillips et al. 2001) and selection (Doroszuk et al. 2008; Hine et al. 2011; Careau et al. 2015). Empirical studies of the stability of **G** among diverging populations in nature provide mixed results. Several have inferred little change in its shape or orientation (Puentes et al. 2016; Delahaie et al. 2017; McGlothlin et al. 2018; Hangartner et al. 2020; Henry and Stinchcombe 2023a), while others find differences in **G** across diverging populations (Hine et al. 2009; Walter et al. 2018; McGlothlin et al. 2022). When assessing difference in **G** among nine geographic populations of *Drosophila serrata* for a set of sexually selected traits, Hine et al. (2009) found that divergence in **G** was driven primarily by one trait combination which reflected changes in the amount of genetic variance among populations (i.e., a change in the size of **G**). In contrast, when comparing **G** among populations of four ecotypes of *Senecio lautus*, Walter et al. (2018) found divergence in multiple independent trait combinations that reflected changes in

both the amount of genetic variance and the covariance structure between ecotypes (i.e. changes in the size and orientation of **G**). Further empirical studies are needed to quantify the frequency and magnitude of divergence in **G** in natural systems.

D. subquinaria is a species that provides an ideal opportunity to investigate questions concerning **G** matrix stability and its role in mediating adaptive diversification. This species occurs in western N. America and its range partially overlaps that of the closely related eastern species, *Drosophila recens*. Evidence suggest that this region of sympatry, just east of the Rocky Mountains, is a secondary contact zone initiated within the last 12,000 years since the end of the Wisconsin glaciation (Jaenike et al. 2006; Bewick and Dyer 2014). Genetic incompatibilities between *D. subquinaria* and *D. recens* cause F1 hybrid male sterility in both directions of crosses. In addition, cytoplasmic incompatibility from a *Wolbachia* infection in *D. recens* causes a further ~90% offspring mortality when *D. subquinaria* females mate with *D. recens* males (Jaenike et al. 2006), generating strong, but not complete, postzygotic isolation. Consistent with the reinforcement of prezygotic isolation in response to low hybrid fitness, *D. subquinaria* exhibit a pattern of reproductive character displacement in which sexual (i.e., behavioral) isolation with *D. recens* is much stronger in sympatry than it is in allopatry (Jaenike et al. 2006; Bewick and Dyer 2014; Dyer et al. 2018). A manipulative evolution experiment further supports the hypothesis that this divergence was an evolutionary response to selection in sympatry arising from reproductive interactions with *D. recens* (see chapter 2).

The divergence in sexual isolation between *D. subquinaria* populations from sympatric vs. allopatric regions is mirrored by a similar pattern of character displacement in a suite of (epi)cuticular hydrocarbons or CHCs. In a survey of multiple geographic populations, different multivariate CHC profiles distinguished flies from sympatry vs. allopatry (Dyer et al. 2013), consistent with divergent selection on these or correlated traits. CHCs are integral to controlling the permeability of the cuticle to water loss and so play a key role in desiccation tolerance in *Drosophila* and other insects (Howard and Blomquist 2005; Kwan and Rundle 2009). However, CHCs also serve as contact pheromones that function as a chemosensory means of communication (Ferveur 2005), often leading to trade-offs between viability and sexually selected trait optimums (Berson et al. 2019). I expect genetic correlations among this set of traits as they share much of the same developmental pathways (Blomquist and Bagnères 2010). In *D. subquinaria*, manipulative perfuming experiments and multiple behavioral assays have demonstrated that female mate preferences and species discrimination are based in part on variation in CHCs profiles (Curtis et al. 2013; Dyer et al. 2013; Giglio and Dyer 2013), and female preferences for these traits have likewise diverged between sympatry and allopatry (Rundle and Dyer 2015; Dyer et al. 2018). A laboratory evolution experimental also reveals that CHCs can evolve in response to experimental sympatry with *D. recens* (see chapter 2), consistent with adaptive divergence of these traits in nature.

Here I take advantage of this divergence to study the quantitative genetic basis of variation in CHCs and how this may have impacted, and evolved in response to, the character displacement that has occurred in this suite of traits. I start by characterizing multivariate divergence in CHCs among six replicate populations of *D. subquinaria*,

recently collected from the sympatric and allopatric regions (three populations from each). From this I estimate **D**, the (co)variance matrix of population-differences in mean phenotypes, and $\mathbf{d}_{max}$ (the leading eigenvector of **D**), the trait combination for which there is the greatest divergence among population multivariate means. Using a replicate breeding design within each population, I also estimate the additive genetic (co)variance structure of these traits (i.e., **G**). Then, using these data, I first investigate whether there is evidence of divergence in **G** between allopatric and sympatric regions. I then employ several analytical methods to identify possible aspects of changes in **G**, first characterizing divergence holistically via eigenanalysis of the difference matrix between region-specific **G** matrices (i.e., $\mathbf{G}_{sympatry} - \mathbf{G}_{allopatry}$), followed by separate investigations of how variances and covariances each may have diverged. Next, I explore the extent of any association between **G** and **D**, specifically the alignment between $\mathbf{g}_{max}$ within each region and $\mathbf{d}_{max}$. Lastly, to shed light on whether character displacement has structured the evolution of genetic (co)variance structure in this system, I test whether the major axes of divergence in **G** between regions align with the major axis of among-population trait divergence ($\mathbf{d}_{max}$).

3.2 Methods

Populations and Breeding Design

Between 2019 and 2021, adult *D. subquinaria* females were collected from three geographic populations from the region west of the Rocky Mountains that is allopatric to *D. recens,* and from three geographic populations from the region in central Canada that is sympatric with *D. recens* (Fig. S3.1). Flies were collected by setting baits consisting of

small piles of white button mushrooms (*Agaricus bisporus*) in damp, forested habitat following Dyer et al. (2018), from which individuals were captured using fine-mesh nets and stored in vials with an agar-sugar medium. Females were identified, isolated, and shipped to the University of Ottawa from which isofemale lines were established (see Table 3.1 for population details). *D. subquinaria* and *D. recens* females cannot be reliably distinguished visually, so species identity of each isofemale line was confirmed using molecular markers following Bewick & Dyer (2014). *D. recens* lines were subsequently discarded and *D. subquinaria* isofemale lines were maintained via non-overlapping generations in a common environment (20° C, 75% humidity, 12 hr. L:D cycle), in glass vials with 1.3 g of Instant Drosophila Medium (Formula 4-24; Carolina Biological Supply Company, Burlington NC), 6 mL of water, a piece of folded blotting paper as a pupation substrate, and a piece of fresh commercial white button mushroom.

Outbred populations for breeding designs were created by combining all isofemale lines from a given geographic population via two generations of a crossing scheme in which lines were numbered and then each was separately crossed with each of its two "neighboring" line numbers (e.g. male from line #1 with female from line #2, male from line #2 with female from line #3, etc.), followed by one generation of mass mating. Breeding designs were initiated in the subsequent generation by collecting virgin adults and holding them separately by sex for 7 days until sexually mature. Breeding designs were performed in blocked pairs of one population from the region of sympatry and one from the region of allopatry to avoid confounding block and region. For each population, 150-200 sires were each mated to each of four unique dams in sequence over a period of 4 days by confining a given sire with a single female for 24 hours. Following

their 24-hour mating period, dams were held individually in vials for egg laying for 7 days, after which they were discarded. The setup of sires was further sub-blocked over a period of 2 days in block 1 for logistic purposes. *D. subquinaria* is challenging to maintain in the lab compared to other model *Drosophila* species, and I successfully sampled from an average of 113.2 ± 8.9 SE sires per population (Table 3.1), with an average of 2.88 ± 0.03 SE half-sib families per sire. The breeding design for the Deer Creek population (allopatric region), originally conducted in block 2, was repeated in block 3 by creating a new outbred population from the isofemale lines to increase the number of sires.

Virgin male offspring were collected from each half-sib family using light CO_2 anesthesia and were held in vials with 1.3 g instant medium in groups of 5-10 for 7 days to reach sexual maturity. CHCs were then extracted from three males per half-sib family by separately placing individuals in 100 µl of hexane for 3 min., followed by 1 min. of vortexing before the fly was removed and discarded. Extractions were performed in the mornings within 5 h of incubator lights turning on, alternating between the two populations from the different regions every 100 extractions to minimize temporal confounds given that CHCs can vary in a circadian pattern (Gershman et al. 2014).

CHCs were quantified via gas chromatography with flame ionization detection on an Agilent 6890N dual-channel "fast" (220 V oven) gas chromatograph (Agilent Technologies, Wilmington, DE) following previously established methods (Curtis et al. 2013; Rundle and Dyer 2015). CHCs identifications were made via qualitative alignment of chromatographic peaks with those in Curtis et al. (2013) in which compounds were identified via mass spectrometry (Fig. 3.1; Table S3.1). The area under each of 10 peaks

(i.e., unique CHC compounds) was then integrated to quantify their abundances. These ten compounds are a subset of those identified by Curtis et al. (2013) in male *D. subquinaria* which focuses on the most abundant CHCs that could be most reliably integrated; 11-*cis*-Vacenyl acetate (cVa) and several tri-acylglycerides were excluded because these compounds transfer to females during copulation and hence are not amenable to sexual selection analyses, a component of a larger project that will be addressed in a subsequent publication.

There is substantial technical error associated with quantifying absolute compound abundances via gas chromatography, so relative abundances (i.e., proportional concentrations) of each compound were calculated by dividing the area under a given peak by the total area under all ten peaks for a given individual. To break the unit-sum constraint inherent in relative abundances, proportions were logcontrast transformed by dividing the proportional area of each peak by the proportional area of an arbitrarily selected common divisor (peak 5 in this case) and taking the $\log_{10}$ of this quotient (Aitchinson 1983). The resulting nine logcontrast CHCs are unitless and on the same scale, and since they reflect proportions, further standardization is neither necessary nor appropriate (Hansen and Houle 2008; Hansen et al. 2011).

Characterizing phenotypic divergence

I assessed multivariate phenotypic divergence among all six populations using a nested linear model in R (i.e., a MANOVA), where the nine logcontrast CHC traits were fit as a multivariate response, with experimental block, GC channel, and population as fixed

effects (added in that order). Following Martin et al. (2008), the variance-covariance matrix of divergence in population means (the **D** matrix) was calculated by extracting the sums of squares and cross-product matrix for population ($SSCP_P$), dividing it by its degrees of freedom to find the mean squares matrix ($MS_P = SSCP_P/4$), and finally dividing the mean squares matrix by *nf*, where *nf* is an approximation of the number of individuals sampled from each population in a balanced design. Eigenanalysis of **D** gives the linear combinations of traits describing divergence in multivariate phenotypic means among the populations.

G-matrix estimation

I estimated a single overall **G**-matrix, as well as separate **G**-matrices by regions (i.e., sympatric vs. allopatric) and by population, by implementing separate hierarchical mixed effects models in MCMCglmm (Hadfield 2010). Each **G**-matrix had 9 traits (45 parameters) and the model included fixed effects of population (where applicable) and gas chromatograph channel (two levels), and random effects of sire and dam nested within sire (Lynch and Walsh 1998; Chenoweth et al. 2010). Random effects were all modeled as unstructured covariance matrices. Traits were multiplied by a factor of 10 to help with model convergence, which was factored out afterward by dividing each variance-covariance matrix by 100. **G** was estimated as four times the sire-level covariance matrix. Using a Bayesian framework allowed me to carry forward uncertainty around the estimated **G**-matrices in the form of posterior distributions. For the single overall **G**, and separate **G**-matrices by region, I checked the posterior mean estimates against **G**-matrices estimated via restricted maximum likelihood (using ASREML-R;

Butler et al. 2023) with the same model structure by assessing the correlation between corresponding eigenvectors of **G**. Correlations were greater than 0.99 for the first three leading eigenvectors (accounting for >80.6% variance) in all cases, and the resulting **G**-matrices were qualitatively very similar (unpublished results).

My Bayesian models used weakly informative inverse Wishart priors for estimating variances and covariances, where the distribution was described by the variance (V = diag(9)) and the belief parameter (nu = 8.002), and a normal prior for the fixed effects. I tested a variety of alternative priors, including using the observed phenotypic variance, and found that different prior specifications had minimal effect on the estimated **G**-matrices. The prior I selected had the lowest average DIC score (Deviance Information Criterion) among those tested. My MCMCglmm models ran for 505,000 iterations, with a burn in of 5,000 and a sampling interval of 500 iterations, resulting in 1,000 posterior estimates. 95% of the estimated parameters had an effective sample size greater than 90% of the total number of samples for the overall **G**, and **G** by region models, and autocorrelation was less than 0.05 for all parameters. However, at the population level where sample sizes are reduced, **G**-matrices were estimated with much less confidence, tending to have low effective sample sizes and autocorrelations greater than 0.05 for many parameters. Unlike at the higher levels (see Results), genetic variance at the population level could not be distinguished from null sampling variation for 26 of 54 ‘traits’ across the six populations (i.e., 6 populations × 9 traits/population), so subsequent analyses focus on the significance of the overall **G** and its divergence between regions.

Null models

To generate suitable null distributions for use in significance testing of **G** and its divergence between regions, I ran corresponding models to those above but using breeding design data that were randomized within each population. Each individual's logcontrast CHC values and their associated GC channel identifier were shuffled intact (i.e. as a single unit) across sires and dams. This means that, in each randomized data set, a given sire would have the same number of dams, and a given dam the same number of offspring, as in the original data set, but the identity (i.e. CHC trait values) of the offspring would change. This approach preserved the nested structure of the original data, but ensured that any estimated genetic variance would reflect sampling variation only. This is the same conceptual approach used by Walter et al. (2018) and Henry and Stinchcombe (2023). For each of 1,000 randomized datasets, I ran a short chain MCMCglmm model, with the same structure and priors as the one for observed data. Models were run for 15,000 iterations, a burn in of 5,000, and a sampling interval of 100. I then took the 100^{th} posterior sample from each run to create null posteriors of **G** for each region and for a single overall **G**. Like for the observed data models, 95% of the estimated parameters had an effective sample size greater than 90% of the total number of samples for both models, and autocorrelation was less than 0.05 for all parameters. Morrisey et al. (2019) notes that randomizing individuals over significant effects will inflate null distributions and are therefore overly pessimistic, so I avoid this by keeping fixed effects and individual traits together during the randomization.

For a given **G**, significant additive genetic variances (V_A) in a logcontrast CHC was inferred if the 95% Highest Probability Density (HPD) interval around its mean

posterior estimate did not overlap with the HPD interval for the null estimate for the same trait from the corresponding model fit to the randomized data, indicating greater genetic variance than expect via sampling variation alone. I also assess the number of eigenvectors of each **G** with statistical support by comparing the 95% HPD intervals of their eigenvalues with that of the eigenvalue of eigenvectors from the corresponding null **G** estimated with randomized data.

Testing for divergence in G between regions

The divergence in **G** between sympatry and allopatry was assessed by comparing the fit of a model specifying a single overall **G** to one that estimated **G** separately by region using DIC scores. Differences in DIC > 2 are often considered strong support for improved model fit (Spiegelhalter et al. 2002; Puentes et al. 2016), and I employed a more conservative rule of ΔDIC > 4. I corroborated this with likelihood ratio tests of the corresponding nested models fit via ASREML-R, specifying either a single overall sire-level (co)variance or separate sire-level (co)variances for each region. In each case, dam and residual matrices were fit separately by region with unstructured (co)variances.

Characterizing between-region divergence in G

Given strong support for divergence in **G** between regions (see Results), I then proceeded to characterize differences between sympatric vs. allopatric **G**-matrices via an eigenanalysis of their difference matrix. Subtracting $\mathbf{G}_{\text{sympatry}}$ from $\mathbf{G}_{\text{allopatry}}$ finds a new 9 x 9 matrix $\mathbf{G}_{\text{difference}}$, eigenanalysis of which finds the linear trait combinations for which

genetic variances and covariances differ the most between regions (Sztepanacz and Rundle 2012). This is equivalent to a tensor analysis which identifies the axes in which (co)variances differ between two or more matrices by calculating a fourth-order genetic covariance tensor (Hine et al. 2009; Aguirre et al. 2014; Walter et al. 2018). Given two matrices being compared here, the first eigentensor of the fourth-order genetic covariance tensor, normally termed **E**1, is equivalent to $\mathbf{G}_{\text{difference}}$. The tensor analysis approach has the advantage that it can be applied to more than two matrices but given that I am concerned with only two region-specific G matrices, I focus on the conceptually simpler approach of analyzing the difference matrix. I tested the significance of eigenvectors of $\mathbf{G}_{\text{difference}}$ ($\mathbf{e}_i$) in explaining divergence in the (co)variance structure between **G**-matrices by comparing the 95% HPD interval around each corresponding eigenvalue to the intervals around eigenvalues from $\mathbf{G}_{\text{difference, null}}$, the difference matrix between the region-specific null **G**-matrices estimated from randomized data.

I then investigated how exactly differences in genetic variances and covariances (correlations) contributed to divergence in **G** between regions. To quantify differences in the total amount of genetic variance, I divided the posterior distribution of the trace (i.e., sum of the variances) of $\mathbf{G}_{\text{sympatry}}$ by the posterior distribution of the trace of $\mathbf{G}_{\text{allopatry}}$. The significance of any deviation of this ratio from a value of one, indicating more or less genetic variance in sympatry compared to allopatry (ratios > 1 or < 1, respectively), was determined by comparing the overlap between the 95% HPD interval around this ratio with that of the corresponding null ratio calculated with posterior estimates from randomized **G**-matrices. Differences in trait-specific genetic variances were likewise assessed via overlap (or not) of the 95% HDP intervals for a given trait in the sympatric

compared to $\mathbf{G}_{\text{allopatry}}$. I can also quantify the amount of genetic variance in each region for the linear trait combinations underlying changes in **G** as identified by eigenanalysis of the difference matrix. To do this, I projected each of the first two eigenvectors of $\mathbf{G}_{\text{difference}}$ through the original **G**-matrices for allopatry and sympatry using:

$$V_i = e_i^T \boldsymbol{G} e_i,$$

where V_i is the variance in **G** associated with e_i, i.e. the *i*th eigenvector of $\mathbf{G}_{\text{difference}}$ (Walsh and Blows 2009). I constructed 95% HPD intervals around these variance estimates by carrying forward the posterior distributions for each **G** and for each eigenvector of $\mathbf{G}_{\text{difference}}$.

Next, I tested whether there were significant differences in genetic covariance (correlation) structure in isolation of changes in genetic variance. This was done by comparing the fit of a model that estimated separate variances and correlations by region to one that constrained correlations to be estimated as a single matrix for both regions but allowed variances to be estimated separately. Fitting genetic correlations is appropriate here because variances differ (see Results), and correlations are standardized in this respect. These nested models were fit in ASREML-R, allowing region-specific unstructured correlation matrices for the dam and residual level, and were then compared with a likelihood ratio test. I employed REML because there is not an equivalent method to constrain correlations in MCMCglmm. I also examined differences between allopatric vs. sympatric **G**-matrices in the trait-by-trait correlations by assessing the overlap (or not) of their 95% HPD intervals.

Lastly, sympatric vs. allopatric **G**-matrices were compared via Krzanowski's subspace analysis, which is a commonly employed geometric approach to comparing the dimensionality of two or more matrices (Krzanowski 1979). First, the primary subspace for each matrix is calculated from a set of eigenvectors of **G** that account for the majority of total genetic variance. Following Aguirre *et al.* (2014), I used eigenvectors that accounted for >90% of the genetic variance (the first three eigenvectors in this case, which is also the minimum number of eigenvectors with statistical support for each **G;** see Results). A common subspace, **H**, was then calculated as:

$$\boldsymbol{H} = \sum_{t=1}^{p} A_t \, A_t^T,$$

where A_t is the subset of eigenvectors for each of the p (two in this case) **G**-matrices, and the superscript T indicates the matrix transposition. The eigenvectors of **H** will have eigenvalues between 0 and p (i.e., the number of original matrices; two in this case), and higher values indicate greater similarity of their primary subspaces. I compared the posterior mean and 95% HPD interval of the eigenvalues of **H** to those calculated from the null distributions of **G** to assess whether there was any evidence of more or less similarity than expected by chance. Higher values would indicate that the **G**-matrices for *D. subquinaria* in each region have genetic variance concentrated in the same multivariate trait subspaces, while lower values would indicate differences in the trait subspaces with genetic variance.

Alignment between changes in genetic (co)variance and phenotypic divergence

To assess the alignment between the primary axis of phenotypic divergence among populations ($\mathbf{d}_{max}$, the first eigenvector of **D**) and the axis of greatest genetic variance ($\mathbf{g}_{max}$), I calculated the angle between these vectors for each region (i.e. sympatric and allopatric $\mathbf{g}_{max}$). For each case I used the appropriate posterior distribution of $\mathbf{g}_{max}$ estimates, and calculated the posterior mean correlation and 95% quantiles. To test the significance of this angle, I compared it to the distribution of angles between 1,000 pairs of random vectors with *n* samples (where *n* is the number of traits) drawn from a normal distribution, then normalized to unit length (Marsaglia 1972). An angle close to 0 would indicate strong alignment between phenotypic divergence and the 'genetic line of least resistance', indicating high genetic variance for the trait combination underlying divergence between sympatry and allopatry.

I can also ask whether changes in genetic (co)variance between regions are associated with the major axis of among-population phenotypic divergence. The leading eigenvectors of the difference matrix ($\mathbf{G}_{difference}$) describe orthogonal trait combinations that underly the greatest changes in genetic (co)variances between region-specific **G**-matrices. I assessed how these vectors aligned with phenotypic divergence among population means by calculating each angle between e_i and $\mathbf{d}_{max}$. I can again test the significance of this correlation by comparing it to a null distribution of angles between 1,000 pairs of random unit vectors and assessing the overlap between the 95% quantiles.

3.3 Results

Characterizing phenotypic divergence

There was significant divergence among populations in mean CHC phenotypes (MANOVA: Wilks' $\lambda = 0.248$, $F_{4,1819} = 85.01$, $p < 0.0001$). $\mathbf{d}_{max}$, the first eigenvector of **D**, accounted for the majority (79.5%) of variance among population means, and primarily reflects variation in the relative abundance of longer chain CHCs (Table 3.2). $\mathbf{d}_2$ accounts for most of the remaining variance (18.1%). Patterns of differentiation among populations in $\mathbf{d}_{max}$ and $\mathbf{d}_2$ reveals that $\mathbf{d}_{max}$ corresponds to divergence between sympatric vs. allopatric regions, while $\mathbf{d}_2$describes an unknown aspect of among-population differentiation (Fig. 3.2). $\mathbf{d}_{max}$ here is extremely similar to the first canonical variate calculated from a MANOVA distinguishing CHCs by region (vector correlation = 0.98), confirming that $\mathbf{d}_{max}$, the primary axis of differences among these six populations, essentially reflects divergence between sympatry vs. allopatry.

Estimating G and testing for divergence between regions

I first estimated the genetic (co)variance structure of CHCs via a single, overall **G**-matrix (Fig. 3.3A). I found that all traits have additive genetic variance (V_A) greater than one would expect from sampling variation alone (Fig. 3.3B). The genetic covariances/correlations between pairs of traits are generally positively, with the exception of logcontrast CHC17 which is negatively correlated with most other traits. The first four eigenvectors of the overall **G** explained statistically significant amounts of variation, as inferred by the non-overlapping HPD intervals of their eigenvalues with that of the eigenvalue of the corresponding eigenvectors of the null **G** estimated with randomized data (Fig. S3.2, Table S3.2).

Divergence in **G** was tested by comparing a model that fit a single overall **G**-matrix to one that fit separate **G**-matrices by region. There was strong evidence for divergence, with a lower DIC score for the separate **G**-matrices model than for the single-**G** model ($\triangle$DIC = -75.7), indicating a significantly better fit to the data. I corroborated these results with a likelihood ratio test of the corresponding models in ASREML-R, which indicated a better model fit with separate sire level covariances and thus significant structuring of **G** by region ($\chi^2_{45\ df} = 818.7$, $p < 0.0001$).

Given evidence of divergence in (co)variance structure between regions, I then estimated **G** separately for the sympatric and allopatric populations (Fig. 3.4). With the exceptions of logcontrast CHC7 and CHC12 in sympatry, all traits had significant V_A in both regions (i.e., 95% HPD interval of each point estimate did not overlapped with its that of its corresponding null estimate representing sampling variation), demonstrating additive genetic variance for CHCs in both regions (Fig. 3.5A). Visual examination of the two **G** matrices revealed some noticeable differences (significant for most comparisons, see below for details), including smaller point estimates of V_A in sympatry as compared to allopatry for most (but not all) traits, a general weakening of positive additive genetic correlations in sympatry compared to allopatry, and negative correlations involving logcontrast CHC17 in allopatry becoming positive in sympatry (Fig. 3.4). The first four eigenvectors of $\mathbf{G}_{\text{sympatry}}$ had statistical support (Fig. 3.6A, Table S3.2), while the first three had support for $\mathbf{G}_{\text{allopatry}}$ (Fig. 3.6B, Table S3.4).

Characterizing between-region divergence in G

Given significant divergence in **G** between regions, I used eigenanalysis of the difference matrix ($\mathbf{G}_{difference}$) to investigate these differences quantitatively. Eigenanalysis of $\mathbf{G}_{difference}$ revealed that the first three eigenvectors accounted for more variation than expected by sampling variation alone (Fig. S3.4), and identified a single linear trait combination ($\mathbf{e}_1$) that explained the majority of the differences between the **G**-matrices (71.1%). $\mathbf{e}_1$ primarily groups a set of six logcontrast CHCs (all with substantial negative loadings), and contrasts these with logcontrast CHC17 which has a weaker positive loading (Table 3.3). The second eigenvector of $\mathbf{G}_{difference}$ ($\mathbf{e}_2$) accounts for much of the remaining variation between the **G**-matrices (17.0%), and it primarily differentiates logcontrast CHC17 and CHC19, which have strong positive loadings, from the other traits, which have negative or weakly positive loadings. This analysis is sensitive to differences in both variances and covariances, so I examined these separately to further characterize the divergence of **G** and to aid interpretation of $\mathbf{G}_{difference}$ and its eigenvectors.

Comparing the traces of the sympatric and allopatric **G** matrices revealed that total genetic variance in sympatry was 44% of that in allopatry. This reduction in total V_A in sympatry was significant: the 95% HPD interval around the ratio of the sympatric to allopatric traces did not overlap with one (the expectation under equal variances), nor with the 95% HPD interval around the corresponding null ratio from the sympatric and allopatric **G**-matrices constructed from randomized data (Fig. 3.5B). Examination of the individual traits reveals six for which the point estimate of V_A in sympatry is significantly less than that in allopatry (i.e., the 95% HPD intervals do not overlap), while there is a single trait (logcontrast CHC17) where the reverse is true (Fig. 3.5A). The remaining two

traits have variances that do not differ significantly. These differences in trait-specific genetic variances between sympatry and allopatry generally match the major axis of differentiation in genetic (co)variance structure captured by $\mathbf{e}_1$, which has strong negative loadings for the six traits with significantly greater variance in allopatry, a positive loading for the one trait with significantly greater variance in sympatry (logcontrast CHC17), and smaller contributions for the two traits for which variance did not differ significantly (logcontrast CHC2 and CHC19). $\mathbf{e}_2$ further clusters logcontrast CHC17 and CHC19, the only two traits with point estimates indicating greater variance in sympatry than allopatry. This indicates that changes in levels of genetic variance are a major contributor to the divergence of **G**. Projecting $\mathbf{e}_1$ and $\mathbf{e}_2$ through the sympatric and allopatric **G** matrices quantifies the amount of additive genetic variation in these two trait combinations for which **G** differs most. Consistent with difference in V_A being important to the divergence of **G**, there is significantly more genetic variance in allopatry than in sympatry for $\mathbf{e}_1$, while the reverse is true for $\mathbf{e}_2$ (Fig. 3.7).

Turning to genetic correlations, divergence between the sympatric and allopatric **G** matrices in these was also significant overall: a likelihood ratio test indicated a significantly better fit for a model that allowed genetic correlations to differ by region compared to one that constrained all of the correlations to be the same ($\chi^2_{36\ df} = 707.4$, $p < 0.0001$). Examination of the additive genetic correlations revealed 20 that were significantly higher in allopatry than sympatry (i.e., 95% HPD intervals do not overlap), seven for which the reverse was true (i.e., significantly weaker in allopatry than sympatry), and nine that did not differ significantly (Fig. S3.3). These changes in correlation structure are apparent in the second eigenvector of $\mathbf{G}_{\text{difference}}$ (i.e., $\mathbf{e}_2$; Table

3.3). $\mathbf{e}_2$ has strong contributions of logcontrast CHCs 13, 17 and 19, and it is these traits that are involved in some of the most substantial changes in genetic correlations: correlations involving CHC17 are almost exclusively negative in allopatry but positive in sympatry, generally strong and positive correlations with CHC13 in allopatry weaken in sympatry and become negative in two cases, and correlations involving CHC19 also tend to weaken in sympatry compared to allopatry.

As an additional means to gain insight into divergence in **G**, I used Krzanowski's subspace analysis to assess overlap in the primary subspaces of the sympatric and allopatric **G** matrices. The eigenvalues of the three eigenvectors of **H** were close to the upper bound of two, suggesting that the subspaces were highly similar between the region-specific **G**-matrices (Fig. 3.8). However, the 95% HPD intervals for these estimates overlapped with the corresponding null estimates calculated from the subspace comparison of the sympatric and allopatric **G**-matrices estimated from randomized data. This indicates that there is no more (or less) similarity between the observed allopatric and sympatric **G**-matrices than expected due to sampling variation alone. Eigenvalues decreased somewhat for the higher eigenvectors of **H**, indicating somewhat less similarity, but values still did not differ from those of the null comparisons (Fig. 3.8).

Alignment between changes in genetic (co)variance and phenotypic divergence

I found no evidence of a significant alignment between the primary axis of phenotypic divergence among populations ($\mathbf{d}_{max}$) and the axis of greatest genetic variance ($\mathbf{g}_{max}$) in allopatry: the angle was large (61.9º, 95% CI: 60.8º - 63.5 º) and did not differ from that

expected via sampling variation alone (Fig. 3.9A). In contrast, the angle between $\mathbf{d}_{max}$ and $\mathbf{g}_{max,\ sympatry}$ (i.e. the axis of greatest genetic variation in sympatry) was much smaller (18.9º, 95% CI: 16.9º - 20.8º) and was significantly less than expected via sampling variation (Fig. 3.9A). This indicates an alignment between the major axis of divergence and the 'genetic line of least resistance' within sympatric populations (i.e., high genetic variation in sympatry in the trait combination that has diverged).

Finally, I found mixed evidence for significant associations between phenotypic divergence ($\mathbf{d}_{max}$) and changes in genetic (co)variance between regions (Fig. 3.9B). The angle between the leading eigenvector of the difference matrix ($\mathbf{e}_1$) and $\mathbf{d}_{max}$, the primary axis of phenotypic divergence among populations, was 68.6º (95% CI: 66.9º - 70.3º), which does not differ from the null distribution of angles (mean angle: 72.9º, 95% CI: 58.2º - 87.7º). However, there was evidence of some alignment between $\mathbf{e}_2$ and $\mathbf{d}_{max}$, as the angle there of 31.0º (95% CI: 28.4º - 34.4º) was smaller and did not overlap with the null distribution of angles (mean angle: 73.2º, 95% CI: 58.8º - 87.6º).

3.4 Discussion

The genetic (co)variance structure for a suite of traits can play a major role in structuring evolutionary divergence, including biasing the response to selection (Lande 1979; Arnold et al. 2001; Walsh and Blows 2009). As divergence proceeds, changing allele frequencies can in turn alter the underlying (co)variance structure (Agrawal et al. 2001; Eroukhmanoff 2009). There is a great deal of interest in understanding the interplay between patterns of genetic (co)variation within populations and phenotypic divergence

between them, and empirical data addressing this for populations in nature have only recently started to accumulate. Here I used replicate breeding designs to quantify the additive genetic (co)variance structure (**G**-matrix) of a suite of cuticular contact pheromones (CHCs) in six geographic populations of *D. subquinaria* that are either allopatric or sympatric to the closely related *D. recens*. I found that among-population phenotypic divergence is dominated by differences between these regions (i.e., sympatry vs. allopatry), consistent with previous studies (Dyer et al. 2013; Rundle and Dyer 2015). I also show that **G** has also diverged between the regions, and that the major axis of genetic variation ($\mathbf{g}_{max}$) in sympatry is aligned with the primary axis of phenotypic divergence.

Characterizing among-population phenotypic divergence

The major axis of phenotypic divergence in the relative concentrations of CHCs among these populations is one which separates those from the sympatric region from those from the allopatric region. This divergence is part of a pattern of reproductive character displacement that has evolved following secondary contact within the past ~12,000 years (Shoemaker et al. 1999; Jaenike et al. 2006), where sexual isolation of *D. subquinaria* against *D. recens* is stronger in sympatry than in allopatry. Behavioural assays and manipulative experiments demonstrate that CHCs are used in mate choice in this species, that female preferences for these traits (and hence the sexual selection this generates) differs between regions, and that the divergence of CHCs strengthens behavioural isolation (Dyer et al. 2013; Rundle and Dyer 2015). A lab evolution experiment has also confirmed that CHCs can evolve in response to selection generated by the presence of *D.*

recens in at least some contexts (Jarvis et al. 2024). Taken together, these studies strongly infer the adaptive divergence of CHCs in response to differences in selection between regions.

Divergence in genetic (co)variance structure

Our analyses comparing the (co)variance structure of CHCs uncovered four main results. First, there is compelling evidence that **G** has diverged between sympatric vs. allopatric populations. Second, changes in both genetic variances and covariances contribute to this divergence. Third, the orientation of **G** has shifted, and the major axis of genetic variation in sympatry ($\mathbf{g}_{max,\ sympatry}$) is more closely aligned to the primary axis of phenotypic divergence ($\mathbf{d}_{max}$) than expected by chance, while $\mathbf{g}_{max,\ allopatry}$ is not. Fourth, the trait combination for which **G** differs the most ($\mathbf{e}_1$, the first eigenvector of the difference matrix $\mathbf{G}_{difference}$, accounting for 71.1% of the divergence in **G**) is not particularly aligned with $\mathbf{d}_{max}$, but the second eigenvector, $\mathbf{e}_2$ (accounting for 17.0% of the divergence in **G**) is. I discuss these results in more detail below.

Divergence in **G** was demonstrated by a significantly (and substantially) improved fit of a model allowing separate **G**-matrices by region compared to one that fit a single overall **G**. Eigenanalysis of the difference matrix ($\mathbf{G}_{difference}$), which holistically characterizes the differences in (co)variance structure between these regions, also identified three trait combinations that accounted for more variation (i.e., had a larger eigenvalue) than the corresponding eigenvectors of the null difference matrix calculated from **G**-matrices estimated with randomized data. In studies using quantitative genetic

models to predict responses to selection in diverging populations, **G** is often assumed to be constant (Duputié et al. 2012). How often this assumption holds, and over what timeframe, are critical empirical questions. These results add to growing evidence that **G** can evolve over even moderate timescales (Doroszuk et al. 2008; Eroukhmanoff and Svensson 2011; Walter et al. 2018; Chantepie et al. 2024), and emphasizes both the importance and challenge of incorporating variation in **G** into studies of divergence.

Ultimately, patterns of genetic (co)variance arise from the joint effects of the microevolutionary processes of selection, genetic drift, mutation, and gene flow (Steppan et al. 2002). How susceptible **G** is to change via these processes depends on the extent to which the underlying distribution of genetic effect sizes strays from that of the infinitesimal model (Lande 1980; Barton and Turelli 1987; Reeve 2000; Blows et al. 2004). Lab experiments have contributed importantly to our understanding of how these processes can impact **G** matrices (e.g., Phillips et al. 2001; Nosil et al. 2006; Doroszuk et al. 2008; Careau et al. 2015), but empirical studies of natural populations have provided varied and sometimes unexpected results. Some find remarkable stability in **G** despite likely variation in microevolutionary processes (Puentes et al. 2016; Henry and Stinchcombe 2023a), while others find evidence of divergence that suggests the action of drift (Hine et al. 2009) and/or selection (Eroukhmanoff and Svensson 2011; Walter et al. 2018; Chantepie et al. 2024). Here, I found divergence in **G** between regions where *D. subquinaria* are experiencing different selection regimes. Although I cannot rule out a contribution of other processes such as drift and mutation, these results add to a growing body of evidence consistent with variation in selection across environments playing a key role in driving divergence in **G** matrices.

A key aspect of my study design was the sampling of replicate populations within both the sympatric and allopatric regions. When characterizing phenotypic divergence (i.e. **D**), for example, this provided insight into between vs. within region variation. Unfortunately, I was unable to extend this strength to the quantitative genetic analyses; while region-specific estimates of **G** were well supported (i.e., robust model diagnostics and significant genetic variances for nearly all traits), population-level **G**-matrices were not, presumably due to smaller sample sizes decreasing power. This can be an issue in quantitative genetics studies where large sample sizes, particularly at the sire level, are necessary to estimate genetic (co)variances of a set of traits with precision, and similar studies have also encountered this problem (Walter et al. 2018; Henry and Stinchcombe 2023a). This issue can be exacerbated when working with a non-model species with low fecundity in the lab, limiting the ability to conduct large breeding designs. This underscores the challenges involved in generating robust quantitative genetics datasets for species undergoing adaptive divergence in nature, and it is likely one reason why these kinds of studies have been rare.

Given clear differences in **G** between regions, I then characterized how variances and covariances contribute to this. There was a substantial reduction in overall additive genetic variation, and hence evolutionary potential, in sympatry compared to allopatry: the sum of the trait variances in sympatry (i.e. the trace of **G**) was 44% of that in allopatry (Fig. 3.5A). This resulted from a reduced variance in sympatry for most, but not all, individual logcontrast CHCs (Fig. 3.5B). These differences were reflected in the first eigenvector of $\mathbf{G}_{\text{difference}}$ ($\mathbf{e}_1$), the trait combination that accounts for the majority of the divergence in **G** between regions, which had much higher genetic variance in allopatry

than sympatry ($\mathbf{e}_{1,\text{ allopatry}} = 0.532$, HPD interval: 0.422 – 0.654; $\mathbf{e}_{1,\text{ sympatry}} = 0.0695$, HPD interval: 0.0445 – 0.0975), and for which the trait loadings corresponded quite closely to the trait-specific differences in variance. An overall reduction in genetic variation could be the result of drift following founder events and/or reduced populations sizes in sympatry. *D. subquinaria* and *D. recens* are nearly identical morphologically and have no known ecological differences (they can be collected from the same mushrooms at the same time of year), so it is possible that habitats in sympatry support smaller populations of each species. However, contrary to an overall reduction in variation, there are two traits for which genetic variance is higher in sympatry (one significantly so), and this is reflected in $\mathbf{e}_2$, the second eigenvector of $\mathbf{G}_{\text{difference}}$, which has more genetic variance in sympatry than allopatry ($\mathbf{e}_{2,\text{ allopatry}} = 0.0285$, HPD interval: 0.0189 – 0.0403; $\mathbf{e}_{2,\text{ sympatry}} = 0.137$, HPD interval: 0.0971 – 0.179). Different patterns of variance change could also arise as a product of selection, with an increase occurring if initially rare variants (e.g., alleles previously held at mutation-selection balance) were brought under positive selection in sympatry, rising to intermediate frequencies and thereby increasing genetic variance in the direction of divergence (Turelli 1988; Agrawal et al. 2001; Blows and Higgie 2003; Conner et al. 2011; Uesugi et al. 2017). Hine et al. (2009) found a similar increase in genetic variance for a combination of CHC traits in a subset of populations of *Drosophila serrata* at the northern edge of a geographic cline, and interpret this as reflecting a combination of selection and drift acting on traits with a relatively simple genetic basis (i.e., impacted by a few genes of large effect; see also Sztepanacz and Blows 2017).

I also found evidence for changes in trait associations, which were tested via differences in genetic correlational structure given demonstrated differences in genetic variances. These differences were extensive: 95% HPD intervals did not overlap for 27 of the 36 correlations, 8 of which changed sign. The second eigenvector of $\mathbf{G}_{\text{difference}}$ appeared to capture some of these changes, with high loadings for the three traits for which correlational structure differed the most. Random mating was imposed for three of the four generations immediately prior to offspring being phenotyped in the breeding designs. This will reduce the impact of long-distance linkage disequilibrium (i.e., maintained by selection), and suggests that observed correlations are likely the result of physical linkage and/or pleiotropy. Divergence in trait correlations could simply reflect a secondary outcome of changes in trait-specific allele frequencies or could be the result of differing patterns of correlational selection between regions, a question that has not been studied in this system to date.

There is much interest in the extent to which patterns of genetic (co)variation influence divergence, and in particular whether $\mathbf{g}_{\text{max}}$ biases evolutionary responses over extended time scales (Hansen and Houle 2008; Agrawal and Stinchcombe 2009; Kirkpatrick 2009). If covariances are particularly strong such that genetic variation is concentrated in one or a few trait combinations (i.e. **G** is ill-conditioned) and these are not well aligned with selection, then **G** is expected to more strongly constrain evolutionary responses (Schluter 1996; Chenoweth et al. 2010). Alternatively, if **G** can evolve during adaptive divergence, such constraints may change, and could even be weakened (Eroukhmanoff 2009). A pattern of alignment between phenotypic divergence and $\mathbf{g}_{\text{max}}$ may arise in a few ways: if divergence occurs via drift (Lande 1979; Phillips et

al. 2001), if selection happens to align with $\mathbf{g}_{max}$, or if $\mathbf{g}_{max}$ evolves to come into alignment with selection (Eroukhmanoff 2009; Walter et al. 2018). In this case, multiple lines of evidence imply the adaptive divergence of CHCs between allopatry and sympatry, but $\mathbf{g}_{max}$ for these traits differs between regions, inconsistent with a simple scenario of divergence along a static line of least resistance. That the association of $\mathbf{d}_{max}$ with $\mathbf{g}_{max}$ has substantially increased in sympatry suggests that **G** may have evolved in these populations to bring it into greater alignment with the direction of divergence. This is supported by the association between the second eigenvector of $\mathbf{G}_{difference}$ ($\mathbf{e}_2$, which captures a component of differences in **G**), with $\mathbf{d}_{max}$ (Fig. 3.9B) and hence with $\mathbf{g}_{max,\ sympatry}$ ($\mathbf{e}_2 \times \mathbf{g}_{max,\ sympatry}$ angle: 31.3°; 95% CI: 27.8° - 34.6°). An estimate of the ancestral **G** for these populations is lacking, but if the allopatric region more closely reflects the pre-secondary contact state, these results suggest a scenario in which **G** has evolved in sympatry, at least in part, due to selection-induced increases in genetic variation for a particular trait combination. However, the major axis of differentiation in **G**, as described by $\mathbf{e}_1$, is not aligned with $\mathbf{d}_{max}$, suggesting other processes were involved in the divergence of **G**. Ultimately, such an interpretation assumes that $\mathbf{d}_{max}$ reflects selection, which is not necessarily the case. Independent estimates of selection in these populations will be an important next step for gaining additional insight into the dynamics of adaptive divergence in this system.

Finally, as an alternative, and commonly employed, technique for comparing **G** matrices I also performed a Krzanowski subspace analysis. This failed to detect any signal of significantly less (or more) sharing of the subspaces that contained the majority of the genetic variance in each region than that expected by sampling variation alone. The

null comparisons of **G** matrices created from the randomized data yielded point estimates for the eigenvalues of the first few eigenvectors of **H** that were quite high, providing ample opportunity for the real values (i.e., from the observed **G**-matrices) to be lower. However, the observed point estimates did not differ significantly from these and, in fact, indicated somewhat greater similarity than expected by chance. This contrasts directly with the substantial differences in **G** detected via the other methods I employed. Several recent studies investigating the stability of **G** during divergence over short-to-moderate timescales have detected no difference using Krzanowski's subspace analysis (Puentes et al. 2016; Henry and Stinchcombe 2023a), one of which also found separate evidence of divergence in **G** via other methods (Aguirre et al. 2014). This suggests that this approach may lack power when applied in this context, and that conclusions from earlier applications of this approach, particularly in the absence of corresponding null model comparisons, may be problematic.

In conclusion, I have characterized phenotypic divergence in CHCs between the sympatric and allopatric regions for *D. subquinaria* and found strong evidence that the genetic (co)variance structure for these traits has diverged as well. Not only has $\mathbf{g}_{max}$ shifted in orientation between regions, coming into greater alignment with phenotypic divergence in sympatry, but an axis of divergence in **G** is also highly aligned with phenotypic divergence. Together, these results suggest that **G** has evolved in sympatry following colonization from the ancestral allopatric zone, and therefore may not have substantially constrained divergence. These changes in **G** could be the result of selection in sympatry, and/or drift; multivariate estimates of selection on CHCs in each region would be helpful in evaluating these. Incorporating estimates of selection with our

understanding of the genetic (co)variance structure for these traits would also allow us to better explain the observed phenotypic divergence.

3.5 Tables

Table 3.1 Population collection and breeding design details. Separate breeding designs were performed for Deer Creek in blocks 2 and 3.

Population	Block	Latitude, Longitude	Region	Collection Years	# Isofemale Lines	# Sires	# Half-sib Families
Canmore	1	51.1, -115.3	sympatric	2019	12	78	228
Jocko	1	47.2, -113.8	allopatric	2019/2020	28	97	271
Sibbald	2	51.05, -114.9	sympatric	2021	21	113	311
Deer Creek	2, 3	46.8, -113.9	allopatric	2019/2020	14	92 (B2), 43 (B3)	212 (B2), 103 (B3)
Opal	3	50.8, -115.2	sympatric	2021	25	125	344
Lolo	3	46.8, -114.4	allopatric	2019/2020	19	129	358

Table 3.2 Characteristics of the first three eigenvectors of **D**, the among-population (co)variance matrix of trait means for logcontrast CHCs (LCs). Loadings greater than ± 0.25 are bolded for ease of interpretation.

			Trait loadings								
Eigen-vector	**Eigen-value**	**% Variance**	**LC2**	**LC7**	**LC11**	**LC12**	**LC13**	**LC14**	**LC17**	**LC19**	**LC20**
$\mathbf{d}_{max}$	0.092	79.5%	**-0.262**	-0.046	-0.071	-0.069	0.022	**-0.390**	**-0.652**	**-0.461**	**-0.360**
$\mathbf{d}_2$	0.021	18.1%	**0.396**	0.180	**-0.344**	0.164	**0.367**	-0.037	**0.414**	**-0.405**	**-0.442**
$\mathbf{d}_3$	0.002	2.1%	**0.679**	**-0.573**	0.185	0.111	0.207	-0.087	**-0.314**	0.106	0.060

Table 3.3 The eigenvalues, percent variance (calculated as the proportion of the sum of the absolute eigenvalues), and logcontrast CHC (LC) trait loadings for the eigenvectors of the difference matrix ($\mathbf{G}_{difference}$). Bold denotes loadings greater than ± 0.25.

eigenvector	eigenvalue	% σ^2	LC2	LC7	LC11	LC12	LC13	LC14	LC17	LC19	LC20
$\mathbf{e}_1$	0.465	71.1	-0.105	**-0.384**	**-0.424**	**-0.358**	**-0.310**	**-0.325**	0.141	-0.178	**-0.355**
$\mathbf{e}_2$	0.111	17.0	0.064	-0.005	-0.039	-0.092	**-0.267**	0.093	**0.446**	**0.418**	0.235
$\mathbf{e}_3$	0.045	6.9	-0.075	0.162	-0.053	0.026	-0.047	0.010	0.027	-0.019	-0.057
$\mathbf{e}_4$	0.014	2.2	-0.006	0.014	0.000	-0.002	-0.040	-0.027	-0.034	-0.014	0.046
$\mathbf{e}_5$	0.007	1.1	0.051	0.021	-0.031	0.009	-0.037	-0.025	-0.021	-0.029	0.051
$\mathbf{e}_6$	0.006	0.9	0.129	0.043	-0.049	-0.001	-0.052	-0.037	-0.042	-0.033	0.055
$\mathbf{e}_7$	0.003	0.4	0.235	0.084	-0.048	-0.008	-0.095	-0.073	-0.052	-0.060	0.063
$\mathbf{e}_8$	0.002	0.3	0.164	0.064	-0.047	0.016	-0.070	-0.061	-0.034	-0.038	0.041
$\mathbf{e}_9$	0.000	0.0	-0.070	0.009	0.005	0.035	-0.122	-0.100	-0.182	0.180	-0.006

3.6 Figures

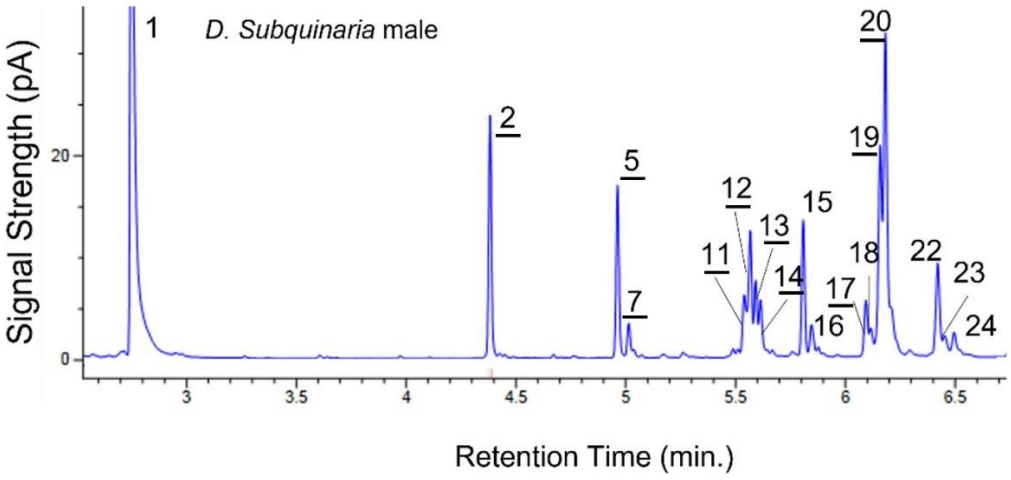

Figure 3.1 GC chromatographic profile of a male *D. subquinaria*. Each peak is a separate epicuticular compound (i.e. CHC), the integrated area of which reflects its concentration. Numbers correspond to those Curtis et al. (2013) found to be reliably present across all sampled individuals. The ten CHCs integrated in the current study are underlined here and are listed in Table S3.1, along with the chemical identities of all compounds.

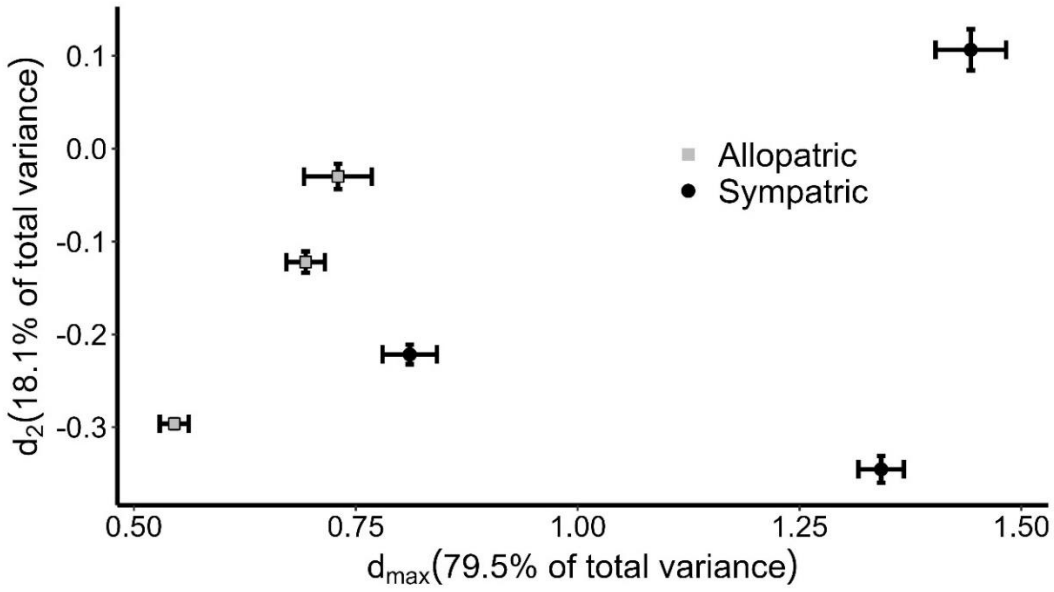

Figure 3.2 Variation among populations in mean values (± 95% CI) of $\mathbf{d}_{max}$ and $\mathbf{d}_2$. Allopatric population marked as gray squares, sympatric populations as black circles.

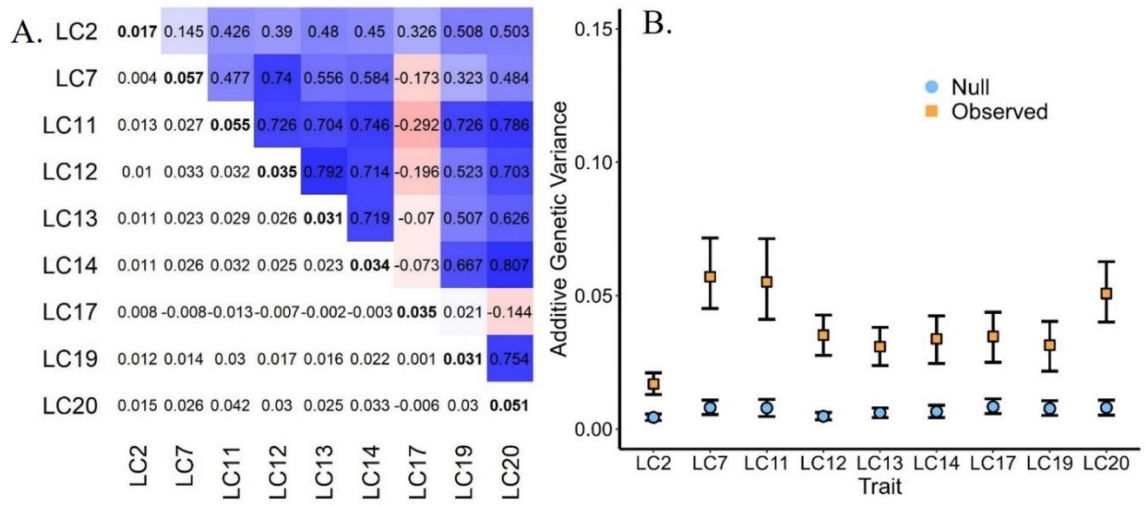

Figure 3.3 (A) The overall G-matrix for logcontrast CHCs (LCs), with genetic variances for each trait along the diagonal (bolded), genetic covariances in the lower left, and a heatmap of the corresponding genetic correlations in the upper right. **(B)** Additive genetic variance (V_A) for each trait as estimated by an MCMCglmm sire mixed effects model with observed data (orange squares) and randomized data (blue circles). Error bars are 95% highest probability density (HPD) intervals.

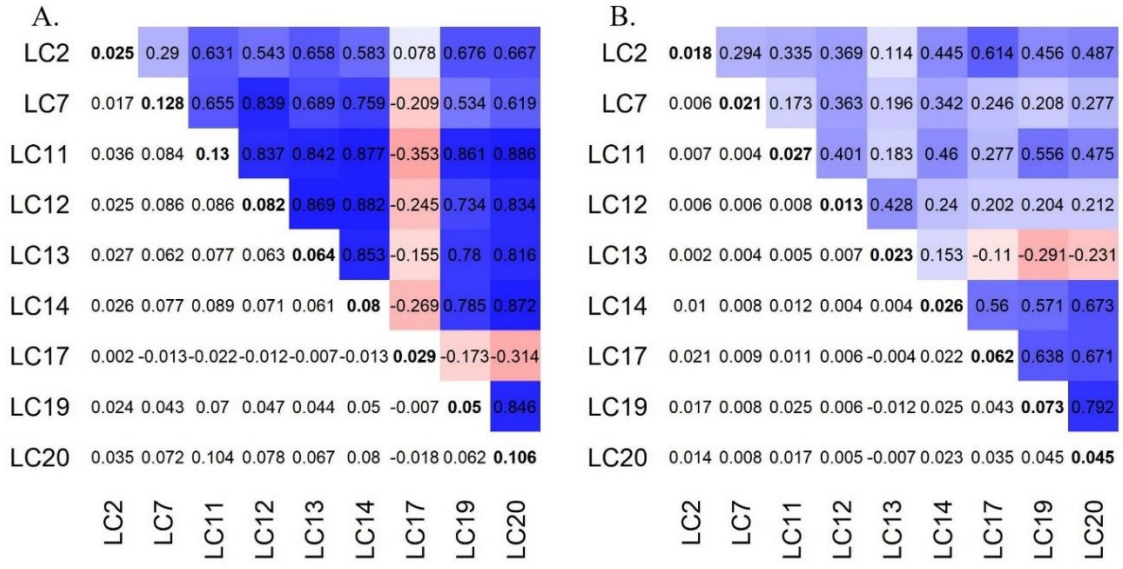

Figure 3.4 Posterior mean G-matrices for logcontrast CHCs (LCs) fit separately for populations from **(A)** allopatry and **(B)** sympatry. Genetic variances for each trait are along the diagonal (bolded), between-trait genetic covariances are in the lower left, and the corresponding genetic correlations are expressed as heatmaps in the upper right.

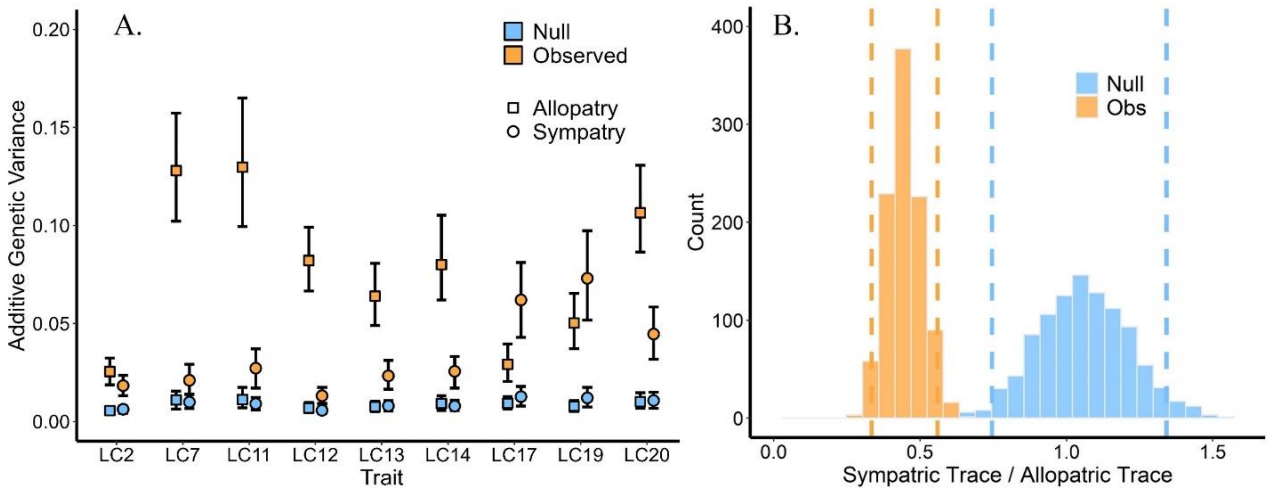

Figure 3.5 (A) Additive genetic variance (V_A) for each logcontrast CHC (LC) in allopatry (squares) and sympatry (circles), as estimated by an MCMCglmm sire mixed effects model with observed data (orange) and randomized data (blue). Error bars are 95% HPD intervals. **(B)** Posterior distributions for the ratio of sympatric trace to allopatric trace for observed (orange) and null (blue) data, where a value of one would indicate identical **G**-matrix size. The dashed lines are the 95% HPD intervals.

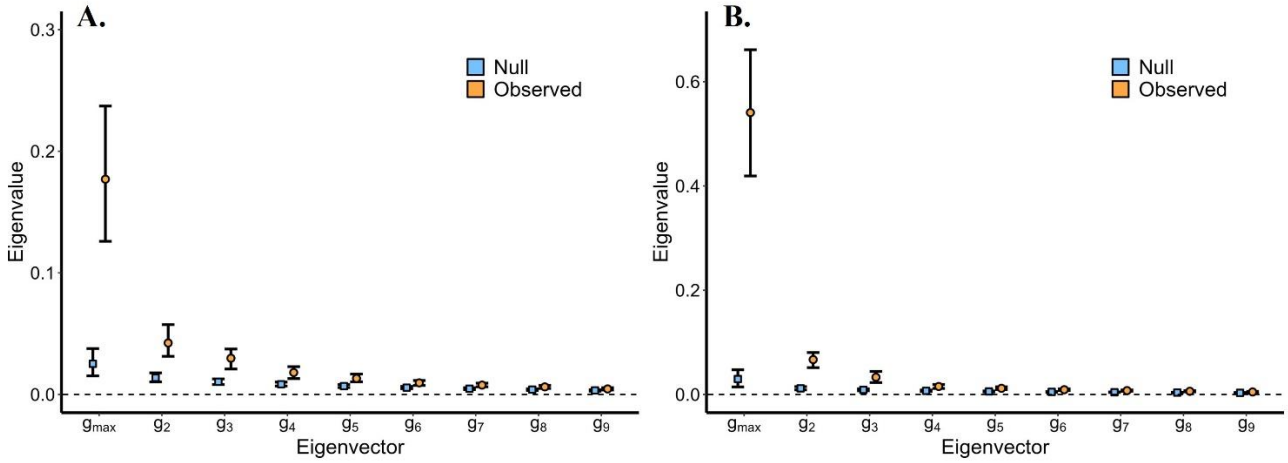

Figure 3.6 Eigenvalues for each eigenvector of the sympatric **G (A)** and allopatric **G (B)**, as estimated by an MCMCglmm sire mixed effects model with observed data (orange) and randomized data (blue). Error bars are 95% HPD intervals.

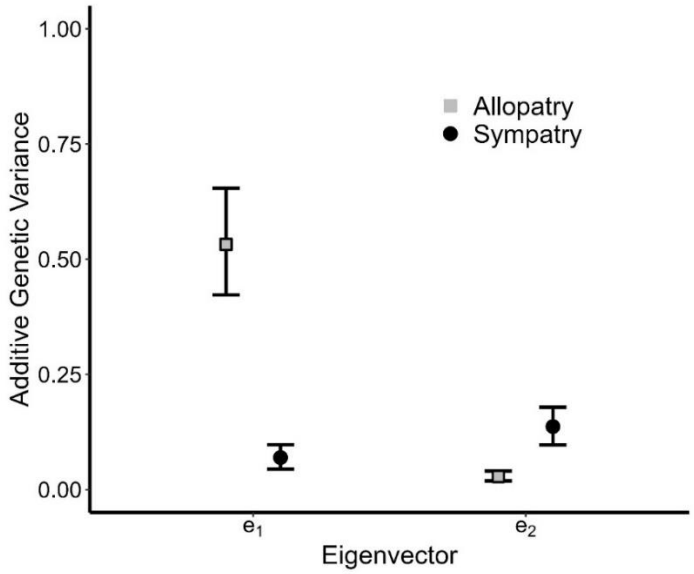

Figure 3.7 The mean additive genetic variance (V_A) in allopatry (grey squares) and sympatry (black circles) for the first two eigenvectors of the difference matrix ($\mathbf{G}_{difference}$). Error bars are the 95% HPD intervals.

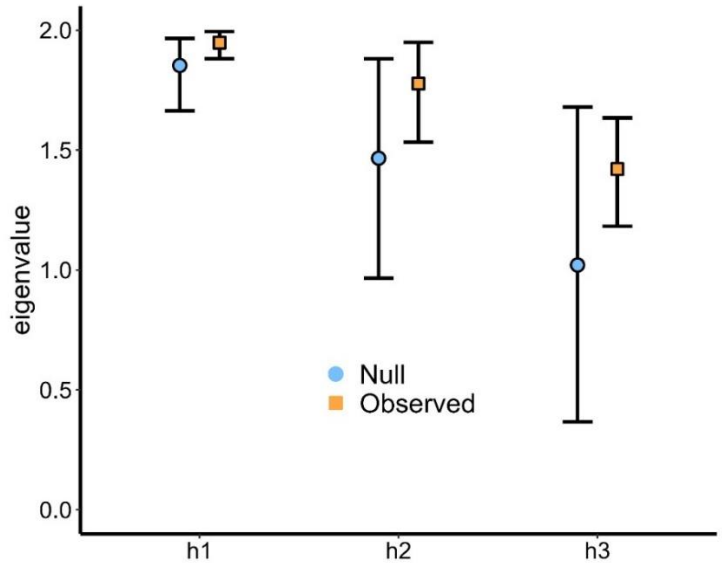

Figure 3.8 Results of a Krzanowski subspace comparison showing the eigenvalues for each eigenvector of the shared subspace **H** between sympatric and allopatric **G**-matrices estimated with observed data (orange squares) and randomized data (blue circles). Error bars are the 95% HPD intervals.

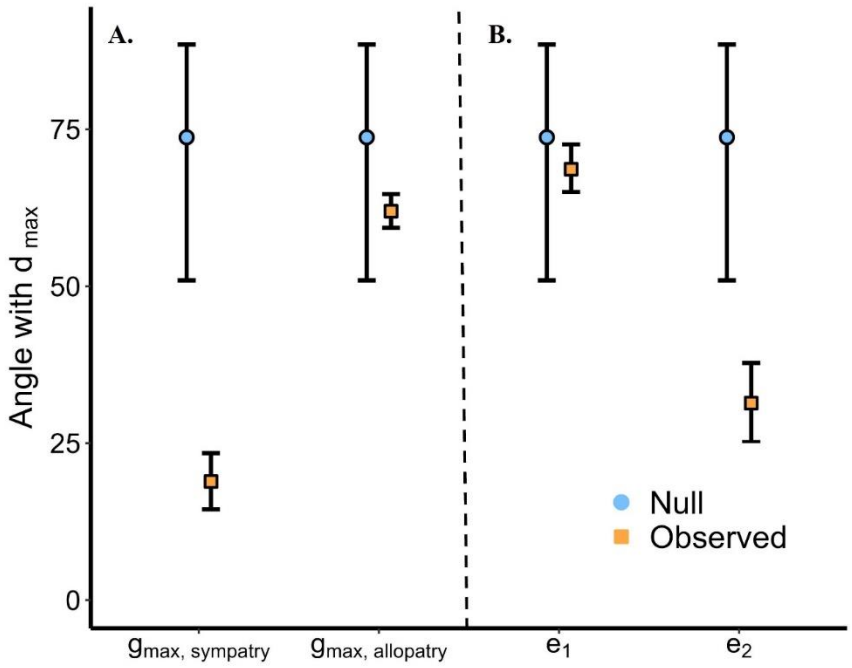

Figure 3.9 The posterior mean angles for **(A)** d_{max} with $\mathbf{g}_{max}$ for each region, and for **(B)** d_{max} with each of the first two eigenvectors of $\mathbf{G}_{difference}$. Angles between observed data in orange squares, and angles between 1000 pairs of random unit vectors blue circles. The error bars are the 95% quantiles.

Chapter 4: The contribution of sexual selection and genetic constraints to phenotypic divergence among natural populations of *Drosophila subquinaria*

The following thesis chapter is a modified version of a manuscript in preparation for submission to the *Journal of Evolutionary Biology*. The manuscript was modified to meet thesis formatting requirements. This work was done in collaboration with my supervisor, Howard Rundle. I took the lead role in all aspects of the work including study design, data collection and processing, analyses, interpretation, and writing.

4.1 Introduction

Divergent selection between contrasting environments is a major driver of phenotypic diversification and speciation in nature (Schluter 2000; Coyne and Orr 2004; Rundle and Nosil 2005; Richardson et al. 2014), but understanding past evolution and predicting future trajectories is complicated because selection generally acts on suites of traits which may be genetically correlated (Lande and Arnold 1983; Walsh and Blows 2009). This means that responses to selection are determined not only by the multivariate direction of selection, but also by the underlying genetic architecture of the traits (i.e. their genetic variances and covariances). This can constrain or bias evolutionary trajectories (Schluter 1996; Chenoweth et al. 2010), and in some cases it may actually facilitate divergence (Eroukhmanoff 2009; Walter et al. 2018). The extent to which genetic architecture shapes divergence of natural populations is not well understood, mainly because it is challenging to collect the necessary data on phenotypes, genetic architecture, and selection for a set of populations.

The evolutionary response to selection of a suite of traits ($\mathbf{\Delta z}$) depends on multivariate selection ($\boldsymbol{\beta}$) and the pattern of genetic variances and covariances of the traits ($\mathbf{G}$) that arise from pleiotropy and linkage. This relationship can be described by the multivariate breeder's equation, $\mathbf{\Delta z} = \mathbf{G}\boldsymbol{\beta}$ (Lande 1979). As with a response to selection in a single trait, the amount of genetic variance in the direction of selection is a key determinant of the magnitude of the response. If the major axis of genetic variance in $\mathbf{G}$ ($\mathbf{g}_{max}$) does not align with selection, then the evolutionary response will be biased towards a suboptimal direction (Schluter 1996; Arnold et al. 2001). However, if $\mathbf{g}_{max}$ does align with selection, or evolves to come into alignment, evolution can be accelerated.

It is still an open question as to what extent genetic constraints bias evolutionary trajectories, with some studies finding that adaptive divergence (Mitchell-Olds 1996; Schluter 1996; Bégin and Roff 2004; Chenoweth et al. 2010) or drift (Phillips et al. 2001) has been heavily constrained, while others find that divergence has proceeded with minimal influence from **G** (McGuigan et al. 2005). There is also mounting evidence that **G** itself can evolve over relatively short time scales, potentially reducing the impact of any genetic constraints if $\mathbf{g}_{max}$ comes into alignment with $\boldsymbol{\beta}$ (Doroszuk et al. 2008; Eroukhmanoff and Svensson 2011; Walter et al. 2018; Chantepie et al. 2024). For example, using experimental evolution with populations of *Drosophila serrata*, Blows and Higgie (2003) found that directional selection led to increased genetic variance in a set of reproductive traits, and that this was likely a case of rare alleles rising in frequency under persistent directional selection. An analysis of the relative strength of genetic constraints reported in the literature finds a wide range of effects, from strong constraints to facilitation of adaption (Agrawal and Stinchcombe 2009), indicating that the specific shape of the (co)variance structure and estimates of selection are needed to make inferences for any given case. However, a common shortcoming of many studies on how genetic (co)variances structure evolutionary trajectories is a lack of empirical estimates of selection and how this differs among populations, as well as information concerning among-population variation in **G**.

To better understand divergence among a set of populations in nature, we need robust estimates of the among-population divergence in multivariate phenotypes, together with population-specific estimates of selection and the quantitative genetic architecture of the traits. Population divergence can be due to differences in local selection and/or to

differences in the shape and orientation of **G**, which may filter responses to selection or drift in different directions. Evaluating the relative contributions of differences in selection and **G** to patterns of phenotypic divergence in nature is therefore a substantial empirical challenge, and such data are unsurprisingly limited. Several studies have made use of complimentary data sets on selection and genetic architecture, but these typically either measure selection in only one population and/or assume a constant **G** (e.g., Simonsen and Stinchcombe 2010; Lewis et al. 2011; Teplitsky et al. 2014; Reddiex and Chenoweth 2021; Oh and Shaw 2022; Henry and Stinchcombe 2023).

Here, I use a framework developed by Chenoweth et al. (2010), based on the work of Zeng (1988), that incorporates estimates of genetic (co)variances, multivariate phenotypic means, and selection from replicate populations. This approach seeks to explain divergence in a suite of traits among a set of populations, as described by the variance-covariance matrix of among-population multivariate phenotypic means (**D**). To this end, it assesses the relative contributions of differences in selection (quantified by **B**, the among-population variance-covariance matrix quantifying differences in directional selection gradients, i.e. $\boldsymbol{\beta}$s), genetic constraints as quantified by a single overall **G,** and differences in **G** among populations, in explaining the multivariate trait divergence quantified in **D**. It does this by comparing **D** first with **B**, and then with two among-population (co)variance matrices quantifying how trait means would respond to population-specific directional selection filtered either through a single, common **G**, or through population-specific **G** matrices. Applying this framework to sexually selected traits that have diverged between populations of *Drosophila serrata* in nature, Chenoweth et al. (2010) showed that divergence was biased away from the direction of sexual

selection by the genetic (co)variance structure, and further that accounting for among-population variation in **G** better explained the observed pattern of divergence. As far as I know, there have been no other studies that have fully applied this approach, likely due to the extensive nature of the data required.

I implement this approach to investigate how sexual selection, genetic architecture, and potential among-population differences in these constraints, contribute to an observed pattern of phenotypic divergence in a suite of contact pheromones among populations of *Drosophila subquinaria*. This species, which inhabits western N. America, is found both allopatric to, and sympatric with, the closely related species eastern N. American *Drosophila recens*. The sympatric region, on the eastern side of the Rocky Mountains, is a secondary contact zone formed in within the last 12,000 years (Jaenike et al. 2006; Bewick and Dyer 2014). Crosses between *D. subquinaria* and *D. recens* result in genetic incompatibilities that cause F1 hybrid male sterility, as well as a further ~90% offspring mortality when *D. subquinaria* females mate with *D. recens* males due to a cytoplasmic incompatibility from a *Wolbachia* infection carried by *D. recens* (Jaenike et al. 2006). Evidence indicates that selection in sympatry, generated by the presence of *D. recens*, has produced a pattern of reproductive character displacement in which sexual (i.e., behavioral) isolation with *D. recens* is much stronger in sympatry than it is in allopatry (Jaenike et al. 2006; Bewick and Dyer 2014; Dyer et al. 2018). This is further supported by a manipulative evolution experiment that recreated the reproductive character displacement by bringing *D. subquinaria* from allopatry into artificial sympatry with *D. recens* (see chapter 2). The high fitness cost to hybridization would imply reinforcement, although other negative heterospecific interactions, such as signal

interference, are also potential sources of selection generating reproductive character displacement.

A suite of (epi)cuticular hydrocarbons or CHCs, show a similar pattern of character displacement in *D. subquinaria* populations from sympatric vs. allopatric regions. These quantitative traits differ among populations from each region such that sympatric and allopatric flies each have a distinct multivariate CHC profile (Dyer et al. 2013). CHCs play a key role in desiccation tolerance in *Drosophila* and other insects (Howard and Blomquist 2005; Kwan and Rundle 2009), but are also contact pheromones that function as a means of chemosensory communication (Ferveur 2005). This often leads to trade-offs between viability and sexually selected trait optimums (Berson et al. 2019). In *D. subquinaria*, manipulative perfuming experiments and multiple behavioral assays have demonstrated that female mate preferences and species discrimination are based in part on variation in CHCs profiles (Curtis et al. 2013; Dyer et al. 2013; Giglio and Dyer 2013), and that sexual selection for these traits differs between sympatric and allopatric populations (Bewick and Dyer 2014; Rundle and Dyer 2015). A laboratory evolution experiment reveals that CHCs can evolve in response to experimental sympatry with *D. recens* (see chapter 2), consistent with selection generating divergence in these traits in nature. Additive genetic (co)variances of CHCs have also diverged between sympatry and allopatry (i.e., there are differences in **G** between regions), with changes in the amount and orientation of genetic variation, and the sign of some covariances (see chapter 3). These changes are associated with phenotypic divergence of CHCs between regions, suggesting that **G** may have evolved to come into alignment with selection in sympatry, however I heretofore lacked estimates of selection to test this more directly.

Here, I integrate estimates of sexual selection and genetic (co)variance structure for a set of six populations of *D. subquinaria*, three from sympatry and three from allopatry, with a focus on understanding the observed phenotypic divergence (i.e., character displacement) of CHCs. Genetic (co)variance structure for these populations was previously described in chapter 3. Here I add novel estimates of sexual selection on CHCs within each population. With these data in hand, I quantify the extent to which sexual selection on CHCs is divergent between regions. I then apply the methodology of Chenoweth et al (2010) to test whether among-population divergence in CHCs is best explained by sexual selection alone, by sexual selection and a single common **G**, or by sexual selection and region- or population-specific estimates of **G**. In doing this, I extend the Chenoweth et al. (2010) approach to incorporate not only uncertainty in the estimates of **G** (as was originally presented), but also in the estimates of selection and **D**, providing more appropriate insight into the significance (or not) of the difference components (see also Oh and Shaw 2022). Lastly, I assess the potential for genetic covariances to constrain (or promote) future adaptation in these populations by comparing the rate of adaptation under two scenarios – with vs. without genetic covariances – following the approach of Agrawal and Stinchcombe (2009).

4.2 Methods

Procedures concerning population collections, the breeding designs, characterization of multivariate phenotypic divergence, and **G** matrix estimation were previously described in Chapter 3. An overview is provided below; for additional details see Chapter 3

methods. Detailed methods are provided here for the mating trials, estimation procedures for selection gradients, and the implementation of the Chenoweth et al. (2010) method.

Populations and Breeding Design

I established isofemale lines in the lab using adult *D. subquinaria* females collected between 2019 and 2021 from three populations in the region west of the Rocky Mountains that is allopatric to *D. recens*, and from three populations in the region in central Canada that is sympatric with *D. recens* (see Table 4.1 for population details). Lines were maintained in 4 mL glass vials with food medium in a common environment (20° C, 75% humidity, 12 hr. L:D cycle) with non-overlapping generations.

Breeding designs were conducted using outbred stocks for each population created by combining the isofemale lines collected from the wild and had an average of 113.2 ± 8.9 SE sires per population, with 2.88 ± 0.03 SE half-sib families per sire. CHCs were extracted from three male offspring per half-sib family and quantified via gas chromatography with flame ionization detection on an Agilent 6890N dual-channel "fast" (220 V oven) gas chromatograph (Agilent Technologies, Wilmington, DE) following previously established methods (Curtis et al. 2013; Rundle and Dyer 2015). After identifying CHCs via qualitative alignment of chromatographic peaks with those in Curtis et al. (2013), the area under each of ten peaks (i.e., unique CHC compounds that could be reliably quantified) was then integrated to quantify their abundances. For statistical analyses, relative abundance (i.e., proportional concentrations) of each peak was taken and subsequently logcontrast transformed by dividing the proportional area of each peak by the proportional area of an arbitrarily selected common divisor (peak 5 in

this case) and taking the $\log_{10}$ of this quotient (Aitchinson 1983). The resulting nine logcontrast CHCs no longer have a unit-sum constraint and are all on the same scale, so these were used with no further standardization.

Quantifying phenotypic divergence and G-matrix estimation

D, the variance-covariance matrix of among-population multivariate phenotypic means, was estimated following the methods of Martin et al. (2008) using a MANOVA, where the nine logcontrast CHC traits were fit as a multivariate response, with experimental block, GC channel, and population as fixed effects. **D** was extracted from the sums of squares and cross-product matrix for the population effect, thereby controlling for the block and channel nuisance variables.

A single overall **G**-matrix was estimated, as well as separate **G**-matrices by regions (i.e., sympatric vs. allopatric) and by population, by implementing separate hierarchical mixed effects models in MCMCglmm (Hadfield 2010). Each **G**-matrix was 9×9 (45 parameters) and the models included fixed effects of population (where applicable) and gas chromatograph channel (two levels), random effects of sire and dam nested within sire, and a random effect of residual variance (Lynch and Walsh 1998; Chenoweth et al. 2010). **G** was estimated as four times the sire-level covariance matrix in each case and the posterior distribution of each **G** was carried forward in subsequent analyses to quantify uncertainty in their estimation. There was significant divergence in **G** between regions (based on comparison of nested mixed effect models fitting either a single or region-specific sire level (co)variance matrix, and assessing the significance of the

leading eigenvectors of the difference matrix, $\mathbf{G}_{sympatric} - \mathbf{G}_{allopatric}$; chapter 3), but there was insufficient power to confidently estimate population-level **G** matrices, so among-population differences in **G** within regions were not tested.

Mating trials

I conducted 145.5 ± 19.8 SE binomial mate choice trials per population concurrently with the breeding designs using the same block structure (Table 4.1; Fig. S4.1). Each trial had a single virgin female from a focal mixed population (created using the same procedure as the breeding designs; see Chapter 3) and two virgin males drawn from a reference population that was unique to a given block. To enhance phenotypic variation and hence power to detect selection, reference populations were created by mixing the isofemale lines from the corresponding sympatric and allopatric populations being tested in that block via two generations of a crossing scheme in which lines were numbered and then each was separately crossed with each of its two "neighboring" line numbers (e.g. male from line #1 with female from line #2, male from line #2 with female from line #3, etc.) in the first generation, followed by one generation of mass mating. Virgin males were collected from the reference population, and virgin females from the two mixed focal populations, using light CO_2 anesthesia, and were subsequently held for 7 days, separately by sex, with 10 individuals/vial.

Following the methods of Rundle and Dyer (2015), for each trial the focal female was placed in a standard drosophila vial with blended a mushroom-agar food medium that stimulates female mating behavior, and then two males from the reference population

(but separate holding-vials) were simultaneously added by aspiration, without any CO_2 anesthesia. Vials were then observed until intromission occurred with one of the two males, at which point all flies were anesthetized with light CO_2. To preserve independence among males in the data set, CHCs were extracted from either the successful or unsuccessful male from a given trial (randomly predetermined, 50:50 ratio) and the other male, and the female, were discarded. If mating did not occur within 30 min., the trial was terminated, and the flies were discarded. All mating trials and CHC extractions were conducted in the mornings within 4 h of incubator lights turning on, alternating between the two populations from the different regions every ten trials to minimize temporal confounds given that CHCs can vary in a circadian pattern (Gershman et al. 2014). Trials were conducted over four days within each block. I then processed, integrated, and logcontrast transformed these CHC samples using the same methods as described above, and I retained the same set of nine logcontrast CHCs that were used in previous analyses.

Selection and genetic constraints in population divergence

To quantify sexual selection separately within each population, I used general linear models, fit via least squares, that regressed mating success (fitness) on the nine logcontrast CHCs (Lande and Arnold 1983; Rundle et al. 2008) in a given population. The partial regression coefficients for each trait form a vector of selection gradients (***β***) quantifying multivariate sexual selection on this suite of traits for a given population. The among-population sexual selection variance-covariance matrix (**B**) was calculated directly from the ***β*** vectors of the six populations as the variances and covariances of the

gradients for the nine traits. Given the binomial nature of the response variable (mating success), overall significance of sexual selection, as well as the significance of the selection gradients on the individual traits, was evaluated via a generalized version of the same model, specifying a binomial distribution with a logit link function (Rundle et al. 2008). I tested the significance of block, gas chromatography channel, and assay day in all models, but none of these had significant effects on mating success, and so they were excluded from all models (including the general linear models estimating selection). The generalized linear models were fit via maximum likelihood as there were no random effects to account for.

To test whether sexual selection on CHCs varied significantly between regions (i.e. allopatry vs. sympatry), I used a generalized model, fit via maximum likelihood, that combined the data for all six populations. Region and its interaction with each trait were included as fixed effects, and this model was compared against a reduced model that excluded the interaction effects using a likelihood ratio test. To test for differences in sexual selection among the three populations within a given region, I again fit separate nested generalized models with the nine logcontrast CHCs for the sympatric and allopatric region with either population and its interaction with each trait, or with the interactions removed in the reduced model. A likelihood ratio test comparing the full and nested models within each region was used to assess overall significance. I checked for nonlinear selection in each population by including the second-order polynomials for all traits in the model (i.e., the quadratic and correlational terms). Nonlinear selection could not be confidently estimated in all cases due to a lack of power, but when it was estimated, it never resulted in a significant improvement when comparing nested models

that included vs. excluded the second-order terms with a likelihood ratio test (Table S4.1).

To assess the similarity between observed phenotypic divergence and predicted responses to selection under potential patterns of genetic architecture, we first calculate a series of three different among-population (co)variance matrices ($\mathbf{R}_G$) that represent the variances and covariances among the trait-specific predicted responses. These each used a different **G** matrix to estimate the population-specific responses to selection that were then used to calculate $\mathbf{R}_G$. In particular, I used either a single, overall **G** for all populations ($\mathbf{R}_{single}$), separate **G** matrices for each region ($\mathbf{R}_{region}$), or separate **G** matrices for each population ($\mathbf{R}_{pop}$). Following Chenoweth et al. (2010) and Zeng (1988), $\mathbf{R}_{single}$ was calculated as:

$$\mathbf{R}_{single} = \mathbf{G}_{single}\mathbf{B}\mathbf{G}_{single},$$

where $\mathbf{G}_{single}$ is the overall additive genetic variance-covariance matrix for all populations, and **B** is the among population variance-covariance matrix of sexual selection gradients. $\mathbf{R}_{region}$ and $\mathbf{R}_{pop}$ were calculated slightly differently. First, by applying the multivariate breeder's equation $\boldsymbol{\Delta Z}_k = \mathbf{G}_k\boldsymbol{\beta}_k$ in each population using the region- or population-specific **G** and the vector of sexual selection gradients ($\boldsymbol{\beta}$) for that population. Then, taking the set of six individual predicted responses to selection ($\boldsymbol{\Delta Z}_k$), I calculated the variance-covariance matrix among these populations in the predicted response. This method can also be used to estimate $\mathbf{R}_{single}$ and gives the same results as above.

Next, to quantify the roles of selection and genetic constraints in explaining the observed pattern of phenotypic divergence (**D**), I compared **D** to each of these variance-covariance matrices (i.e. **B**, $\mathbf{R}_{single}$, $\mathbf{R}_{region}$, $\mathbf{R}_{pop}$) using Krzanowski's subspace analysis

(Krzanowski 1979). This is a commonly employed geometric approach to comparing the similarity in orientation of two or more matrices (Blows et al. 2004). This method is appropriate here as the matrices being compared are all of the same dimensions but characterize different information (Chenoweth et al. 2010). First, the primary subspace for each matrix is calculated from a set of eigenvectors that account for the majority of total variance. Following Chenoweth et al. (2010), I used eigenvectors that accounted for at least 95% of the variance (the first three eigenvectors in this case). A common subspace, **S**, is then calculated as:

$$\mathbf{S} = \mathbf{L}^{T}\mathbf{M}\mathbf{M}^{T}\mathbf{L},$$

where **L** and **M** are each the subset of eigenvectors for a matrix, and the superscript *T* indicates the matrix transposition. This projection takes the best-matching vectors from the two matrix subspaces being compared and computes the angles between them as the projection matrix **S**. The eigenvectors of **S** will have eigenvalues between 0 and *p*, where *p* is the number of original matrices (two in this case), while the sum of the eigenvalues of **S** ($\Delta\lambda_{\mathbf{S}(\mathbf{L},\mathbf{M})}$) will be bounded by the number of eigenvectors used in the subset of the original matrices (three in this case). Higher values indicate greater similarity of the primary subspaces of the two matrices, while a value of zero would indicate the best matching eigenvectors of those subspaces are orthogonal. I calculated **S** and $\Delta\lambda_{\mathbf{S}(\mathbf{L},\mathbf{M})}$ for separate comparisons of **D** with **B**, with $\mathbf{R}_{\text{single}}$, with $\mathbf{R}_{\text{region}}$, and with $\mathbf{R}_{\text{pop}}$. In each case, I carried forward uncertainty in the estimation of these parameters, using the posterior distribution for each **G** and via 1,000 bootstrap samples of **D** or $\boldsymbol{\beta}$, the latter created by resampling individuals within each population with replacement and then recalculating **D** or $\boldsymbol{\beta}$ for each iteration. Propagating uncertainty in all parameters allowed us to calculate

95% quantile intervals around each estimate of matrix similarity, providing a means of statistical inference for changes in the degree of similarity for each comparison that incorporated all sources of uncertainty.

To further assess the alignment between the direction of multivariate selection and the major axes of phenotypic divergence and genetic variance, I calculated the pairwise angles in degrees between either an average overall $\boldsymbol{\beta}$ or population $\boldsymbol{\beta}$s with $\mathbf{d}_{max}$ and each of $\mathbf{g}_{max,\ single}$, $\mathbf{g}_{max,\ sympatry}$ and $\mathbf{g}_{max,\ allopatry}$. As $\mathbf{d}_{max}$ and $\mathbf{g}_{max}$ vectors have no inherent directionality, angles greater than 90º are subtracted from 180º. For each case I used the appropriate distribution of uncertainty, either the posterior distribution for eigenvectors of **G**s or 1000 bootstrap samples of $\mathbf{d}_{max}$ and $\boldsymbol{\beta}$. From these I calculated the 95% quantile intervals around each angle. An angle close to 0 would indicate strong alignment.

Finally, I use the approach proposed by Agrawal and Stinchcombe (2009) to assess the relative strength of genetic covariances to constrain future adaptive divergence. First, I calculated the predicted responses to selection in each population via the multivariate breeder's equation. As we have evidence that **G** has diverged between regions (chapter 3) while population-level **G** matrices are not well estimated, I used region-specific **G** matrices to calculate the predicted responses to selection. This is then repeated but with all of the genetic covariances set to zero. In each case, the change in fitness for the mean phenotype over a single generation is then calculated as:

$$\Delta \boldsymbol{W}(\bar{\boldsymbol{z}}) = \Delta \bar{\boldsymbol{z}}^{\boldsymbol{T}} \boldsymbol{\beta},$$

where $\bar{\mathbf{z}}$ is the mean multivariate phenotype, $\boldsymbol{\beta}$ is the population vector of sexual selection gradients, and $\boldsymbol{\Delta W(\bar{z})}$ is the rate of adaptation. This involves a number of assumptions,

including that curvature in the fitness surface is weak and therefore that the fitness of the mean phenotype is similar to the mean fitness of the population (in other words, that nonlinear selection is weak). Here, I did not detect evidence of nonlinear selection (Table S4.1), although my power to do so was low. Lastly, I calculated the ratio (R) of the rates of adaptation under these two scenarios (i.e., the **G** matrix including observed genetic covariances vs. when the genetic covariances are zero). This ratio provides a relative measure of the strength of genetic constraints, where values of R less than one indicate the rate of adaptation is slowed by the pattern of genetic covariances, while a value greater than one implies that adaptation is facilitated. Significance was assessed by carrying forward the uncertainty in estimates of **G** and $\boldsymbol{\beta}$ to calculating the 95% quartiles for each ratio.

4.3 Results

Characterizing phenotypic divergence

Mean CHC phenotype diverged significantly among populations (MANOVA: Wilks' λ = 0.248, $F_{4,1819} = 85.01$, $p < 0.0001$). The first eigenvector of **D**, $\mathbf{d}_{max}$, accounted for the majority (79.50%) of variance among population means (Table 4.2), and primarily reflects variation in the relative abundance of longer chain CHCs and logcontrast CHC2. Most of the remaining variance (18.10%) is accounted for $\mathbf{d}_2$, the second eigenvector of **D**. Patterns of differentiation among populations in $\mathbf{d}_{max}$ and $\mathbf{d}_2$ reveal that $\mathbf{d}_{max}$ corresponds to divergence between sympatric vs. allopatric regions, while $\mathbf{d}_2$ describes an unknown aspect of among-population differentiation that is not associated with region (see Chapter 3, Table 3.2).

Characterizing sexual selection

Linear sexual selection on CHCs was significant overall in five of the six populations, with from one to five of the selection gradients on individual logcontrast CHCs showing significance depending on the population (Table 4.3). Selection coefficients on three logcontrast CHCs showed some evidence that they differed between regions (i.e., sympatry vs. allopatry), with one of these significant and the other two marginally nonsignificant (Table 4.4). However, accounting for this variation did not significantly improve the overall model fit ($\chi^2_{d.f.=9} = 12.48$, $p = 0.187$). These three logcontrast CHCs are longer chain hydrocarbons that also differed in mean value between regions (Table S4.2), and in the case of LC20, it was under both stronger positive selection and at higher relative abundance in allopatry. I found significant variation in selection overall among populations within regions, both in sympatry ($\chi^2 = 33.45$, $d.f. = 18$, $p = 0.015$) and in allopatry ($\chi^2 = 52.70$, $d.f. = 18$, $p < 0.0001$), however, population here is confounded with experimental block and so may reflect variation due to collection year and/or lab conditions at the time of the mating trials. Among all six populations, variation among-populations in sexual selection gradients was dominated by a single trait combination ($\mathbf{b}_{max}$), which accounted for 77.03% of the variance, while $\mathbf{b}_2$ accounted for most of the remainder (Table 4.2).

Relative contributions of selection and genetic (co)variance to divergence

The observed pattern of divergence (**D**) in CHCs was not well explained by differences among populations in sexual selection alone ($\Delta\lambda_{\mathbf{S(D,\,B)}}$ = 0.720 of a possible 3, or 24.0% of the maximum; Table S4.3; Fig. 4.1). This was substantially improved (approximately doubled) by filtering sexual selection through a single overall **G** to estimate the among-population variation in the predicted responses to selection ($\Delta\lambda_{\mathbf{S(D,\,Rsingle)}}$ = 1.431, or 47.7% of the maximum), although the improvement was marginally non-significant as inferred by the small overlap of the 95% quantiles. There was a small further mean improvement when accounting for divergence in **G** between regions (i.e. predicting the response using separate **G** matrices in allopatry vs. sympatry: $\Delta\lambda_{\mathbf{S(D,\,Rregion)}}$ = 1.495, or 49.8% of the maximum). Uncertainty was somewhat reduced, however, resulting in a significant improvement over sexual selection alone (Fig. 4.1). Incorporating population specific **G** matrices reduced similarity with observed divergence ($\Delta\lambda_{\mathbf{S(D,\,Rpop)}}$ = 1.287, or 42.9% of the maximum), likely because population specific **G**s were poorly estimated. Sexual selection gradients, whether an average $\boldsymbol{\beta}$ for all populations or population-specific $\boldsymbol{\beta}$s, were not aligned with the major axes of phenotypic divergence ($\mathbf{d}_{max}$) or genetic variance ($\mathbf{g}_{max}$): each of the pairwise angles were between 70º and 85º, and their 95% quantiles overlapped 90º.

As an overall measure of the extent to which genetic covariances may constrain further evolutionary responses to the observed sexual selection, I calculated the predicted responses to selection under two scenarios that included the observed genetic covariances among traits or set these to zero. Results revealed that the rate of adaptation would be constrained by the pattern of genetic covariances (Fig. 4.2), as I found estimates of R, the response ratio, below one for all populations. This was particularly

strong for the sympatric populations, where the 95% quantiles around R were well below 1 and indicated that on average the rate of adaptation would be slowed by around 54.3%. In allopatry, two out of three populations showed no significant evidence of a strong constraint as their 95% quantiles overlapped 1, although both point estimates were negative; the other population, Deer Creek, appears to be heavily constrained by the allopatric **G** matrix (slowed by 77.2%).

4.4 Discussion

I estimated linear sexual selection gradients on male CHCs in the same six populations of *D. subquinaria* for which I have data on phenotypic divergence and genetic (co)variances and found significant selection in five of six populations. While the pattern of selection differed among populations, sexual selection alone did not well explain the pattern of divergence in CHC phenotypes among populations. Accounting for genetic (co)variances, whether assuming a single overall **G**-matrix or region-specific **G**-matrices, substantially improved explanatory power for the observed pattern of divergence, suggesting that genetic constraints have biased evolutionary divergence. These results also suggest that future adaption under this sexual selection regime is likely constrained, particularly in the sympatric populations. There, the ratios of the rate of adaption under the observed pattern of genetic (co)variances compared to the rate under a hypothetical scenario without genetic covariances (denoted R) were significantly below one in all three populations, indicating that further adaptation would be slowed.

Our finding that accounting for genetic (co)variances among theses traits improves our ability to explain the pattern of among-population phenotypic divergence suggests that genetic constraints have played a role in shaping divergence. Chenoweth et al. (2010) similarly found that variation in selection only weakly resembled the observed pattern of CHC divergence in *D. serrata*, and that accounting for genetic (co)variances significantly improved the resemblance. If the major axes of genetic variation are not aligned with the direction of selection, then the response to selection will be altered, and the magnitude of the response will be constrained (Schluter 1996; Arnold et al. 2001). Here, the average direction of sexual selection was not well aligned with among-population divergence in CHC phenotypes, nor with the major axes of genetic variance. In chapter 3, I found that the size and orientation of **G** differed between regions and that in sympatry, **G** had changed to align with the direction of among-population phenotypic divergence. There, I speculated that selection may have induced **G** to come into alignment, but I do not find evidence of that here. It is possible that **G** is responding to a different source of selection, or that I have not accurately characterized sexual selection in these populations. Alternatively, divergence in **G** between regions may be the result of founder effects in sympatry, followed by drift along $\mathbf{g}_{max}$ resulting in phenotypic divergence. However, this is not consistent with other work on this system, where there is strong evidence that reproductive character displacement is a direct result of selection induced by the presence of *D. recens* (Rundle and Dyer 2015; Jarvis et al. 2024).

While I found that the pattern of sexual selection varied among populations, I did not find convincing evidence that it differed strongly between regions. This contrasts somewhat with previous studies, as Rundle and Dyer (2015) found that average linear

sexual selection on CHC trait combinations (identified with PCA) was divergent between females from sympatry vs allopatry. Dyer et al. (2014) also demonstrated that sympatric *D. subquinaria* females (in contrast with allopatric females) were more likely to mate with males from either species that had been perfumed with the CHCs of sympatric male *D. subquinaria*. While the test of overall differences in sexual selection between regions was not significant, three individual CHCs showed some evidence of differences in selection (CHCs 14, 19, and 20). In particular, the relative concentration of CHC 20 was under much stronger, positive selection in allopatry than in sympatry. Unfortunately, it is not straightforward to compare these results to those of Rundle and Dyer (2015) because they use a larger set of male CHC traits (17 vs 9 in this study), they expressed CHCs as principal component trait combinations to assess sexual selection (because multicollinearity among this larger set of traits was very strong), and they used a different transformation (center log-ratios instead of logcontrasts) that is more appropriate when working with principal components (Aitchison 1982; Rundle and Dyer 2015). It is possible that I excluded CHC traits that, although lower in relative abundance (and hence providing less precise estimates), were important to differences in selection. In Rundle and Dyer (2015), CHC 20 loaded positively on PC4 and was under positive selection in allopatry, which is consistent with the results here. However, as these are different trait combinations it is difficult to draw conclusions from this. It is also important to note that populations within a region are confounded with experimental block in the current data (Table 4.1), so differences among populations may partially reflect variation due to differences in lab conditions at the time of the mating trials and collection year.

Here, and in previous studies on *D. subquinaria* reproductive character displacement, I have only attempted to assess variation in sexual selection. This is largely because it is assumed that sexual selection is easier to quantify in the lab and is a key component of reproductive character displacement in this species. However, as CHCs are important in moisture regulation (Howard and Blomquist 2005), they are potentially targets of natural selection as well. Lab evolution experiments have shown strong divergence in CHC traits for *D. subquinaria* populations exposed to desiccation treatments (Kwan and Rundle 2009). If desiccation stress differ between regions, this would generate selection on CHCs that may not align with sexual selection and would mean that we are missing a key component of total selection. Additionally, if there is some form of reproductive interference by *D. recens* that alters female preferences or causes plastic changes in male *D. subquinaria* CHC signals, then I may not be accurately estimating sexual selection on CHCs as all mating trials were done in the absence of *D. recens,* as well as other fly species that commonly co-occur with *D. subquinaria* on mushrooms in the wild. If the social environment alters male sexual displays and/or female mate choice for these, these estimates may not recreate sexual selection in the wild. This is an issue shared by Chenoweth et al. (2010), as well as a large body of work on mate choice of CHCs (Blows et al. 2004; Chenoweth and Blows 2005; Rundle et al. 2008).

Our results suggest that future evolutionary responses to sexual selection in these populations is likely to be constrained by the pattern of genetic (co)variances, particularly in sympatry. There, the predicted rate of adaptation was roughly 54% slower when accounting for region-specific genetic (co)variances, a relatively strong genetic

constraint on adaptation. Henry and Stinchcombe (2023) also found evidence that the rate of adaptation was constrained in three of the four studied populations of *Ipomoea hederacea* (ivyleaf morning glory), but only on average by 30.4%. Agrawal and Stinchcombe (2009) showed that across the literature reviewed, genetic constraints were about as equally likely as cases where genetic (co)variances facilitate adaptation, with a mean R ratio of 0.89. The strong constraint on future adaptation here could reflect a depletion of genetic variance in the direction of sexual selection through its persistent action (e.g. Blows et al. 2004). Several studies demonstrate that sexual selection can cause the evolution of antagonistic pleiotropy, where further increases in sexual fitness, come at a cost to nonsexual fitness (Hine et al. 2011; Delcourt et al. 2012). This could account for the lack of alignment we see between our estimates of sexual selection and the trait combination with the greatest genetic variance.

Following the implementation in Chenoweth et al. (2010), I used Model II from Zeng (1988) which assumes that stabilizing selection is absent and allows for divergence to be influenced by the structure of genetic (co)variances over short or moderate timescales. This importance of genetic (co)variances in shaping responses to selection is supported in general because of the abundance of traits in nature with genetic variance, despite the action of selection, which should deplete variance (Walsh and Blows 2009). In contrast, Model I from Zeng (1988) assumes the action of both directional and stabilizing selection and predicts that, over longer time periods, the pattern of divergence will come to reflect the among-population variation in directional selection (i.e. that constraints become less important over extended time periods). Some theory suggests that, due to the ultimately high dimensionality of **G** for many species, genetic constraints

will be rare and that even traits with low genetic variance can evolve over the long-term (Arnold 2023). Implementing Model I requires reliable estimates of stabilizing (nonlinear) selection. With nine traits, and hence 36 quadratic gradients, substantial replication is needed to obtain robust estimates of nonlinear selection, which I simply did not have. It remains unclear at what time scale the influence of genetic (co)variances on divergence may be reduced, although some studies have found divergence along highly conserved genetic lines of least resistance over 20-40 million years (McGlothlin et al. 2018). Better replicated estimates of selection would nevertheless be useful in allowing Model 1 to be investigated.

In conclusion, to my knowledge this is only the second full implementation of the approach proposed by Chenoweth et al. (2010), which incorporates variation among populations in phenotypes, selection, and genetic (co)variances. Building on this original application, we carry through uncertainty in each of these parameters to better assess the significance of each comparison between the observed pattern of divergence and variation in selection or predicted responses accounting for genetic (co)variance. I found that accounting for genetic (co)variance, and in particular its variation between regions, improved our ability to explain the pattern of divergence in male *D. subquinaria* CHCs, suggesting an important role of the **G**-matrix in shaping evolutionary responses. I also found that future adaptation in sympatry is likely to be similarly constrained, as the genetic 'line of least resistance' is not aligned with selection. However, I have not estimated sexual selection in the presence of *D. recens* and therefore may be missing an important aspect of selection in this species.

4.5 Tables

Table 4.1 Population collection details. The number of mating trials is the number retained in the final analysis. Block is the experimental block for which mating trials were conducted for a pair of populations (see chapter 3).

Population	Block	Latitude, Longitude	Region	Collection Years	# Isofemale Lines	# Mating Trials
Canmore	1	51.1, -115.3	sympatric	2019	12	194
Jocko	1	47.2, -113.8	allopatric	2019/2020	28	189
Sibbald	2	51.05, -114.9	sympatric	2021	21	83
Deer Creek	2	46.8, -113.9	allopatric	2019/2020	14	82
Opal	3	50.8, -115.2	sympatric	2021	25	137
Lolo	3	46.8, -114.4	allopatric	2019/2020	19	188

Table 4.2 The first two eigenvectors of matrices characterizing among population variation in CHCs (**D**), sexual selection on these traits (**B**), predicted responses to selection assuming a constant **G** ($\mathbf{R}_{single}$), and predicted responses to selection assuming **G** varies by region ($\mathbf{R}_{region}$).

	D		**B**		$\mathbf{R}_{single}$		$\mathbf{R}_{region}$	
Trait	$\mathbf{d}_{max}$	$\mathbf{d}_2$	$\mathbf{b}_{max}$	$\mathbf{b}_2$	$\mathbf{r}_{single,max}$	$\mathbf{r}_{single,2}$	$\mathbf{r}_{region,max}$	$\mathbf{r}_{region,2}$
% variance explained	79.50	18.10	77.03	12.95	89.50	7.62	94.90	2.41
LC2	-0.262	0.396	-0.155	0.010	-0.086	-0.275	-0.153	-0.278
LC7	-0.046	0.180	0.056	-0.183	-0.330	0.644	-0.337	0.692
LC11	-0.071	-0.344	-0.511	-0.108	-0.487	-0.478	-0.492	-0.268
LC12	-0.069	0.164	0.439	0.212	-0.332	0.316	-0.338	0.333
LC13	0.022	0.367	-0.073	-0.250	-0.262	-0.053	-0.290	0.096
LC14	-0.390	-0.037	0.094	0.331	-0.297	-0.048	-0.350	0.060
LC17	-0.652	0.414	-0.226	0.529	0.258	-0.278	0.088	-0.216
LC19	-0.461	-0.405	0.609	-0.374	-0.307	0.108	-0.282	-0.298
LC20	-0.360	-0.442	-0.286	-0.564	-0.467	-0.295	-0.462	-0.336

Table 4.3 Linear (i.e. directional) sexual selection gradients on logcontrast (LC) CHCs for each population as estimated with multiple (general) linear regression, with the associated coefficient of determination (R^2_{adj}). Overall model significance, and significance of individual gradients, were assessed by fitting binomial generalized linear models with a logit link function (see Methods). Bold entries denote significant gradients.

	Allopatry			**Sympatry**		
Population	**Jocko**	**Deer Creek**	**Lolo**	**Canmore**	**Sibbald**	**Opal**
Model significance	**<0.001**	**<0.001**	**0.0239**	**0.003**	0.133	**0.0297**
$\mathbf{R^2_{adj}}$	0.221	0.228	0.051	0.077	0.050	0.061
LC2	-0.136	0.638	0.042	-0.254	-0.685	-0.461
LC7	0.603	0.296	0.406	0.270	-0.213	**0.599**
LC11	-0.543	1.564	**1.928**	0.139	1.141	-0.539
LC12	0.852	-1.491	**-1.388**	-0.800	-0.047	0.316
LC13	0.412	0.687	0.888	0.790	-0.038	0.407
LC14	-0.419	-0.504	**-1.191**	-0.314	-0.252	-0.355
LC17	**-1.054**	**0.738**	-0.745	**-0.570**	-0.037	-0.724
LC19	0.179	**-2.755**	**-1.419**	-0.412	**-0.960**	**0.850**
LC20	0.446	**1.367**	**2.456**	0.842	0.709	0.659

Table 4.4 Results of LRT model comparisons testing differences in sexual selection between regions, and among populations within regions. Binomial generalized linear models with a logit link function were used (see Methods). Bolded entries denote significance.

	Among Allopatric Populations		**Among Sympatric Populations**		**Between Regions**	
	χ^2	*p*	χ^2	*p*	χ^2	*p*
Model significance	**52.70**	**<0.001**	**33.45**	**0.015**	12.49	0.187
LC2	1.79	0.408	0.47	0.790	0.74	0.390
LC7	0.70	0.706	1.86	0.395	0.07	0.789
LC11	**7.10**	**0.029**	3.96	0.138	1.28	0.257
LC12	**6.40**	**0.041**	2.36	0.307	0.36	0.549
LC13	0.39	0.822	0.93	0.630	2.54	0.111
LC14	0.91	0.634	0.05	0.978	3.58	0.058
LC17	**24.08**	**<0.001**	2.18	0.335	0.64	0.423
LC19	**14.11**	**0.001**	**7.58**	**0.023**	3.79	0.052
LC20	3.23	0.199	0.08	0.960	**5.16**	**0.023**

4.6 Figures

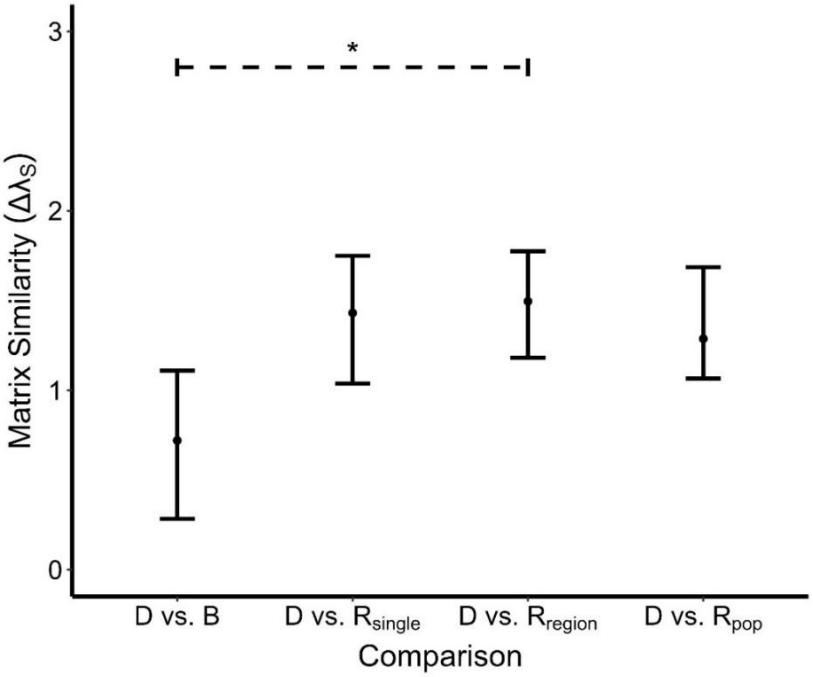

Figure 4.1 Results of Krzanowski comparisons of the major subspaces between among-population phenotypic divergence (**D**) and three variance-covariance matrices characterizing variation in linear sexual selection **B**, or the predicted responses to this selection assuming a single, overall **G** matrix ($\mathbf{R}_{single}$), region-specific **G** matrices, and population-specific **G** matrices ($\mathbf{R}_{pop}$). The metric of matrix similarity ($\Delta\lambda_S$) is a bounded measure between 0 and 3 (number of eigenvectors from original matrices used to construct subspaces) that reflects shared variance in trait space. Error bars are the upper and lower 95% quantiles that account for uncertainty in the estimation of all parameters. Asterisk indicates significance (non-overlapping 95% quantiles).

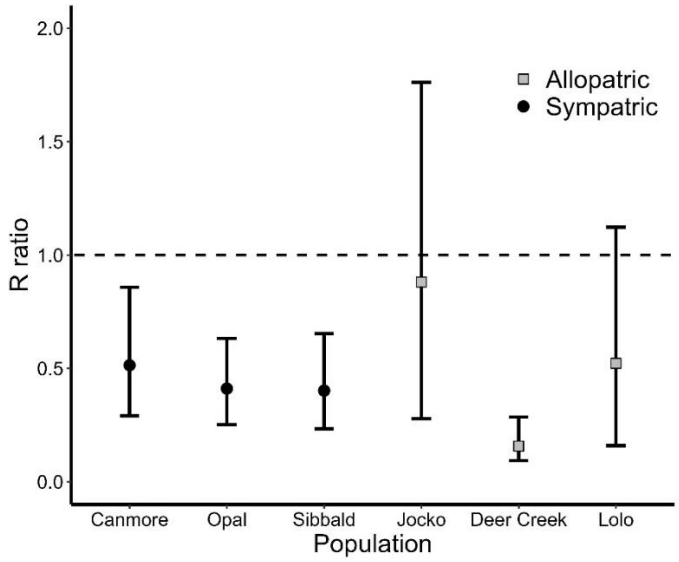

Figure 4.2 Population-specific response ratios (R) of the predicted rate of adaptation under the observed region-specific **G** matrices and the rate under region-specific **G** matrices where all genetic covariances are set to zero. A value of one would indicate no effect of genetic covariances, while values less than one indicate the rate of adaptation will be constrained, and values greater than one indicate that **G** is concentrating genetic variance in the direction of selection and adaption will be facilitated. Error bars are the 95% quartiles that account for the uncertainty in the estimation of **G** and selection.

Chapter 5: General Conclusion

To better understand the past and future of biological variation on our planet, our models of evolution rely on testing hypotheses about mechanisms underlying actual divergence in nature. Here, I have applied experimental evolution and quantitative genetic techniques to help understand a naturally occurring case of reproductive character displacement in the mycophagous fly species *Drosophila subquinaria*. In chapter 2 of this thesis, replicate populations of allopatric *D. subquinaria* were brought into experimental sympatry with *D. recens* during mating and I tracked the evolution of sexual isolation and male cuticular hydrocarbon profiles. Heterospecific mating rates decreased over the course of the experiment and male CHC profiles evolved in the kill-the-hybrids the treatment. This result strongly implies that the existing reproductive character displacement in wild *D. subquinaria* populations was an evolutionary response to selection arising from secondary contact with *D. recens*. In chapter 3 I assessed the divergence between allopatric and sympatric *D. subquinaria* in their pattern of CHC genetic (co)variances, as well as the alignment between the major axes of genetic variance and among-population phenotypic divergence. I found that the orientation of $\mathbf{g}_{max}$ has changed between regions, coming into greater alignment in sympatry with the observed phenotypic divergence. In chapter 4 I estimated sexual selection on male CHCs in each region and assessed the potential for past or future constraints on adaptive divergence in this system. I found evidence indicating that the pattern of genetic correlations has played a key role in shaping divergence, and that further evolutionary responses to sexual selection are likely to be constrained. Here I discuss some of the implications and limitations of this work, including unresolved questions and promising avenues for further research.

It is still an open question as to how RI can be strengthened in sympatry, and how frequently this is an important component of adaptive divergence (Coughlan and Matute 2020; Matute and Cooper 2021b). This phenomenon is typically attributed to reinforcement, where prezygotic isolation is strengthened by selection arising from reduced hybrid fitness, however other mechanisms may generate selection for enhanced RI, namely heterospecific reproductive interference (Gröning and Hochkirch 2008). Experimentally manipulating mating conditions so that there are artificial allopatric and sympatric treatments, and tracking the evolutionary responses of replicate populations is the best way to improve our understanding of the mechanisms that strengthen RI in sympatry. Although there is an extensive literature of studies that have done this (Fry 2009), there are very few that deviate from the "kill-the-hybrids" approach, which limits their relevance for understanding how RI is actually strengthened. My study reported in chapter 2 attempted to addresses this gap by including a treatment where hybrids are allowed to survive and thereby create the opportunity for natural selection to drive the evolution of enhanced RI. Enhanced RI evolved in the *hybrids live* treatment, despite the opportunity for gene flow. This supports the hypothesis that reproductive character displacement in nature has evolved under selection in sympatry, as opposed to an alternative explanation like biased extinction, where weakly isolated populations fuse or go extinct, leaving an apparent pattern of reproductive character displacement. However, it is worth noting that I was not able to identify whether selection for lower heterospecific mating rates was operating prior to and/or after the formation of hybrid zygotes; an issue that goes beyond this study, as few have attempted, let alone succeeded, at distinguishing between these sources of selection. One potential way to gain insight on this issue would

be to test the effects of likely sources of heterospecific reproductive interference (e.g. acoustic playbacks, perfuming mating arenas with heterospecific pheromones) without the physical presence of the other species. If increased reproductive isolation evolves, then it would rule out reduced hybrid fitness (i.e. reinforcement) as the only source of selection driving enhanced RI. This speaks to the value of identifying the traits that mediate sexual interactions and RI, and ultimately to understanding how variation in these impacts fitness.

It is also worth noting that male CHC profiles didn't respond to artificial sympatry in the *hybrids live* treatment, and in the *hybrids killed* treatment, they responded differently than the existing pattern of CHC divergence. This suggests the selection regime on CHCs in the lab is different than in nature. We know that CHCs are targets of natural selection due to their importance in moisture regulation (Kwan and Rundle 2009), and its possible that CHC divergence in real populations is shaped by a combination of sexual selection and ecological constraints. In *Drosophila serrata* populations, which exhibit a similar pattern of reproductive character displacement in sympatry, the presence of the heterospecific species *Drosophila birchii* causes the evolution of CHC profiles away from their sexual selection optimum in the absence of *D. birchii* (Higgie and Blows 2008). This implies there is a fitness cost to the reproductive character displacement of CHC traits. A fitness cost to divergent CHC traits, such as reduced desiccation tolerance or a reduction in sexual fitness, may also exist for sympatric *D. subquinaria*, but it is an idea that has not yet been explored. A potentially interesting experiment could be to repeat the study reported in chapter 2 but include low and high humidity treatments to vary the strength of selection via desiccation tolerance on CHCs. One could also bring

sympatric *D. subquinaria* into the lab and assess whether heterospecific mate discrimination and CHC profiles evolve to resemble allopatric populations in the absence of *D. recens*.

Turning to the results of chapters 3 and 4, on the surface these appear to paint a conflicting picture: the **G**-matrix for CHCs has evolved so that it better aligns with the direction of phenotypic divergence in sympatry, which could imply that selection has shaped **G** and thus divergence in CHC traits has not been constrained by the underlying pattern of genetic (co)variances. However, as I found in chapter 4, sexual selection on these traits does not explain the pattern of phenotypic divergence very well and, in fact, selection is not well aligned with divergence or the major axes of genetic variance. Taken together, these results point to a scenario where **G** has evolved in response to a source of selection that I have not estimated here (or estimated accurately), one that has played a significant role in structuring divergence between regions. Alternatively, **G** could have diverged through founder effects in sympatry, involving a particular subset of alleles that differed in pleiotropy, and subsequent drift along $\mathbf{g}_{max}$ has led to the observed phenotypic divergence. However, as I found in chapter 2, sympatry with *D. recens* induces the evolution of reproductive isolation and divergence in male CHC traits, making neutral processes alone an unlikely explanation. It is also possible that linkage disequilibrium is a key component of genetic covariances in the wild, and that this linkage partially broke down in the lab during the three generations of random mating used in constructing the outbred stocks for each population and when conducting the breeding design. Still, this would not result in the pattern of divergence in **G** between regions we find, nor the importance of **G** in explaining the observed phenotypic divergence.

A key assumption in chapters 3 and 4 is that the allopatric populations have genetic (co)variances and sexual selection that is reflective of the ancestral populations before secondary contact, which is not necessarily the case. The **G**-matrix and sexual selection gradients may have shifted over time with changes in local biotic and abiotic conditions, for example. Attempts could be made to reconstruct the ancestral the **G**-matrix using phylogenetic techniques to assess whether the allopatric **G**-matrix is representative (see McGlothlin et al. 2018). This would require extensive estimation of population **G**-matrices across both the allopatric and sympatric regions, and potentially of closely related species like *D. recens* and *D. transversa*. However, reliable phylogenies are challenging to construct in this species complex due to incomplete lineage sorting and ongoing gene flow between *D. subquinaria* and *D. recens* in sympatry (Ginsberg et al. 2019).

Another interesting opportunity for future work would be an experiment that brought allopatric *D. subquinaria* populations into the lab, exposed them to artificial sympatry with *D. recens* as in chapter 2, and tracked the evolution of the **G**-matrix for CHC traits (see Careau et al. 2015). This could provide useful insights into how rapidly **G** may have diverged, although, as we know, CHCs are likely under different natural selection pressures in the lab. Chenoweth et al. (2010) similarly found that selection alone did not explain divergence in CHC traits among populations of *D. serrata*. There and in this study, it may be the case that sexual selection estimated in the lab does not reflect the pattern in the wild. This could be because the lab environment differs in many ways from natural conditions, including a constant, relatively high humidity, ample resource availability, and altered social context of mating. *D. recens* is not present during

the mating trials in my work, an issue that could be addressed in a future study that estimates sexual selection on CHC's while manipulating the opportunity for heterospecific reproductive interference (i.e. presence/absence of *D. recens*). Additionally, plasticity in male CHCs in response to the presence of *D. recens* could be assessed to better understand whether this could be a crucial factor in mating interactions. Estimating natural (viability) selection on CHCs would be challenging but may ultimately be a crucial component of selection on these traits. CHCs are integral to regulating moisture loss through the insect's cuticle and are likely under selection pressures in the relatively arid environments that this species is often found in. The method of assaying CHCs employed here, gas chromatography-mass spectrometry, is lethal to the individual fly, precluding the estimation of any form of viability selection. However, there are techniques which could potentially allow non-lethal sampling of CHCs, specifically direct analysis in real-time (DART) mass spectrometry (Yew et al. 2008). This could hypothetically allow us to assess CHCs for a group of flies, then expose them to a desiccation treatment and track survival. More preliminary work would need to be done to determine whether this is a reliable method for *D. subquinaria*.

Lastly, more work could be done on characterizing CHC divergence and the evolution of RI in the *D. subquinaria* species complex more generally. Within the allopatric region for this species, there is CHC divergence between populations that are on either side of the Coast Mountains (Bewick and Dyer 2014), and there is a clinal pattern in the strength of RI, with increasing strength as populations near the sympatry zone (Dyer et al. 2018). In my studies, allopatric populations were collected from the inland portion of the allopatric region, as it has an abiotic environment that more closely

resembles the sympatric region which should reduce variation in natural (viability) selection. Repeating these studies with coastal populations of allopatric *D. subquinaria* could provide insight on whether there are meaningful differences in how populations across the allopatric range respond to artificial sympatry and in the pattern of genetic (co)variances for CHCs. Further, there is another closely related species, *D. transversa*, whose range extends across northern Eurasia to the Kamchatka Peninsula (Sidorenko 2009; Ginsberg et al. 2019). It is speculated that there may be a sympatry zone between *D. transversa* and *D. subquinaria* in Alaska or eastern Russia, although its currently unknown (Humphreys et al. 2016). Although RI has been found to be low between populations of these species that are allopatric to each other, this may not be the case if a sympatric zone could be found. It would be interesting to assess whether they demonstrate the same pattern of reproductive character displacement we find between *D. subquinaria* and *D. recens* and could provide a unique opportunity to test the consistency of my findings here with an analogous system.

Quantifying how evolutionary trajectories are shaped by the underlying pattern of genetic correlations has important implications for evolutionary geneticists seeking to understand fundamental processes of adaptive evolution and speciation, and for animal and plant breeders seeking to improve economically important traits. My thesis has attempted to address key gaps in our knowledge of how reproductive isolation and genetic correlations evolve, and how they influence multivariate trait evolution in nature. Studies incorporating estimates of multivariate phenotypes, their genetic covariances, and selection gradients are exceedingly rare, and ones with sufficient replication to incorporate non-linear selection are essentially absent. This is perhaps unsurprising given

the considerable effort required to get robust estimates of these parameters. Opportunities to collaborate and pool resources among researchers in order to accomplish more ambitious goals in this field will be vital to expanding our understanding of how genetic architecture and selection interact to produce the vast array of biological variation on our planet.

References

Agrawal, A. F., E. D. Brodie, and L. H. Rieseberg. 2001. Possible consequences of genes of major effect: transient changes in the G-matrix. Genetica 112–113:33–43.

Agrawal, A. F., and J. R. Stinchcombe. 2009. How much do genetic covariances alter the rate of adaptation? Proceedings of the Royal Society B: Biological Sciences 276:1183–1191. Royal Society.

Aguirre, J. D., E. Hine, K. McGuigan, and M. W. Blows. 2014. Comparing G: multivariate analysis of genetic variation in multiple populations. Heredity 112:21–29.

Aitchison, J. 1982. The statistical analysis of compositional data. Journal of the Royal Statistical Society. Series B (Methodological) 44:139–177. [Royal Statistical Society, Wiley].

Arnold, S. J. 2023. Evolutionary Quantitative Genetics. Oxford University Press.

Arnold, S. J., R. Bürger, P. A. Hohenlohe, B. C. Ajie, and A. G. Jones. 2008. Understanding the evolution and stability of the G-matrix. Evolution 62:2451–2461.

Arnold, S. J., M. E. Pfrender, and A. G. Jones. 2001. The adaptive landscape as a conceptual bridge between micro- and macroevolution. Genetica 8:9–32.

Barrett, R. D. H., S. M. Rogers, and D. Schluter. 2008. Natural selection on a major armor gene in threespine stickleback. Science 322:255–257.

Barton, N. H., and M. Turelli. 1987. Adaptive landscapes, genetic distance and the evolution of quantitative characters. Genet Res 49:157–173.

Bates, D., M. Mächler, B. Bolker, and S. Walker. 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67:1–48.

Bégin, M., and D. A. Roff. 2004. From micro- to macroevolution through quantitative genetics: Positive evidence from field crickets. Evolution 58:2287–2304.

Berson, J. D., M. Zuk, and L. W. Simmons. 2019. Natural and sexual selection on cuticular hydrocarbons: A quantitative genetic analysis. Proceedings of the Royal Society B: Biological Sciences 286.

Bewick, E. R., and K. A. Dyer. 2014. Reinforcement shapes clines in female mate discrimination in *Drosophila subquinaria*. Evolution 68:3082–3094.

Blomquist, G. J., and A.-G. Bagnères. 2010. Insect Hydrocarbons: Biology, Biochemistry, and Chemical Ecology. Cambridge University Press.

Blows, M., S. F. Chenoweth, and E. Hine. 2004. Orientation of the Genetic Variance-Covariance Matrix and the Fitness Surface for Multiple Male Sexually Selected Traits. The American Naturalist, doi: 10.1086/381941. The University of Chicago Press.

Blows, M. W. 2007. A tale of two matrices: Multivariate approaches in evolutionary biology. Journal of Evolutionary Biology 20:1–8.

Blows, M. W., and M. Higgie. 2003. Genetic Constraints on the Evolution of Mate Recognition under Natural Selection. The American Naturalist 161:240–253. [The University of Chicago Press, The American Society of Naturalists].

Blows, M. W., and K. McGuigan. 2015. The distribution of genetic variance across phenotypic space and the response to selection. Molecular Ecology 24:2056–2072.

Blows, M., and B. Walsh. 2009. Spherical Cows Grazing in Flatland: Constraints to Selection and Adaptation. Pp. 83–101 *in* J. van der Werf, H.-U. Graser, R. Frankham, and C. Gondro, eds. Adaptation and Fitness in Animal Populations: Evolutionary and Breeding Perspectives on Genetic Resource Management. Springer Netherlands, Dordrecht.

Bonduriansky, R., M. A. Mallet, D. Arbuthnott, V. Pawlowsky-Glahn, J. J. Egozcue, and H. D. Rundle. 2015. Differential effects of genetic vs. environmental quality in *Drosophila melanogaster* suggest multiple forms of condition dependence. Ecology Letters 18:317–326.

Butler, D. G., B. R. Cullis, A. R. Gilmour, B. J. Gogel, and R. Thompson. 2023. ASReml-R Reference Manual Version 4.2. VSN International Ltd., Hemel Hempstead, HP2 4TP, UK.

Careau, V., M. E. Wolak, P. A. Carter, and T. Garland. 2015. Evolution of the additive genetic variance–covariance matrix under continuous directional selection on a complex behavioural phenotype. Proceedings of the Royal Society B: Biological Sciences 282:20151119. Royal Society.

Chantepie, S., A. Charmantier, B. Delahaie, F. Adriaensen, E. Matthysen, M. E. Visser, E. Álvarez, E. Barba, M. Orell, B. Sheldon, E. Ivankina, A. Kerimov, S. Lavergne, and C. Teplitsky. 2024. Divergence in evolutionary potential of life history traits among wild populations is predicted by differences in climatic conditions. Evolution Letters.

Chenoweth, S. F., and M. W. Blows. 2005. Contrasting mutual sexual selection on homologous signal traits in *Drosophila serrata*. American Naturalist 165:281–289.

Chenoweth, S. F., H. D. Rundle, and M. W. Blows. 2010. The Contribution of Selection and Genetic Constraints to Phenotypic Divergence. The American Naturalist 175:186–196.

Conner, J. K., K. Karoly, C. Stewart, V. A. Koelling, H. F. Sahli, and F. H. Shaw. 2011. Rapid Independent Trait Evolution despite a Strong Pleiotropic Genetic Correlation. The American Naturalist 178:429–441. The University of Chicago Press.

Cooley, J. R., D. C. Marshall, K. B. R. Hill, and C. Simon. 2006. Reconstructing asymmetrical reproductive character displacement in a periodical cicada contact zone. Journal of Evolutionary Biology 19:855–868.

Cortot, J., J.-P. Farine, M. Cobb, C. Everaerts, and J.-F. Ferveur. 2022. Factors affecting the biosynthesis and emission of a *Drosophila* pheromone. Journal of Experimental Biology 225:jeb244422.

Coughlan, J. M., and D. R. Matute. 2020. The importance of intrinsic postzygotic barriers throughout the speciation process. Phil. Trans. R. Soc. B 375.

Coyne, J. A., and H. A. Orr. 1989. Patterns of speciation in *Drosophila*. Evolution 43:362–381.

Coyne, J. A., and H. A. Orr. 1997. "Patterns of speciation in *Drosophila*" Revisited. Evolution 50:8.

Coyne, J. A., and H. A. Orr. 2004. Speciation. Sinauer Associates, Inc.

Curtis, S., J. L. Sztepanacz, B. E. White, K. A. Dyer, H. D. Rundle, and P. Mayer. 2013. Epicuticular compounds of *Drosophila subquinaria* and *D. recens*: identification, quantification, and their role in female mate choice. Journal of Chemical Ecology 39:579–590.

Delahaie, B., A. Charmantier, S. Chantepie, D. Garant, M. Porlier, and C. Teplitsky. 2017. Conserved G-matrices of morphological and life-history traits among continental and island blue tit populations. Heredity 119:76–87. Nature Publishing Group.

Delcourt, M., M. W. Blows, J. D. Aguirre, and H. D. Rundle. 2012. Evolutionary optimum for male sexual traits characterized using the multivariate Robertson-Price Identity. Proc Natl Acad Sci U S A 109:10414–10419.

Dodd, D. M. B. 1989. Reproductive isolation as a consequence of adaptive divergence in *Drosophila pseudoobscura*. Evolution 43:1308–1311. [Society for the Study of Evolution, Wiley].

Doroszuk, A., M. W. Wojewodzic, G. Gort, and J. E. Kammenga. 2008. Rapid Divergence of Genetic Variance-Covariance Matrix within a Natural Population. The American Naturalist 171:291–304. The University of Chicago Press.

Duputié, A., F. Massol, I. Chuine, M. Kirkpatrick, and O. Ronce. 2012. How do genetic correlations affect species range shifts in a changing environment? Ecology Letters 15:251–259.

Dyer, K. A., E. R. Bewick, B. E. White, M. J. Bray, and D. P. Humphreys. 2018. Fine-scale geographic patterns of gene flow and reproductive character displacement in *Drosophila subquinaria* and *Drosophila recens*. Molecular Ecology 27:3655–3670.

Dyer, K. A., B. E. White, J. L. Sztepanacz, E. R. Bewick, and H. D. Rundle. 2013. Reproductive character displacement of epicuticular compounds and their contribution to mate choice in *Drosophila subquinaria* and *Drosophila recens*. Evolution 68:1163–1175.

Edmands, S. 2002. Does parental divergence predict reproductive compatibility? Trends in Ecology & Evolution 17:520–527.

Eroukhmanoff, F. 2009. Just How Much is the G-matrix Actually Constraining Adaptation? Evol Biol 36:323–326.

Eroukhmanoff, F., and E. I. Svensson. 2011. Evolution and stability of the G-matrix during the colonization of a novel environment. Journal of Evolutionary Biology 24:1363–1373.

Estes, S., and S. J. Arnold. 2007. Resolving the paradox of stasis: models with stabilizing selection explain evolutionary divergence on all timescales. Am Nat 169:227–244.

Ferveur, J.-F. 2005. Cuticular hydrocarbons: their evolution and roles in *Drosophila* pheromonal communication. Behav Genet 35:279–295.

Fox, J., S. Weisberg, B. Price, D. Adler, D. Bates, G. Baud-Bovy, and B. Bolker. 2019. car: Companion to Applied Regression. R package version 3.0-2. Website: https://cran. r-proje ct. org/web/packa ges/car/index. html [accessed 01 June 2019].

Fry, J. D. 2009. Laboratory Experiments on Speciation. P. *in* Experimental Evolution. University of California Press.

Funk, D. J., P. Nosil, and W. J. Etges. 2006. Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. Proc Natl Acad Sci U S A 103:3209–3213.

Gerhardt, H. C. 1994. Reproductive character displacement of female mate choice in the grey treefrog, *Hyla chrysoscelis*. Animal Behaviour 47:959–969. Elsevier Science, Netherlands.

Gershman, S. N., E. Toumishey, and H. D. Rundle. 2014. Time flies: Time of day and social environment affect cuticular hydrocarbon sexual displays in *Drosophila serrata*. Proc Biol Sci 281:20140821.

Gibbs, A. G., A. K. Chippindale, and M. R. Rose. 1997. Physiological mechanisms of evolved desiccation resistance in *Drosophila melanogaster*. The Journal of Experimental Biology 200:89–100.

Giglio, E. M., and K. A. Dyer. 2013. Divergence of premating behaviors in the closely related species *Drosophila subquinaria* and *D. recens*. Ecology and Evolution 3:365–374.

Ginsberg, P. S., D. P. Humphreys, and K. A. Dyer. 2019. Ongoing hybridization obscures phylogenetic relationships in the *Drosophila subquinaria* species complex. J Evol Biol 32:1093–1105.

Gould, S. J. 1966. Allometry and Size in Ontogeny and Phylogeny. Biological Reviews 41:587–638.

Gould, S. J. 2002. The Structure of Evolutionary Theory. Harvard University Press.

Gourbière, S., and J. Mallet. 2010. Are species real? The shape of the species boundary with exponential failure, reinforcement, and the "missing snowball." Evolution 64:1–24.

Gröning, J., and A. Hochkirch. 2008. Reproductive interference between animal species. The Quarterly Review of Biology 83:257–282. The University of Chicago Press.

Guillaume, F., and M. C. Whitlock. 2007. Effects of migration on the genetic covariance matrix. Evolution 61:2398–2409.

Hadfield, J. D. 2010. MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. Journal of Statistical Software 33:1–22.

Hairston, N. G., S. P. Ellner, M. A. Geber, T. Yoshida, and J. A. Fox. 2005. Rapid evolution and the convergence of ecological and evolutionary time. Ecology Letters 8:1114–1127.

Hangartner, S., C. Lasne, C. M. Sgrò, T. Connallon, and K. Monro. 2020. Genetic covariances promote climatic adaptation in Australian *Drosophila**. Evolution 74:326–337.

Hansen, T. F., and D. Houle. 2008. Measuring and comparing evolvability and constraint in multivariate characters. J Evol Biol 21:1201–1219.

Hansen, T. F., C. Pélabon, and D. Houle. 2011. Heritability is not Evolvability. Evol Biol 38:258–277.

Henry, G. A., and J. R. Stinchcombe. 2023a. G-matrix stability in clinally diverging populations of an annual weed. Evolution 77:49–62.

Henry, G. A., and J. R. Stinchcombe. 2023b. Strong selection is poorly aligned with genetic variation in *Ipomoea hederacea* : implications for divergence and constraint. Evolution 77:1712–1719.

Higgie, M., and M. W. Blows. 2008. The evolution of reproductive character displacement conflicts with how sexual selection operates within a species. Evolution 62:1192–1203.

Higgie, M., S. Chenoweth, and M. W. Blows. 2000. Natural selection and the reinforcement of mate recognition. Science 290:519–521.

Hine, E., S. F. Chenoweth, H. D. Rundle, and M. W. Blows. 2009. Characterizing the evolution of genetic variance using genetic covariance tensors. Philosophical Transactions of the Royal Society B: Biological Sciences 364:1567–1578. Royal Society.

Hine, E., K. McGuigan, and M. W. Blows. 2011. Natural selection stops the evolution of male attractiveness. Proceedings of the National Academy of Sciences of the United States of America 108:3659–3664.

Hopkins, R., and M. D. Rausher. 2012. Pollinator-mediated selection on flower color allele drives reinforcement. Science 335:1090–1092.

Hoskin, C. J., and M. Higgie. 2010. Speciation via species interactions: the divergence of mating traits within species. Ecology Letters 13:409–420.

Hoskin, C. J., M. Higgie, K. R. McDonald, and C. Moritz. 2005. Reinforcement drives rapid allopatric speciation. Nature 437:1353–1356.

Howard, D. 1993. Hybrid Zones and the Evolutionary Process. Oxford University Press.

Howard, R. W., and G. J. Blomquist. 2005. Ecological, Behavioral, and Biochemical Aspects of Insect Hydrocarbons. Annual Review of Entomology 50:371–393.

Humphreys, D. P., H. D. Rundle, and K. A. Dyer. 2016. Patterns of reproductive isolation in the *Drosophila subquinaria* complex: can reinforced premating isolation cascade to other species? Current Zoology 62:183–191.

Jaenike, J., K. A. Dyer, C. Cornish, and M. S. Minhas. 2006. Asymmetrical reinforcement and *Wolbachia* infection in *Drosophila*. PLoS Biology 4:1852–1862.

Jarvis, W. M. C., N. J. Arthur, H. D. Rundle, and K. A. Dyer. 2024. An experimental test of the evolutionary consequences of sympatry in *Drosophila subquinaria*. Evolution 78:555–565.

Jiggins, C. D., R. E. Naisbit, R. L. Coe, and J. Mallet. 2001. Reproductive isolation caused by colour pattern mimicry. Nature 411:302–305.

Jones, A. G., S. J. Arnold, and R. Burger. 2004. Evolution and stability of the G-matrix on a landscape with a moving optimum. Evolution 58:1639–1654.

Jones, A. G., S. J. Arnold, and R. Bürger. 2003. Stability of the G-Matrix in a Population Experiencing Pleiotropic Mutation, Stabilizing Selection, and Genetic Drift. Evolution 57:1747–1760.

Jones, A. G., S. J. Arnold, and R. Bürger. 2007. The mutation matrix and the evolution of evolvability. Evolution 61:727–745.

Kirkpatrick, M. 2009. Patterns of quantitative genetic variation in multiple dimensions. Genetica 136:271–284.

Krzanowski, W. J. 1979. Between-Groups Comparison of Principal Components. Journal of the American Statistical Association 74.

Kuznetsova, A., P. B. Brockhoff, and R. H. B. Christensen. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. Journal of Statistical Software 82:1–26.

Kwan, L., and H. D. Rundle. 2009. Adaptation to desiccation fails to generate pre- and postmating isolation in replicate *Drosophila melanogaster* laboratory populations. Evolution 64:710–723.

Lande, R. 1979. Quantitative Genetic Analysis of Multivariate Evolution, Applied to Brain: Body Size Allometry. Evolution 33:402–416. [Society for the Study of Evolution, Wiley].

Lande, R. 1980. The Genetic Covariance between Characters Maintained by Pleiotropic Mutations. Genetics 94:203–215.

Lande, R., and S. J. Arnold. 1983. The Measurement of Selection on Correlated Characters. Evolution 37:1210–1226. [Society for the Study of Evolution, Wiley].

Lenth, R. V. 2021. emmeans: estimated marginal means, aka least-squares means. R package version 1.6. 1.

Lewis, Z., N. Wedell, and J. Hunt. 2011. Evidence for strong intralocus sexual conflict in the Indian meal moth, *Plodia interpunctella*. Evolution 65:2085–2097.

Losos, J. B., S. J. Arnold, G. Bejerano, E. D. Brodie, D. Hibbett, H. E. Hoekstra, D. P. Mindell, A. Monteiro, C. Moritz, H. A. Orr, D. A. Petrov, S. S. Renner, R. E. Ricklefs, P. S. Soltis, and T. L. Turner. 2013. Evolutionary Biology for the 21st Century. PLoS Biol 11:e1001466.

Lynch, M., and B. Walsh. 1998. Genetics and analysis of quantitative traits. Sinauer, Sunderland, MA.

Marsaglia, G. 1972. Choosing a Point from the Surface of a Sphere. The Annals of mathematical statistics 43:645–646. Institute of Mathematical Statistics.

Marshall, J. L., M. L. Arnold, and D. J. Howard. 2002. Reinforcement: the road not taken. Trends in Ecology & Evolution 17:558–563.

Matute, D. R. 2010a. Reinforcement can overcome gene flow during speciation in *Drosophila*. Current Biology 20:2229–2233.

Matute, D. R. 2010b. Reinforcement of gametic isolation in *Drosophila*. PLOS Biology 8:e1000341. Public Library of Science.

Matute, D. R. 2013. The role of founder effects on the evolution of reproductive isolation. Journal of Evolutionary Biology 26:2299–2311.

Matute, D. R., and B. S. Cooper. 2021a. Comparative studies on speciation: 30 years since Coyne and Orr. Evolution 75:764–778.

Matute, D. R., and B. S. Cooper. 2021b. Reinforcement alone does not explain increased reproductive isolation in sympatry. doi: 10.1101/2021.05.06.442525.

McGlothlin, J. W., M. E. Kobiela, H. V. Wright, J. J. Kolbe, J. B. Losos, and E. D. Brodie. 2022. Conservation and Convergence of Genetic Architecture in the Adaptive Radiation of *Anolis* Lizards. Am Nat 200:E207–E220.

McGlothlin, J. W., M. E. Kobiela, H. V. Wright, D. L. Mahler, J. J. Kolbe, J. B. Losos, and E. D. Brodie. 2018. Adaptive radiation along a deeply conserved genetic line of least resistance in *Anolis* lizards. Evolution Letters 2:310–322.

McGuigan, K., S. F. Chenoweth, and M. W. Blows. 2005. Phenotypic Divergence along Lines of Genetic Variance. The American Naturalist 165:32–43. The University of Chicago Press.

McKinnon, J. S., S. Mori, B. K. Blackman, L. David, D. M. Kingsley, L. Jamieson, J. Chou, and D. Schluter. 2004. Evidence for ecology's role in speciation. Nature 429:294–298.

Mitchell-Olds, T. 1996. Pleiotropy Causes Long-Term Genetic Constraints on Life-History Evolution in *Brassica rapa*. Evolution 50:1849–1858. [Society for the Study of Evolution, Wiley].

Moose, S. P., J. W. Dudley, and T. R. Rocheford. 2004. Maize selection passes the century mark: a unique resource for 21st century genomics. Trends in Plant Science 9:358–364. Elsevier.

Morrissey, M. B., S. Hangartner, and K. Monro. 2019. A note on simulating null distributions for G matrix comparisons. Evolution 73:2512–2517.

Mousseau, T. A., and D. A. Roff. 1987. Natural selection and the heritability of fitness components. Heredity 59:181–197. Nature Publishing Group.

Nichols, C. D., J. Becnel, and U. B. Pandey. 2012. Methods to assay *Drosophila* behavior. J Vis Exp 3795.

Noor, M. 1995. Speciation driven by natural selection in *Drosophila*. Nature 375:674–675.

Noor, M. A. F. 1999. Reinforcement and other consequences of sympatry. Heredity 83:503–508. Nature Publishing Group.

Nosil, P. 2013. Degree of sympatry affects reinforcement in *Drosophila*. Evolution 67:868–872.

Nosil, P., C. P. Sandoval, and B. J. Crespi. 2006. The evolution of host preference in allopatric vs. parapatric populations of *Timema cristinae* walking-sticks. Journal of Evolutionary Biology 19:929–942.

Oh, K. P., and K. L. Shaw. 2022. Axes of multivariate sexual signal divergence among incipient species: Concordance with selection, genetic variation and phenotypic plasticity. Journal of Evolutionary Biology 35:109–123.

Ortiz-Barrientos, D., A. Grealy, and P. Nosil. 2009. The genetics and ecology of reinforcement. Annals of the New York Academy of Sciences 1168:156–182.

Pfennig, K. S., and D. W. Pfennig. 2009. Character displacement: ecological and reproductive responses to a common evolutionary problem. The Quarterly Review of Biology 84:253–276.

Phillips, P. C., M. C. Whitlock, and K. Fowler. 2001. Inbreeding changes the shape of the genetic covariance matrix in *Drosophila melanogaster*. Genetics 158:1137–1145.

Powell, R. L., and H. D. Norman. 2006. Major Advances in Genetic Evaluation Techniques. Journal of Dairy Science 89:1337–1348.

Puentes, A., G. Granath, and J. Ågren. 2016. Similarity in G matrix structure among natural populations of *Arabidopsis lyrata*. Evolution 70:2370–2386.

R Development Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Reddiex, A. J., and S. F. Chenoweth. 2021. Integrating genomics and multivariate evolutionary quantitative genetics: a case study of constraints on sexual selection in *Drosophila serrata*. Proceedings of the Royal Society B: Biological Sciences 288:20211785. Royal Society.

Reeve, J. P. 2000. Predicting long-term response to selection. Genetics Research 75:83–94. Cambridge University Press.

Revell, L. J. 2007. Testing the genetic constraint hypothesis in a phylogenetic context: A simulation study. Evolution 61:2720–2727.

Rice, W. R., and E. Hostert. 1993. Lab experiments on speciation: What have we learning in 40 years? Evolution 47:1637–1653.

Rice, W. R., and G. W. Salt. 1990. The evolution of reproductive isolation as a correlated character under sympatric conditions: experimental evidence. Evolution 44:1140–1152.

Richardson, J. L., M. C. Urban, D. I. Bolnick, and D. K. Skelly. 2014. Microgeographic adaptation and the spatial scale of evolution. Trends in Ecology & Evolution 29:165–176. Elsevier.

Robinson, R. 2006. Infection Status Drives Interspecies Mating Choices in Fruit Fly Females. PLoS Biol 4(10): e345. https://doi.org/10.1371/journal.pbio.0040345

Roff, D. 2000. The evolution of the G matrix: selection or drift? Heredity 84:135–142.

Rundle, H. D., S. F. Chenoweth, and M. W. Blows. 2008. Comparing Complex Fitness Surfaces: Among-Population Variation in Mutual Sexual Selection in Drosophila serrata. The American Naturalist 171:443–454. The University of Chicago Press.

Rundle, H. D., and K. A. Dyer. 2015. Reproductive character displacement of female mate preferences for male cuticular hydrocarbons in *Drosophila subquinaria*. Evolution 69:2625–2637.

Rundle, H. D., A. Ø. Mooers, and M. C. Whitlock. 1998. Single founder-flush events and the evolution of reproductive isolation. Evolution 52:1850–1855.

Rundle, H. D., and D. Schluter. 1998. Reinforcement of stickleback mate preferences: sympatry breeds contempt. Evolution 52:200–208.

Rundle, H., L. Nagel, J. W. Boughman, and D. Schluter. 2000. Natural Selection and Parallel Speciation in Sympatric Sticklebacks. Science 287:306–308.

Rundle, H., and P. Nosil. 2005. Ecological speciation. Ecology Letters 8:336–352.

Saetre, G.P. 1997. A sexually selected character displacement in flycatchers reinforces premating isolation. Nature 387:589–582.

Sætre, G.-P., T. Moum, S. Bureš, M. Král, M. Adamjan, and J. Moreno. 1997. A sexually selected character displacement in flycatchers reinforces premating isolation. Nature 387:589–592. Nature Publishing Group.

Samuk, K., G. L. Owens, K. E. Delmore, S. E. Miller, D. J. Rennison, and D. Schluter. 2017. Gene flow and selection interact to promote adaptive divergence in regions of low recombination. Molecular Ecology 26:4378–4390.

Schluter, D. 1996. Adaptive Radiation Along Genetic Lines of Least Resistance. Evolution 50:1766–1774.

Schluter, D. 2001. Ecology and the origin of species. Trends in Ecology & Evolution 16:372–380.

Schluter, D. 2009. Evidence for Ecological Speciation and Its Alternative. Science 323:737–741. American Association for the Advancement of Science.

Schluter, D. 2000. The ecology of adaptive radiation. Oxford University Press, Oxford.

Shaw, F. H., R. G. Shaw, G. S. Wilkinson, and M. Turelli. 1995. Changes in Genetic Variances and Covariances: G Whiz! Evolution 49:1260–1267.

Shoemaker, D. D., V. Katju, and J. Jaenike. 1999. *Wolbachia* and the evolution of reproductive isolation between *Drosophila recens* and *Drosophila subquinaria*. Evolution 53:1157–1164.

Sidorenko, V. S. 2009. A review of species of the *Drosophila quinaria* Loew species-group (Diptera Drosophilidae) of the Asian part of Russia and adjacent countries. Entmol. Rev. 89:85–90.

Simonsen, A. K., and J. R. Stinchcombe. 2010. Quantifying Evolutionary Genetic Constraints in the Ivyleaf Morning Glory, Ipomoea hederacea. International Journal of Plant Sciences 171:972–986. The University of Chicago Press.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. Van Der Linde. 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64:583–639.

Steppan, S. J., P. C. Phillips, and D. Houle. 2002. Comparative quantitative genetics: Evolution of the G matrix. Trends in Ecology and Evolution 17:320–327.

Stinziano, J. R., R. J. Sové, H. D. Rundle, and B. J. Sinclair. 2015. Rapid desiccation hardening changes the cuticular hydrocarbon profile of *Drosophila melanogaster*. Comparative Biochemistry and Physiology - Part A: Molecular and Integrative Physiology 180:38–42.

Sztepanacz, J. L., and M. W. Blows. 2017. Artificial Selection to Increase the Phenotypic Variance in gmax Fails. Am Nat 190:707–723.

Sztepanacz, J. L., and H. D. Rundle. 2012. Reduced genetic variance among high fitness individuals: Inferring stabilizing selection on male sexual displays in *Drosophila serrata*. Evolution 66:3101–3110.

Templeton, A. R. 1981. Mechanisms of speciation-a population genetic approach. Annual Review of Ecology and Systematics 12:23–48. Annual Reviews.

Templeton, A. R. 2008. The reality and importance of founder speciation in evolution. BioEssays 30:470–479.

Teplitsky, C., M. Tarka, A. P. Møller, S. Nakagawa, J. Balbontín, T. A. Burke, C. Doutrelant, A. Gregoire, B. Hansson, D. Hasselquist, L. Gustafsson, F. de Lope, A. Marzal, J. A. Mills, N. T. Wheelwright, J. W. Yarrall, and A. Charmantier. 2014. Assessing Multivariate Constraints to Evolution across Ten Long-Term Avian Studies. PLOS ONE 9:e90444. Public Library of Science.

Turelli, M. 1988. Phenotypic evolution, constant covariances, and the maintenance of additive variance. Evolution 42:1342–1347.

Uesugi, A., T. Connallon, A. Kessler, and K. Monro. 2017. Relaxation of herbivore-mediated selection drives the evolution of genetic covariances between plant competitive and defense traits. Evolution 71:1700–1709.

van der Niet, T., S. D. Johnson, and H. P. Linder. 2006. Macroevolutionary data suggest a role for reinforcement in pollination system shifts. Evolution 60:1596–1601.

Walsh, B., and M. W. Blows. 2009. Abundant Genetic Variation + Strong Selection = Multivariate Genetic Constraints: A Geometric View of Adaptation. Annual Review of Ecology, Evolution, and Systematics 40:41–59.

Walter, G. M., J. D. Aguirre, M. W. Blows, and D. Ortiz-Barrientos. 2018. Evolution of Genetic Variance during Adaptive Radiation. The American Naturalist 191:E108–E128.

Welch, J. J. 2017. What's wrong with evolutionary biology? Biol Philos 32:263–279.

White, N. J., R. R. Snook, and I. Eyres. 2020. The past and future of experimental speciation. Trends in Ecology & Evolution 35:10–21.

Yew, J. Y., R. B. Cody, and E. A. Kravitz. 2008. Cuticular hydrocarbon analysis of an awake behaving fly using direct analysis in real-time time-of-flight mass spectrometry. Proc Natl Acad Sci U S A 105:7135–7140.

Yukilevich, R. 2012. Asymmetrical patterns of speciation uniquely support reinforcement in *Drosophila*. Evolution 66:1430–1446.

Supplementary Information: Chapter 2

Table S2.1 The eigenvectors of loadings and percent variance explained for the first three principal components (PCs) of male among-population variation in CLR-transformed relative concentrations of cuticular hydrocarbons (CHCs) sampled at generation 12. CHCs identifications were made via qualitative alignment of CHC chromatograms with those in Curtis et al. (2013).

	CHC identification	**PC1**	**PC2**	**PC3**
% Variance		60.9	17.6	7.3
Peak 1	11-cis-Vaccenyl acetate (cVa)	0.192596	0.084863	0.049663
Peak 2	2-methyl octacosane	-0.21107	-0.1548	0.266666
Peak 5	2-methyl triacontane	0.140598	0.178367	0.293201
Peak 7	5-hentriacontene	-0.09616	-0.15714	0.111012
Peak 11	n-tritriacontene	-0.24046	0.046649	-0.34542
Peak 12	(Z,Z)-5-13-tritriacontadiene	-0.25188	-0.35587	0.091547
Peak 13	(Z,Z)-5-11-tritriacontadiene	-0.36337	0.174723	-0.25666
Peak 14	(Z,Z)-n-n-tritriacontadiene	-0.07332	0.165756	0.327757
Peak 15	Tri-acylglyceride	0.276561	-0.19847	0.020636
Peak 16	Tri-acylglyceride	0.23145	-0.28185	-0.12444
Peak 17	n-methyl tetatriacontane	-0.29936	0.07548	0.051082
Peak 18	n-pentatriacontene	0.209028	0.717324	0.138616
Peak 19	(Z,Z)-n-n-pentatriacontadiene	-0.19148	0.146945	-0.40516
Peak 20	(Z,Z)-5-13-pentatriacontadiene	-0.01722	-0.00954	0.322417
Peak 22	Tri-acylglyceride	0.285027	0.085262	-0.45783
Peak 23	Tri-acylglyceride	0.414378	-0.07848	-0.10809
Peak 24	Tri-acylglyceride	0.278184	-0.22652	-0.00142

Table S2.2 The eigenvectors of loadings and percent variance explained for the first three principal components (PCs) of female among-population variation in CLR-transformed relative concentrations of cuticular hydrocarbons (CHCs) sampled at generation 12. CHCs identifications were made via qualitative alignment of CHC chromatograms with those in Curtis et al. (2013).

	CHC identification	**PC1**	**PC2**	**PC3**
% Variance		49.9	29.6	9.7
Peak 2	2-methyl octacosane	0.026462	0.153214	-0.56172
Peak 5	2-methyl triacontane	-0.16165	-0.44118	-0.0672
Peak 7	5-hentriacontene	0.313631	-0.39827	0.030687
Peak 11	n-tritriacontene	-0.04923	0.338586	0.251785
Peak 12	(Z,Z)-5-13-tritriacontadiene	0.474747	0.017396	-0.09194
Peak 13	(Z,Z)-5-11-tritriacontadiene	0.016435	0.428744	0.448012
Peak 14	(Z,Z)-n-n-tritriacontadiene	0.012533	-0.13945	0.360101
Peak 17	n-methyl tetatriacontane	-0.15415	0.180991	-0.47278
Peak 18	n-pentatriacontene	-0.75542	-0.19529	0.026268
Peak 19	(Z,Z)-n-n-pentatriacontadiene	-0.22759	0.323872	0.045186
Peak 20	(Z,Z)-5-13-pentatriacontadiene	-0.01872	-0.35953	0.224781

Table S2.3 Eigenvectors of loadings and percent variance explained for the first two principal components (PCs) of variation in CLR-transformed cuticular hydrocarbons (CHCs) among four groups of males: allopatric *D. subquinaria*, sympatric *D. subquinaria*, allopatric *D. recens*, and sympatric *D. recens*. Data from Dyer et al. (2013).

CHC	PC1	PC2
% Variance	86.1	12.9
Peak 1	0.0834	-0.0156
Peak 2	0.0988	0.1350
Peak 5	0.1013	0.2576
Peak 7	0.4789	-0.5509
Peak 11	0.0591	-0.0069
Peak 12	0.1860	-0.1588
Peak 13	0.1839	0.2497
Peak 14	-0.3360	0.2303
Peak 15	0.1040	0.2909
Peak 16	0.0428	0.2608
Peak 17	-0.4198	-0.4837
Peak 18	-0.0944	0.0385
Peak 19	-0.3558	-0.2316
Peak 20	-0.4480	0.0599
Peak 22	0.1147	0.0396
Peak 23	0.0885	0.0362
Peak 24	0.1127	-0.1510

Table S2.4 Eigenvectors of loadings and percent variance explained for the first two principal components (PCs) of variation in CLR-transformed cuticular hydrocarbons (CHCs) among four groups of females: allopatric *D. subquinaria*, sympatric *D. subquinaria*, allopatric *D. recens*, and sympatric *D. recens*. Data from Dyer et al. (2013).

CHC	**PC1**	**PC2**
% Variance	74.4	24.6
Peak 2	0.203968	0.055317
Peak 5	0.210754	0.173375
Peak 7	0.494444	-0.50408
Peak 11	0.084187	0.094869
Peak 12	0.238234	-0.14815
Peak 13	0.161912	0.388885
Peak 14	-0.21013	0.416242
Peak 17	-0.36355	-0.48431
Peak 18	0.057814	0.218387
Peak 19	-0.35321	-0.26794
Peak 20	-0.52443	0.057416

Table S2.5 Fixed effect results from separate generalized linear mixed models testing the effects of treatment, generation, and their interaction on the proportion of experimental females mating with heterospecific *D. recens* males and, in separate mating trials, with conspecific, allopatric *D. subquinaria* males (see Methods). Bold denotes $p < 0.05$.

	Treatment			**Generation**			**Treatment × Generation**		
Males	**χ^2**	**df**	***p*-value**	**χ^2**	**df**	***p*-value**	**χ^2**	**df**	***p*-value**
heterospecific (*D. recens*)	0.91	2	0.635	**10.04**	**1**	**0.002**	**11.03**	**2**	**0.004**
conspecific (*D. subquinaria*)	**17.59**	**2**	**0.0002**	0.16	1	0.687	0.11	2	0.948

Table S2.6 Post-hoc pairwise comparison of treatment levels, separately by generation, for heterospecific mate discrimination by experimental females against *D. recens* males. Odds ratios calculated using marginal means, averaging across blocks (see Methods). Treatment effect was non-significant in generation 0 (see main text). Bold denotes $p < 0.05$.

Generation	Treatment comparison	Odds ratio (SE)	*z*-ratio (1 df)	*p*-value
6	*hybrids live*/control	**0.710 (0.094)**	**-2.60**	**0.0251**
	hybrids killed/control	**0.564 (0.074)**	**-4.35**	**< 0.0001**
	hybrids live/*hybrids killed*	1.258 (0.149)	1.94	0.1270
12	*hybrids live*/control	**0.602 (0.081)**	**-3.79**	**0.0004**
	hybrids killed/control	**0.479 (0.064)**	**-5.48**	**< 0.0001**
	hybrids live/*hybrids killed*	1.257 (0.151)	1.90	0.138

Table S2.7 Post-hoc pairwise comparisons of treatment levels for conspecific mate discrimination by experimental females against allopatric *D. subquinaria* males. Odds ratios calculated using marginal means, averaging across generations and blocks (see Methods). Bold denotes $p < 0.05$ (*p*-values are adjusted for multiple comparisons).

Treatment comparison	Odds ratio (SE)	*z*-ratio (1 df)	*p*-value
hybrids live/control	0.823 (0.105)	-1.53	0.276
hybrids killed/control	**0.516 (0.065)**	**-5.28**	**<0.0001**
hybrids killed/*hybrids live*[a]	**0.627 (0.067)**	**-4.37**	**<0.0001**

[a]*hybrids killed*/*hybrids live* is reported to align with presentation of results in main text. The inverse odds ratio (i.e., *hybrids live*/*hybrids killed*), matching those in Table S2.6, is 1.596 ± 0.171 SE (significance is unaffected).

Table S2.8 Fixed effect results from separate tests of treatment, generation, and their interaction on the first three principal components of among-population variation in CLR-transformed relative CHC concentrations in males and females. Principal components were estimated separately by sex and therefore do not correspond to the same traits in males and females (see Tables S2.1, S2.2). Bold denotes $p < 0.05$.

		Treatment			**Generation**			**Treatment × Generation**		
Sex	**Trait**	**F**	**df**	***p*-value**	**F**	**df**	***p*-value**	**F**	**df**	***p*-value**
	PC1	0.10	2, 41.2	0.903	0.03	1, 4.1	0.867	**5.00**	**2, 36.8**	**0.012**
males	PC2	**4.39**	**2, 45.9**	**0.018**	7.26	1, 4.2	0.051	0.61	2, 40.3	0.548
	PC3	1.52	2, 42.0	0.231	0.04	1, 4.0	0.851	1.06	2, 37.5	0.356
	PC1	1.19	2, 38.9	0.316	3.32	1, 4.1	0.141	1.91	2, 35.4	0.163
females	PC2	**15.30**	**2, 39.0**	**< 0.0001**	0.31	1, 4.1	0.605	0.21	2, 34.4	0.810
	PC3	1.65	2, 40.2	0.205	1.16	1, 4.3	0.339	2.08	2, 35.8	0.140

Table S2.9 Post-hoc pairwise comparisons of treatment levels for the principal components of among-population variation in CLR-transformed relative CHC concentrations in males and females (Tables S2.1, S2.2). Comparisons for PC1 in males were performed for generation 12 only given a significant treatment × generation interaction (Table S6) and a significant treatment effect in generation 12 only. Comparisons for other traits (PC2 males, PC2 females) averaged across generations given a main effect of treatment only (Table S2.6). Values are differences in marginal mean (See Methods). Bold denotes $p < 0.05$ (p-values adjusted for multiple comparisons).

Trait/generation	**Treatment comparison**	**Estimate (SE)**	***t* (df)**	***p*-value**
PC 1 males – generation 12	*hybrids live* - control	0.050 (0.209)	0.24(12)	0.969
	hybrids killed - control	**-0.756 (0.209)**	**-3.62 (11.9)**	**0.0092**
	hybrids live - hybrids killed	**0.806 (0.187)**	**4.31 (12)**	**0.0027**
PC2 males – averaged across generations	*hybrids live* - control	0.084 (0.070)	1.20 (37.2)	0.460
	hybrids killed - control	**-0.188 (0.070)**	**-2.69 (37.8)**	**0.0282**
	hybrids live - hybrids killed	**0.272 (0.060)**	**4.43 (39.1)**	**0.0002**
PC2 females – averaged across generations	*hybrids live* - control	-0.114 (0.053)	-2.12 (37.3)	0.092
	hybrids killed - control	**0.254 (0.053)**	**4.77 (38.1)**	**0.0001**
	hybrids live - hybrids killed	**-0.368 (0.042)**	**-8.67 (36.2)**	**< 0.0001**

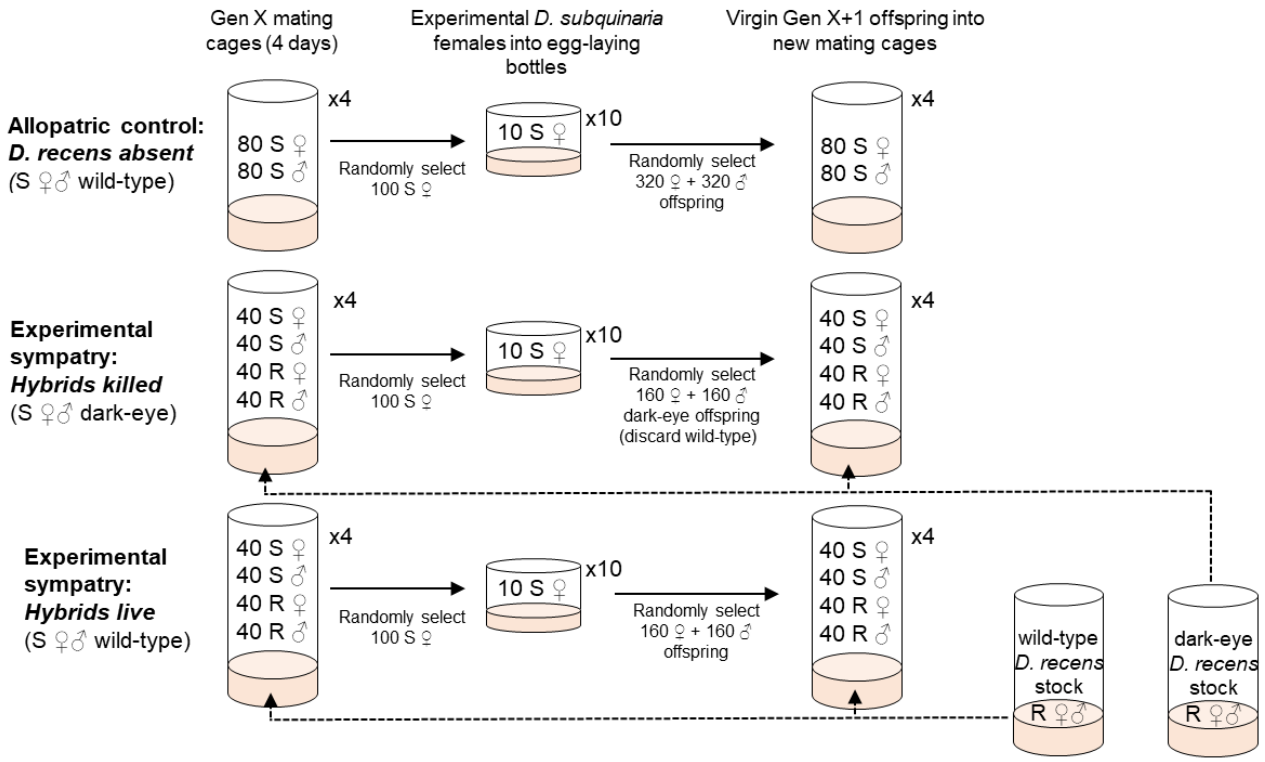

Figure S2.1 Schematic of the design of the evolution experiment, showing one replicate population for each treatment across two generations. S denotes *D. subquinaria* and R indicates *D. recens*.

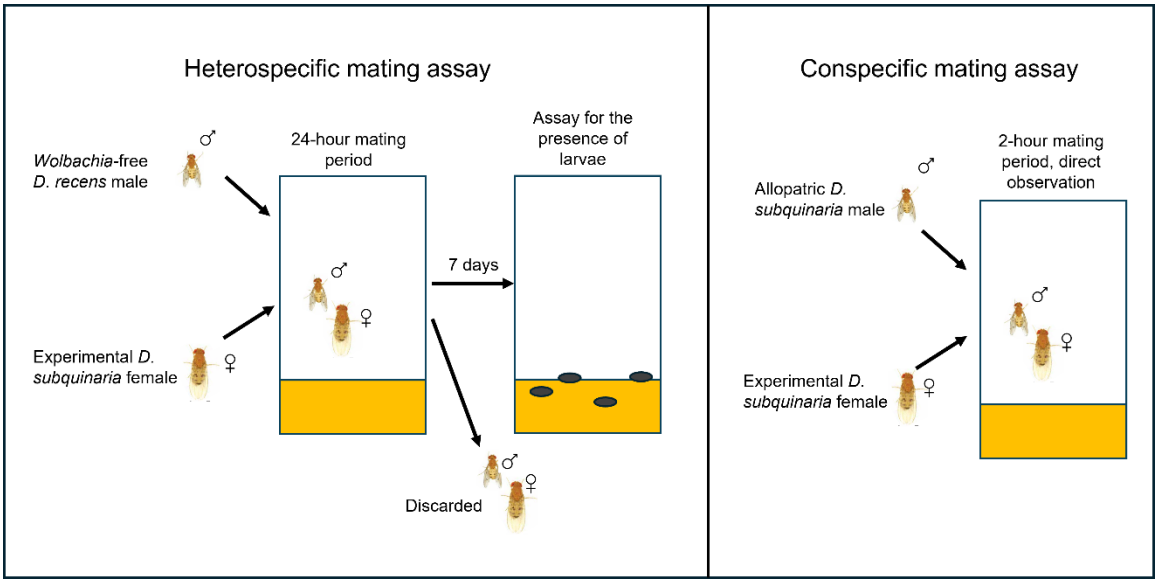

Figure S2.2 Schematic of the design of the mating trials for both heterospecific and conspecific mating assays.

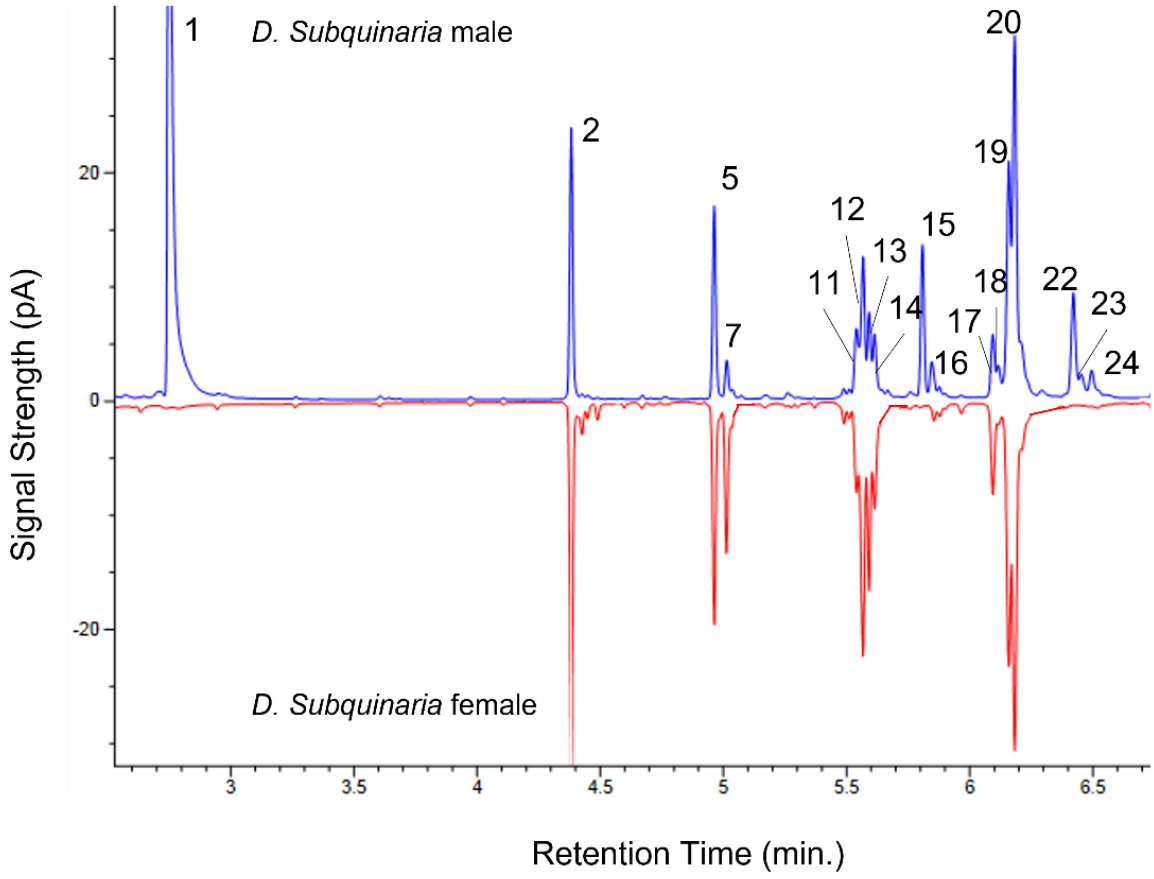

Figure S2.3 Mirrored GC chromatographic profile of a male (above; blue) and female (below; red) *D. subquinaria*. Numbers correspond to those reported in Curtis et al. (2013) and indicate those compounds that were reliably present and are identified in Table S2.1 and S2.2.

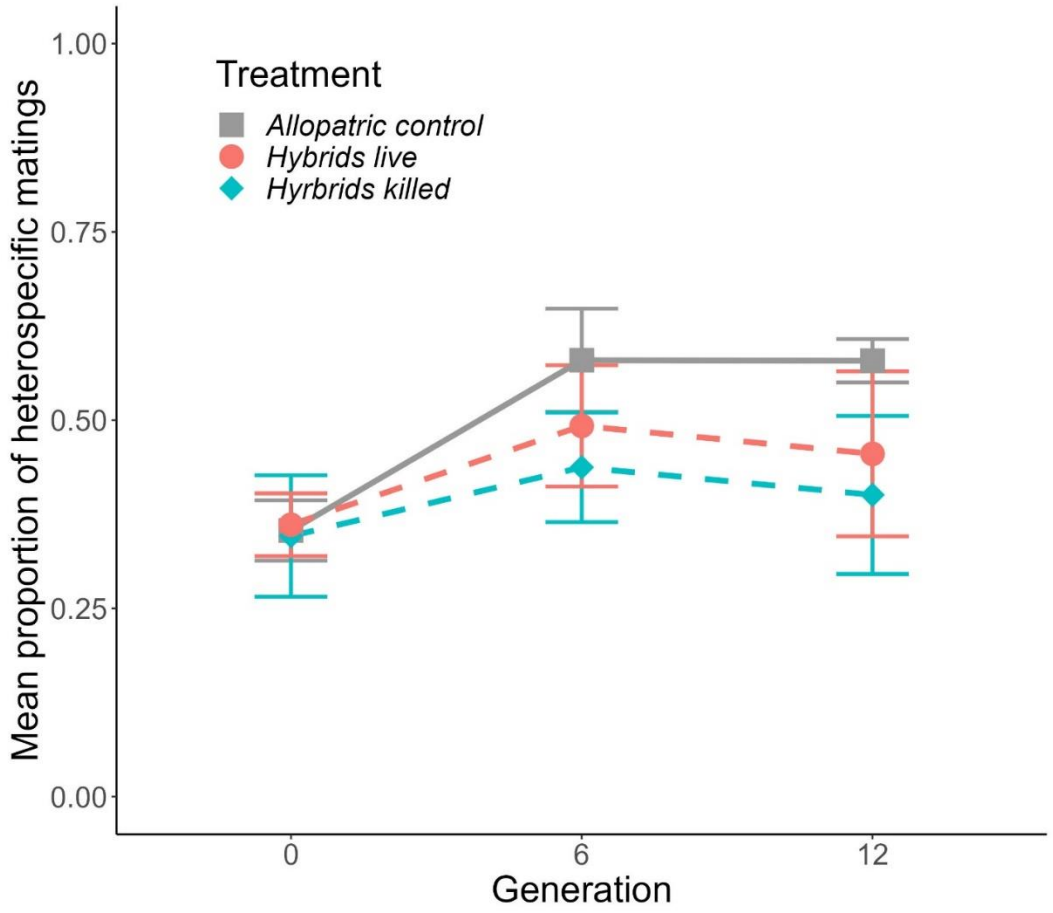

Figure S2.4 Mean heterospecific mating rates (±95% confidence intervals) of experimental *D. subquinaria* females in each experimental treatment at each sampling generation. Confidence intervals are Gaussian approximations treating experimental populations within treatment as replicates.

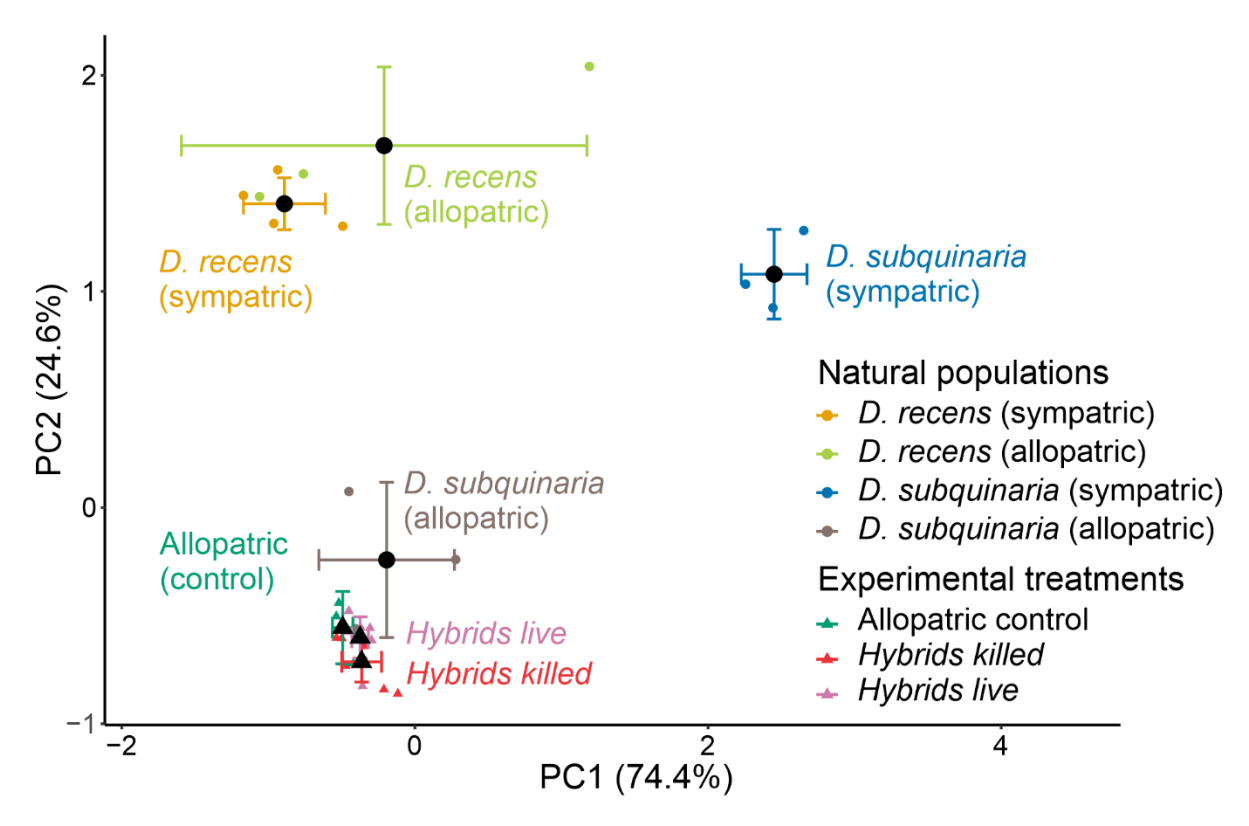

Figure S2.5 Evolutionary response of females from the experimental populations within the trait space that distinguishes sympatric and allopatric populations of *D. recens* and *D. subquinaria* in nature. Principal component analyses were performed separately by sex on data from Dyer et al. (2014) and then experimental population were projected into this space (see Methods). Colored points are means for natural (circles) and experimental (triangles) populations; black points are group means (± 95% confidence intervals), treating populations as replicates. See Fig. 2.4 for males.

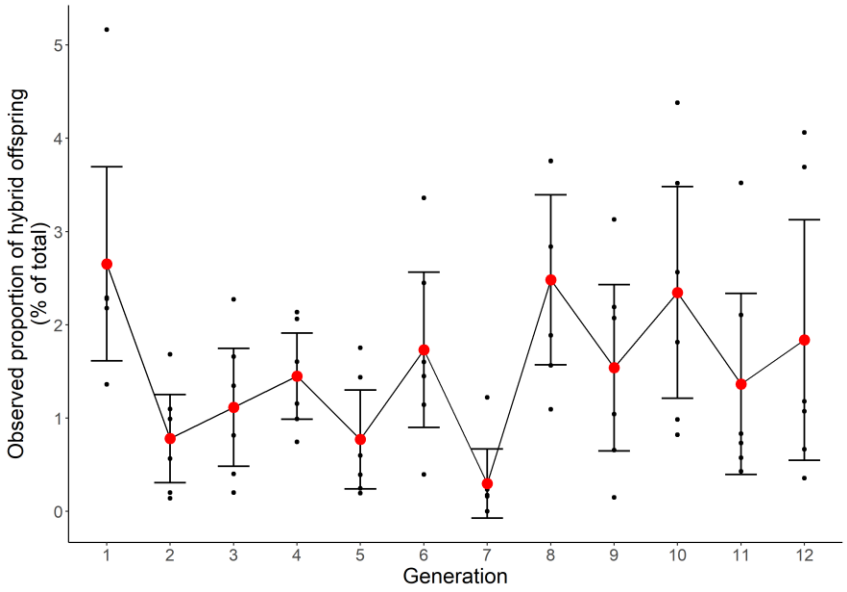

Figure S2.6 The percentage of hybrid offspring observed each generation in the *hybrids killed* experimental sympatry treatment. Hybrid offspring could be identified by their wild-type eyes, whereas pure species *D. subquinaria* had *dark* eyes. Small black points denote individual populations; population means (larger red points) are also shown for each generation with 95% confidence intervals. Hybrids initially comprised about 2.6% of offspring, and there is no indication that this decreased over the 12 generations of the experiment (average 1.5% across all generations and populations, average 1.8% in generation 12). The observed proportion of hybrids is an underestimate of the actual rate of hybrid offspring production because, when *D. subquinaria* females mate with *D. recens* males, ~90% of the embryos die due cytoplasmic incompatibility from a *Wolbachia* infection in *D. recens*. Importantly, these data provide insight into hybrid production during the sympatric maintenance routine of the evolution experiment (i.e., resulting from a 4 day interaction phase involving 40 individuals of each sex and species; see Methods) in a single treatment only (i.e. *hybrids killed*). Similar data are lacking for the *hybrids live* treatment (because hybrids are phenotypically indistinguishable from non-hybrids in this treatment), and analogous data do not exist for the allopatric control because it did not experience a 4 day sympatric interaction phase each generation. There is therefore no way to test for differences in hybrid production across treatments, and no appropriate control for testing for temporal patterns within this one treatment. The latter is important because there is substantial variation in hybridization rates among generations in the primary assay of behavioral isolation, arising from unknown causes. Normalizing relative to the control treatment accounts for this in the main assay, but that is not possible with the data in this figure.

Supplementary Information: Chapter 3

Table S3.1 Chemical identity of the cuticular compounds for the corresponding peak numbers in Fig. 3.1 Compound identifications were made via qualitative alignment of chromatographic peaks with those in Curtis et al. (2013). The 10 CHCs integrated in the current study are indicated (Integrated). Peak 5 was used as the divisor in the logcontrast transformation. cVa and the tri-acylglycerides are not formally CHCs and are likely to co-localized around the ejaculatory bulb and to derive from a different biosynthetic process than the CHCs (Curtis et al. 2013; Cortot et al. 2022).

Peak Number	Compound	Integrated
Peak 1	11-cis-Vaccenyl acetate (cVa)	No
Peak 2	2-methyl octacosane	Yes
Peak 5	2-methyl triacontane	Yes
Peak 7	5-hentriacontene	Yes
Peak 11	n-tritriacontene	Yes
Peak 12	(Z,Z)-5-13-tritriacontadiene	Yes
Peak 13	(Z,Z)-5-11-tritriacontadiene	Yes
Peak 14	(Z,Z)-n-n-tritriacontadiene	Yes
Peak 15	Tri-acylglyceride	No
Peak 16	Tri-acylglyceride	No
Peak 17	n-methyl tetatriacontane	Yes
Peak 18	n-pentatriacontene	No
Peak 19	(Z,Z)-n-n-pentatriacontadiene	Yes
Peak 20	(Z,Z)-5-13-pentatriacontadiene	Yes
Peak 22	Tri-acylglyceride	No
Peak 23	Tri-acylglyceride	No
Peak 24	Tri-acylglyceride	No

Table S3.2 Eigenvectors of the overall **G** matrix, with associated eigenvalues and their 95% HPD intervals. The first four eigenvectors had statistical support as their 95% HPD intervals did not overlap with those of the null **G** estimated from randomized data.

	$\mathbf{g}_{max}$	$\mathbf{g}_2$	$\mathbf{g}_3$	$\mathbf{g}_4$	$\mathbf{g}_5$	$\mathbf{g}_6$	$\mathbf{g}_7$	$\mathbf{g}_8$	$\mathbf{g}_9$
eigenvalue	0.213	0.047	0.035	0.016	0.011	0.008	0.006	0.005	0.004
HPD	0.163, 0.271	0.037, 0.058	0.027, 0.042	0.013, 0.020	0.008, 0.013	0.007, 0.010	0.005, 0.008	0.004, 0.007	0.003, 0.005
Proportion	0.616	0.136	0.101	0.047	0.031	0.024	0.019	0.016	0.011
LC2	-0.127	0.007	-0.003	-0.168	0.029	0.260	0.190	0.212	-0.002
LC7	-0.333	0.040	0.098	0.226	-0.016	0.023	0.041	0.062	0.036
LC11	-0.409	-0.033	-0.052	-0.116	-0.068	-0.066	0.006	0.062	0.007
LC12	-0.323	0.008	0.022	-0.188	-0.004	0.069	-0.064	-0.111	-0.100
LC13	-0.280	0.000	0.003	-0.432	-0.011	-0.086	-0.045	-0.063	0.068
LC14	-0.317	0.006	0.002	0.067	0.051	-0.153	0.002	0.046	-0.057
LC17	0.070	0.043	0.054	0.017	-0.024	-0.082	-0.059	-0.045	0.000
LC19	-0.261	-0.003	-0.027	0.247	-0.077	-0.079	-0.057	-0.097	-0.012
LC20	-0.391	-0.002	-0.029	0.229	0.089	0.128	0.007	-0.028	0.049

Table S3.3 Eigenvectors of the sympatric **G** matrix, with associated eigenvalues and their 95% HPD intervals. The first four eigenvectors had statistical support as their 95% HPD intervals did not overlap with those of the null **G** estimated from randomized data.

	$\mathbf{g}_{max}$	$\mathbf{g}_2$	$\mathbf{g}_3$	$\mathbf{g}_4$	$\mathbf{g}_5$	$\mathbf{g}_6$	$\mathbf{g}_7$	$\mathbf{g}_8$	$\mathbf{g}_9$
eigenvalue	0.177	0.042	0.030	0.018	0.013	0.010	0.008	0.006	0.005
HPD	0.126, 0.237	0.031, 0.057	0.021, 0.037	0.013, 0.023	0.010, 0.017	0.007, 0.011	0.006, 0.009	0.005, 0.007	0.004, 0.006
Proportion	0.574	0.137	0.097	0.058	0.043	0.031	0.025	0.020	0.015
LC2	-0.127	0.007	-0.003	-0.168	0.029	0.260	0.190	0.212	-0.002
LC7	-0.333	0.040	0.098	0.226	-0.016	0.023	0.041	0.062	0.036
LC11	-0.409	-0.033	-0.052	-0.116	-0.068	-0.066	0.006	0.062	0.007
LC12	-0.323	0.008	0.022	-0.188	-0.004	0.069	-0.064	-0.111	-0.100
LC13	-0.280	0.000	0.003	-0.432	-0.011	-0.086	-0.045	-0.063	0.068
LC14	-0.317	0.006	0.002	0.067	0.051	-0.153	0.002	0.046	-0.057
LC17	0.070	0.043	0.054	0.017	-0.024	-0.082	-0.059	-0.045	0.000
LC19	-0.261	-0.003	-0.027	0.247	-0.077	-0.079	-0.057	-0.097	-0.012
LC20	-0.391	-0.002	-0.029	0.229	0.089	0.128	0.007	-0.028	0.049

Table S3.4 Eigenvectors of the allopatric **G** matrix, with associated eigenvalues and their 95% HPD intervals. The first three eigenvectors had statistical support as their 95% HPD intervals did not overlap with those of the null **G** estimated from randomized data.

	$\mathbf{g}_{max}$	$\mathbf{g}_2$	$\mathbf{g}_3$	$\mathbf{g}_4$	$\mathbf{g}_5$	$\mathbf{g}_6$	$\mathbf{g}_7$	$\mathbf{g}_8$	$\mathbf{g}_9$
eigenvalue	0.541	0.067	0.033	0.015	0.012	0.009	0.007	0.006	0.005
HPD	0.419, 0.661	0.051, 0.080	0.023. 0.044	0.012, 0.019	0.010, 0.015	0.007, 0.011	0.006, 0.008	0.005, 0.007	0.004, 0.006
proportion	0.778	0.096	0.048	0.022	0.017	0.013	0.011	0.009	0.007
LC2	-0.134	-0.223	-0.144	-0.041	0.001	0.161	0.255	0.140	0.026
LC7	-0.386	0.614	0.013	0.035	0.001	0.038	0.066	0.038	0.012
LC11	-0.451	-0.212	0.091	0.122	-0.008	0.001	0.022	0.050	0.028
LC12	-0.360	0.144	-0.044	-0.045	0.010	0.018	-0.022	-0.033	-0.010
LC13	-0.306	-0.049	-0.113	-0.028	-0.012	-0.022	-0.057	-0.051	-0.017
LC14	-0.357	0.007	0.004	-0.054	0.003	-0.093	-0.066	-0.030	-0.025
LC17	0.068	-0.011	-0.328	0.015	-0.010	-0.054	-0.073	-0.011	0.007
LC19	-0.255	-0.222	-0.043	0.078	-0.003	-0.054	-0.096	-0.118	-0.040
LC20	-0.402	-0.230	0.024	-0.096	0.006	0.014	-0.004	0.026	0.017

Figure S3.1 The ranges of *D. subquinaria* (black) and *D. recens* (white) as assessed by proportional abundance, modified from Robinson (2006). Sympatric region indicated by cooccurrence of both species (**A**). White square indicates area covered by (**B**), showing the sampling sites for populations of *D. subquinaria* used in chapters 3 and 4.

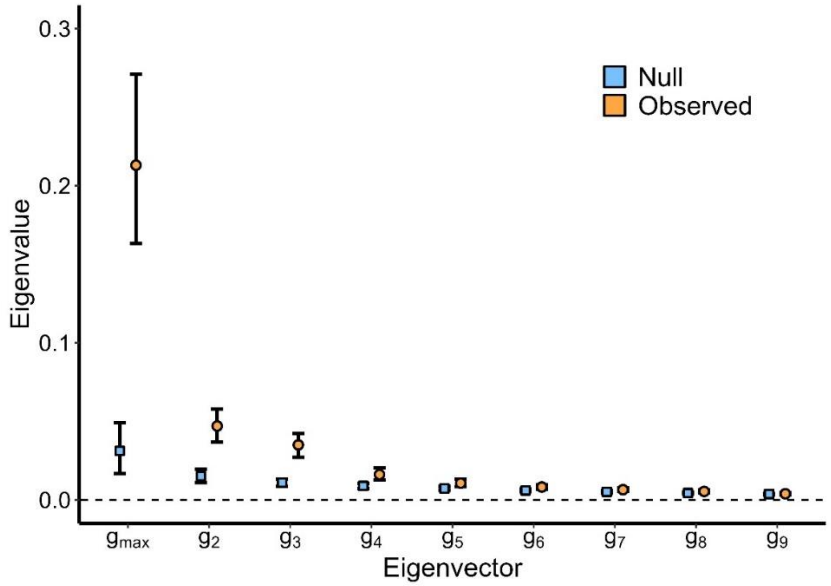

Figure S3.2 Mean (and 95% HPD) for the eigenvalues of each eigenvector of the overall **G**-matrix, as estimated with observed (orange circles) and randomized (blue squares) data. The first four eigenvectors have statistical support as the HPD intervals do not overlap.

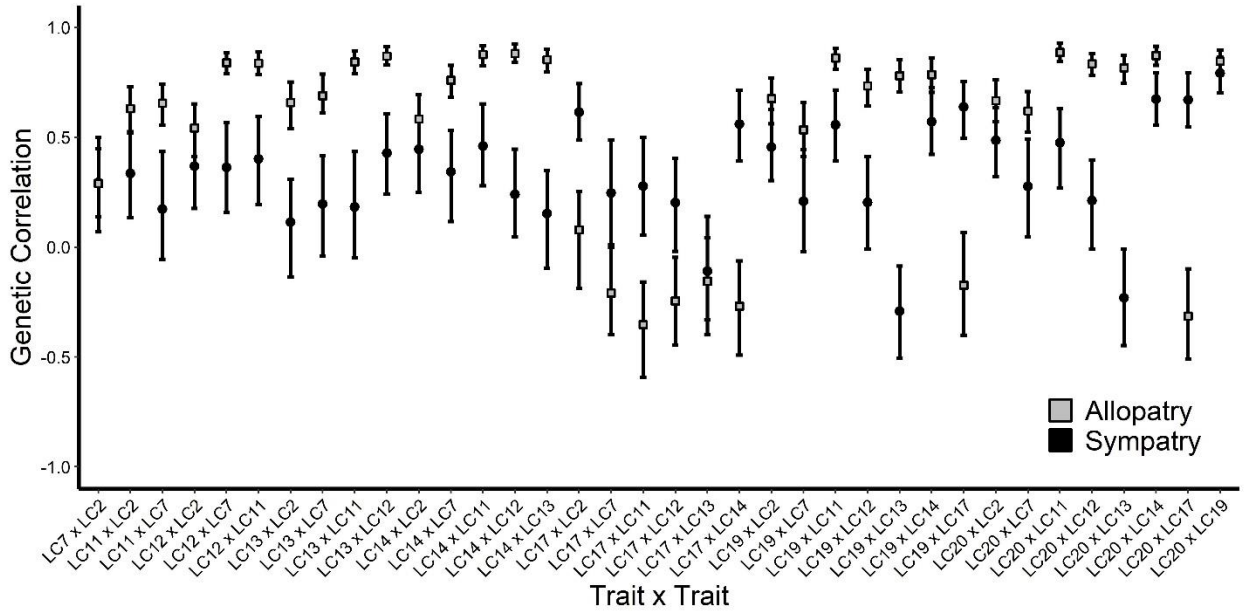

Figure S3.3 The mean and 95% HPD intervals for each trait-by-trait genetic correlation from the allopatric (grey squares) and sympatric (black circles) **G**-matrices.

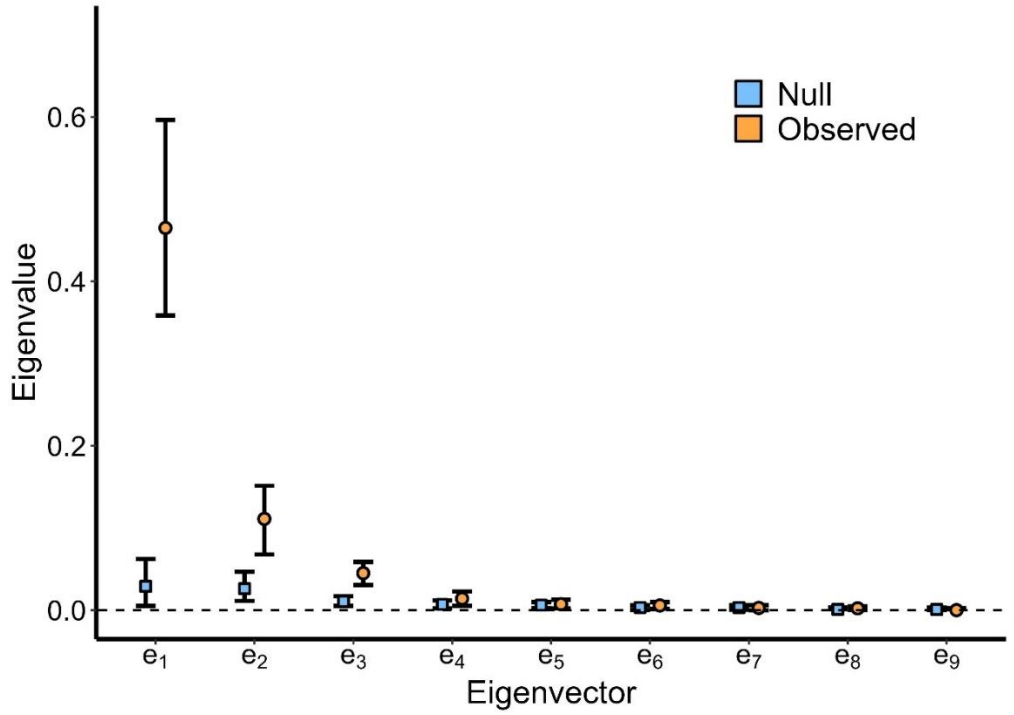

Figure S3.4 The eigenvalues and 95% HPD intervals for each eigenvector of the difference matrix ($\mathbf{G}_{difference}$), for observed (orange squares) and randomized (blue circles) data.

Supplementary Information: Chapter 4

Table S4.1 Details of likelihood-ratio tests comparing models estimating mating success with either linear or non-linear selection. Nonlinear selection could not be estimated in all populations due to a lack of power, although it was nonsignificant in any population where it could be estimated.

Region	Population	df	$\chi 2$	*p*
	Jocko	45	54.75	0.151
Allopatry	**Deer Creek**	45	NA	NA
	Lolo	45	38.23	0.752
	Canmore	45	46.33	0.417
Sympatry	**Sibbald**	45	NA	NA
	Opal	45	57.47	0.1

Table S4.2 Mean differences between regions in logcontrast CHCs that showed some signal of differential sexual selection. The relative abundance of these trait was higher in allopatry in each case.

Trait	Δ mean	*t*	*p*
LC14	**0.180**	**16.61**	**<0.0001**
LC19	**0.221**	**28.28**	**<0.0001**
LC20	**0.232**	**7.10**	**<0.0001**

Table S4.3 Results of separate Krzanowski comparisons quantifying the similarity in shared subspaces of the variance-covariance matrix of among-population phenotypic divergence (**D**) with a series of three variance-covariance matrices characterizing among-population variation in linear selection (**B**), and in the predicted responses to this selection assuming a single, overall **G** matrix ($\mathbf{R}_{single}$), region-specific **G** matrices ($\mathbf{R}_{region}$), or population-specific **G** matrices ($\mathbf{R}_{pop}$).

Comparison matrix	**Matrix similarity ($\Delta\lambda_S$)**	**Upper 95% quantile**	**Lower 95% quantile**
B	0.720	1.109	0.283
$\mathbf{R}_{single}$	1.431	1.749	1.036
$\mathbf{R}_{region}$	1.495	1.774	1.181
$\mathbf{R}_{pop}$	1.287	1.685	1.065

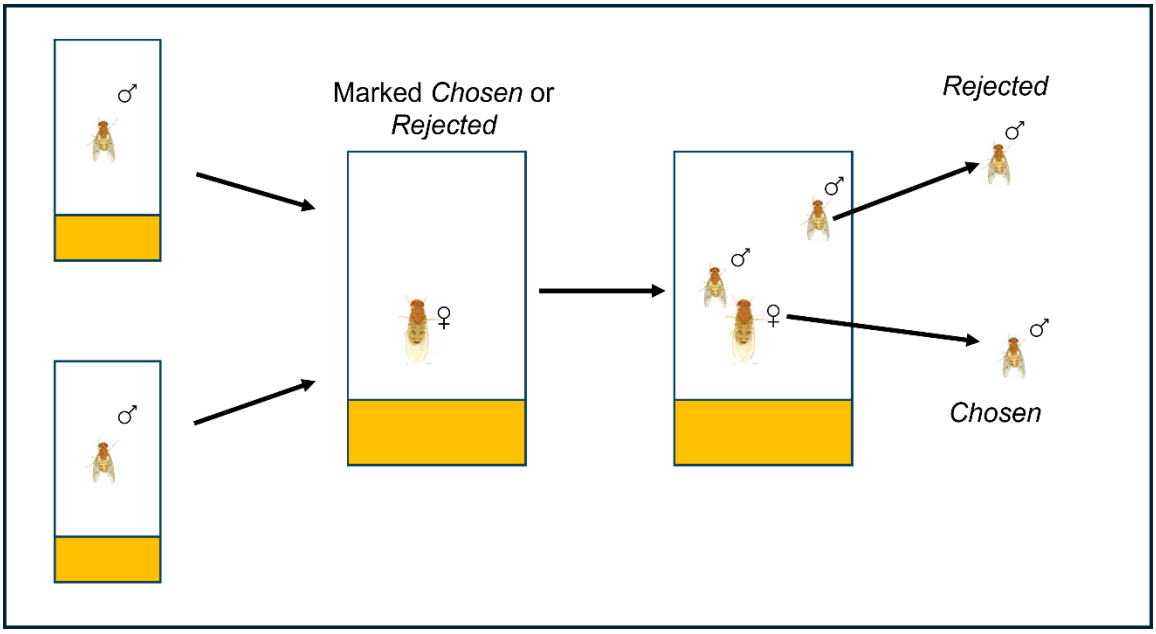

Figure S4.1 Schematic of the design of the mating trials used to estimate sexual selection gradients. Trial vials were either marked chosen or rejected (predetermined randomly) in a 50:50 ratio, and the corresponding male was taken after intromission was observed to have their CHC profile assessed.