The evolutionary ecology of hybridization

by

Ken A. Thompson

B.Sc., The University of Guelph, 2014
M.Sc., The University of Toronto, 2016

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate and Postdoctoral Studies

(Zoology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

May 2021

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

The evolutionary ecology of hybridization

submitted by Ken A. Thompson in partial fulfillment of the requirements for

the degree of Doctor of Philosphy

in Zoology

Examining Committee:

Dolph Schluter, Killam Professor, Zoology, UBC
Supervisor

Darren Irwin, Professor, Zoology, UBC
Supervisory Committee Member

Daniel Matute, Associate Professor, Biology, University of North Carolina at Chapel Hil
External Examiner

Judith Mank, Professor, Zoology, UBC
University Examiner

Jeannette Whitton, Associate Professor, Botany, UBC
University Examiner

Additional Supervisory Committee Members:

Sarah Otto, Professor, Zoology, UBC
Supervisory Committee Member

Loren Rieseberg, Professor, Botany, UBC
Supervisory Committee Member

Abstract

Hybridization is the process that mediates gene flow between sexually reproducing lineages. As such, the factors that determine the fitness of hybrids are critically important for speciation. Although hybridization between natural populations occurs in an explicitly ecological context, where hybrids must compete for resources and attract mates under prevailing ecological conditions, little attention has been paid to the mechanisms that mediate the ecological performance of hybrids. In Chapter 2, I expand upon existing models of speciation and find that standing variation improves hybrid fitness when parents adapt to similar environments, but reduces hybrid fitness when parents adapt to different environments. Chapter 3 systematically reviews the literature to report that F_1 hybrids are often quite mismatched for divergent parental traits, and also uses a field experiment with sunflowers to show that this mismatch reduces fitness via seed count. Chapter 4 also uses data from the literature to demonstrate that divergent adaptation proceeds via pleiotropic alleles. In Chapter 5, I measure morphological traits in the lab to illustrate that the magnitude of trait 'mismatch' in hybrids increases with the phenotypic distance between cross parents. Chapter 6 uses a pond experiment hybridizing 'parallel' ecotypes and reports that hybridization after parallel evolution results both in heterosis and hybrid breakdown. Finally, Chapter 7 uses DNA sequence data to illustrate that the genetic signature of hybrid incompatibilities - excess heterozygosity - is greater when stickleback hybrids are raised in ponds than when they are raised in aquaria. In sum, my thesis demonstrates that ecological selection against hybrids can result from rapid adaptation from standing variation, from mismatched trait combinations that evolve somewhat predictably, and can have a detectable genetic signature. Hybridization after parallel evolution, at least in stickleback, leads to relatively equal measures of hybrid breakdown and heterosis. Ecologically-mediated natural and sexual selection can clearly play a large role in mediating the fitness of hybrids - future work should aim to establish the importance of these processes for speciation more broadly.

Lay Summary

Hybridization—the generation of viable offspring following mating between different species—is increasingly recognized as being common. The fitness of hybrids, that is whether they can successfully survive and reproduce, is key for maintaining the parent species' distinctness. If hybrids readily survive and interbreed, this can cause the parent species to collapse into a hybrid swarm. In this thesis I generate novel hypotheses, syntheses, and results about how ecology acts to determine hybrid fitness. The first half of my thesis uses theory and synthesizes data to identify generalities about hybridization. The second half of my thesis uses original experiments with threespine stickleback fish to test these predictions. In sum, my thesis advances our understanding of the mechanisms through which ecological processes mediate the fitness of hybrids—what I call the 'evolutionary ecology' of hybridization—and contributes to the larger body of work on mechanisms of speciation.

Preface

Although the bulk of chapters are written with co-authors, I use the first-person singular pronoun, 'I' (and 'my'), throughout this thesis so that the language is consistent throughout. It can probably not be overemphasized that this thesis is largely the result of collaboration and my use of the first-person singular should in no way be interpreted as a dismissal of this fact.

A version of Chapter 2 has been published as Thompson, K.A., Osmond, M.M., and Schluter, D. 2019. Parallel genetic evolution and speciation from standing variation. *Evolution Letters*. 3(2): 129–141. The project was conceived by me and I refined the numerical and analytical approaches that were largely implemented by M. Osmond. I wrote the manuscript, analyzed, and plotted the data with regular input from M. Osmond. All co-authors contributed to manuscript edits.

A version of Chapter 3 has been published as Thompson, K.A., Urquhart-Cronish, M., Whitney, K.D., Rieseberg, L.H., and Schluter, D. Patterns, predictors, and consequences of dominance in hybrids. *The American Naturalist* 197(3) (pagination forthcoming). I conceived of the quantitative review and collected data for it with M. Urquhart-Cronish. K. Whitney had the idea to test our hypothesis with the sunflower data and collected and previously published the dataset. L. Rieseberg and D. Schluter provided input on analysis and study design, and all authors contributed to manuscript edits.

A version of Chapter 4 has been published as Thompson, K.A. 2020. Experimental hybridization studies suggest that pleiotropic alleles commonly underlie adaptive divergence between natural populations. *The American Naturalist* 196(1): E16–E22. The paper is sole-authored—I conceived the idea, analyzed the data, and wrote the paper. The data were collected as a part of the literature search in Chapter 3.

Chapter 5 is in preparation for submission with co-primary author A. Chhina and senior author D. Schluter. I conceived of the idea with input from D. Schluter, and conducted the fieldwork and most of the lab work. A. Chhina and I measured fish, analyzed data, and co-wrote the paper which had later input from D. Schluter.

Chapter 6 is in preparation for submission with co-author D. Schluter. The project was conceived by me and developed with input from D. Schluter. I conducted the field and lab work, and collected the data. I analyzed the data and wrote the first draft of the paper with input from D. Schluter.

Chapter 7 is in preparation for submission with co-authors C.L. Peichel, D. Schluter, D. Rennison, A. Albert, T. Vines, A. Greenwood, A. Wark, and M. Schumer. I conceived of the idea, curated data collection, analyzed the data, and wrote the paper with input from C. Peichel, D. Schluter, and M. Schumer. A. Greenwood, A. Wark and C. Peichel generated genetic data for the Paxton Lake aquaria fish. A. Albert and T. Vines generated genetic data for the Priest Lake aquaria fish. D. Rennison contributed additional data for pond-raised Paxton Lake hybrids.

Table of Contents

Abstract . . . iii

Lay Summary . . . iv

Preface . . . v

Table of Contents . . . vi

List of Tables . . . x

List of Figures . . . xi

Acknowledgements . . . xv

1 General Introduction . . . 1
- 1.1 Speciation and the role of hybrid fitness . . . 1
- 1.2 Evidence and mechanisms of extrinsic post-zygotic isolation . . . 2
- 1.3 Key areas requiring research and their motivation . . . 3
- 1.4 Overview of the thesis . . . 4
  - 1.4.1 Chapter 2 . . . 4
  - 1.4.2 Chapter 3 . . . 5
  - 1.4.3 Chapter 4 . . . 5
  - 1.4.4 Chapter 5 . . . 6
  - 1.4.5 Chapter 6 . . . 7
  - 1.4.6 Chapter 7 . . . 7
- 1.5 Remarks . . . 8

2 Parallel genetic evolution and speciation from standing variation . . . 9
- 2.1 Introduction . . . 9
- 2.2 Methods . . . 10
  - 2.2.1 Genotype to phenotype . . . 11
  - 2.2.2 Life-cycle . . . 11
  - 2.2.3 Adaptation to a new environment . . . 13
  - 2.2.4 Quantification of genetic parallelism and hybrid segregation variance & fitness . . . 13
- 2.3 Results . . . 15

2.3.1 Genetic parallelism and phenotypic segregation variance . . . 15
2.3.2 Hybrid fitness . . . 17
2.4 Discussion . . . 21
2.4.1 Key predictions and possible tests . . . 21
2.4.2 Alternative sources of standing variation . . . 22
2.4.3 Possible extensions . . . 23
2.4.4 Concluding remarks . . . 23

3 Patterns, predictors, and consequences of dominance in hybrids . . . 24
3.1 Introduction . . . 24
3.2 I: Patterns & predictors of dominance . . . 25
3.2.1 Methods . . . 25
3.2.2 Results . . . 30
3.3 II: Fitness consequences of parent-bias and mismatch in recombinant sunflowers . . . 31
3.3.1 Methods . . . 32
3.3.2 Results . . . 33
3.4 Discussion . . . 34
3.4.1 Genetic underpinnings of dominance and mismatch . . . 34
3.4.2 Patterns & predictors of dominance . . . 35
3.4.3 Fitness consequences of mismatch . . . 36
3.5 Conclusion . . . 37

4 Experimental hybridization studies suggest that pleiotropic alleles commonly underlie adaptive divergence between natural populations . . . 38
4.1 Introduction . . . 38
4.2 Methods . . . 39
4.3 Results . . . 40
4.4 Discussion . . . 42

5 Adaptive divergence and the evolution of hybrid trait mismatch in threespine stickleback . . . 45
5.1 Introduction . . . 45
5.2 Methods . . . 47
5.2.1 Study system . . . 47
5.2.2 Fish collection and husbandry . . . 47
5.2.3 Phenotype measurements . . . 49
5.2.4 Data analysis . . . 50
5.3 Results . . . 53
5.3.1 Patterns of phenotypic divergence among populations . . . 53
5.3.2 Evolution of trait mismatch . . . 53
5.3.3 Underlying causes of mismatch . . . 56

5.3.4 Patterns of dominance in hybrids . . . 58
5.4 Discussion . . . 59
5.4.1 Causes of divergence-mismatch relationship . . . 60
5.4.2 Causes of recessivity . . . 60
5.4.3 Evolution of dominance for armour . . . 61
5.4.4 Similarities to patterns of 'intrinsic' isolation . . . 62
5.4.5 Caveats and future directions . . . 62

6 Fitness consequences of hybridization after parallel evolution in threespine stickleback . . . 64
6.1 Introduction . . . 64
6.2 Methods . . . 66
6.2.1 Experimental crosses . . . 66
6.2.2 Pond experiment . . . 66
6.2.3 Lab experiment . . . 67
6.2.4 Data analysis . . . 68
6.2.5 Estimating fitness . . . 69
6.3 Results . . . 69
6.3.1 Patterns of survival and growth in ponds . . . 69
6.3.2 Causes of fitness differences among groups in ponds . . . 72
6.3.3 Patterns of growth and survival in aquaria . . . 73
6.4 Discussion . . . 74
6.4.1 Mechanisms underlying fitness differences between parents and hybrid generations in ponds . . . 75
6.4.2 Variation in hybrid fitness among crosses . . . 76
6.4.3 Caveats . . . 76
6.4.4 Progress toward speciation via parallel natural selection? . . . 77

7 Genetic evidence for environment-specific hybrid incompatibilities in threespine stickleback 78
7.1 Introduction . . . 78
7.2 Methods . . . 80
7.2.1 Data sources . . . 80
7.2.2 Marker filtering and estimating heterozygosity . . . 82
7.2.3 Data analysis . . . 83
7.2.4 Heuristic model . . . 83
7.3 Results and discussion . . . 85
7.3.1 Patterns of excess heterozygosity . . . 85
7.3.2 Heuristic model . . . 86
7.3.3 Alternative explanations for excess heterozygosity . . . 86
7.4 Conclusion . . . 89

8 Conclusion . . . 91
8.1 Overview . . . 91
8.2 Areas requiring research . . . 91
8.2.1 Mechanisms underlying ecological incompatibilities . . . 92
8.2.2 Theoretical importance of ecological incompatibilities . . . 93
8.3 A proposal to slightly refine the language of speciation . . . 94

Bibliography . . . 98

Appendices . . . 116

A Appendix for Chapter 2 . . . 117
A.1 Geometric basis for rapid loss of parallelism . . . 117
A.2 Supplementary figures . . . 119

B Appendix for Chapter 3 . . . 140
B.1 Supplementary methods . . . 140
B.1.1 Search strategy . . . 140
B.1.2 Evaluation of studies . . . 141
B.1.3 Data collection . . . 141
B.1.4 Estimating genetic divergence and divergence time . . . 142
B.1.5 Phylogenetic signal . . . 142
B.1.6 Field experiment with sunflowers . . . 143
B.2 Supplementary Tables & Figures . . . 144

C Appendix for Chapter 4 . . . 162
C.1 Supplementary figures . . . 165

D Appendix for Chapter 5 . . . 171
D.1.1 Supplementary methods . . . 171
D.2 Supplementary Tables & Figures . . . 173

E Appendix for Chapter 6 . . . 182
E.1 Supplementary methods . . . 182
E.1.1 Experimental animals . . . 182
E.1.2 Coded wire tagging . . . 183
E.1.3 Estimating fitness via fecundity and overwinter survival . . . 183
E.2 Supplementary Tables & Figures . . . 185

F Appendix for Chapter 7 . . . 192
F.1 Supplementary Figures . . . 192

List of Tables

2.1 Description of parameters and parameter values in parental populations for simulations presented in the main text 12

3.1 Hypothetical examples of possible F_1 trait values, and corresponding values for cross mean $d_{univariate}$, $d_{parent\text{-}bias}$, and $d_{mismatch}$ (parent phenotypes are scaled to [0, 0] and [1, 1]). 26

7.1 Summary of data sources 81

B.1 Description of sunflower traits used to quantify parent-bias and mismatch 145

B.2 Mean $\pm$ SE (n) for parent species and BC_1 hybrids at the common garden site. 146

C.1 Results with alternative choices for data filtering and binning 168

E.1 Summary of fish numbers and survival (re-capture numbers) for the 2020 pond experiment . 185

E.2 Fitness components and relative fitness estimates for cross types within species 186

List of Figures

1.1 Visual overview of underdominance and hybrid incompatibilities on 'intrinsic' fitness landscapes . . . 2

2.1 Visual overview of simulations and concepts . . . 11
2.2 Genetic parallelism and phenotypic segregation variance . . . 16
2.3 The relationship between trait dimensionality (m) and genetic parallelism . . . 17
2.4 The effect of standing variation on mean hybrid fitness. . . . 18
2.5 The effect of standing variation on the distribution of hybrid phenotypes . . . 19
2.6 Properties of alleles fixed during adaptation . . . 21

3.1 Visual overview of how two-dimensional dominance metrics were calculated . . . 27
3.2 Patterns of dominance in F_1 hybrids . . . 31
3.3 Effect of parent-bias and mismatch on fitness in *H. annuus* × *H. debilis* BC_1 hybrid sunflowers growing in the field . . . 33

4.1 Overview of adaptation with pleiotropic alleles and theoretical prediction. . . . 41
4.2 Scatterplot depicting the relationship between phenotypic divergence in parents (statistically divergent traits) and segregation variance in hybrids (statistically indistinguishable traits). . . . 42

5.1 Overview of sampling locations and trait measurements . . . 48
5.2 Visual overview of mismatch calculation and observed mismatch results . . . 55
5.3 Dominance is the primary driver of mismatch in F_1 hybrids, while variance is the primary driver of mismatch in F_2s . . . 56
5.4 F_2, but not F_1, hybrid phenotypic variation increases with the magnitude of phenotypic divergence between parents . . . 57
5.5 Dominance of the freshwater phenotype in F_1 hybrids among traits and populations . . . 59

6.1 Photograph of the study system . . . 66
6.2 Survival and growth of pure species and their F_1 & F_2 hybrids in the experimental ponds . . . 70
6.3 Variation in survival and growth among cross types and crosses in the experimental ponds . . . 72
6.4 Inferring the broad mechanisms underlying patterns of hybrid fitness . . . 73
6.5 Growth and survival in aquaria . . . 74

7.1 Results from simulations illustrating an ecological mechanism underlying the heterozygosity-incompatibility relationship in F_2 hybrids . . . 79
7.2 Patterns of heterozygosity in stickleback hybrids across environments . . . 87
7.3 Heuristic two-way incompatibility model to illustrate sufficient selection strengths to generate excess heterozygosity. . . . 88

8.1 A possible ecological incompatibility in monkeyflowers . . . 94
8.2 Fitness effects of incompatibility loci vary across environments and genetic backgrounds . . . 96

A.1 Genetic parallelism across the continuum of parallel to divergent natural selection (N = 100) 120
A.2 Genetic parallelism across the continuum of parallel to divergent natural selection (N = 1000) 121
A.3 Genetic parallelism across the continuum of parallel to divergent natural selection (N = 5000) 122
A.4 Effect of standing genetic variation on hybrid fitness across the continuum of parallel to divergent natural selection (N = 100) . . . 123
A.5 Effect of standing genetic variation on hybrid fitness across the continuum of parallel to divergent natural selection (N = 1000). . . . 124
A.6 Effect of standing genetic variation on hybrid fitness across the continuum of parallel to divergent natural selection (N = 5000). . . . 125
A.7 Properties of fixed mutations under a variety of parameter combinations (N = 1000). . . . 126
A.8 Simulations under various fitness functions. . . . 127
A.9 Mutation-selection balance and mutation effect sizes in ancestral populations. . . . 128
A.10 Mutation-selection balance and mutation effect sizes in ancestral populations under stronger selection (σanc = 1). . . . 129
A.11 The effects of standing genetic variation on genetic parallelism and phenotypic segregation variance in hybrids under parallel and divergent natural selection. . . . 130
A.12 Effect of standing variation on the pace of adaptation and attainment of mutation-selection-drift balance. . . . 131
A.13 Relationship between genetic parallelism and (A) segregation variance and (B) expected heterozygosity. . . . 132
A.14 Alternative presentation of simulation results across environments: distance between optima (δ). . . . 133
A.15 The effect of population size on the rate of divergence between populations due to drift. . . . 134
A.16 Effect of dimensionality on net segregation variance. . . . 135
A.17 Fraction of overlap of beneficial mutations with parallel selection ($\theta = 0°$) but unequal distance ($d_1 \neq d_2$) . . . 136
A.18 The effect of standing genetic variation (SGV) on relative maximum hybrid fitness across environments. . . . 137
A.19 The relationship between segregation variance and θ for different dimensionalities. . . . 138
A.20 Cartoon illustration of why divergence among populations does not affect whether an allele is beneficial in both of them. . . . 139

B.1 Relationship between SD divergence between parents (SDs in units of smaller parent value) and *P*-value of *t*-test evaluating whether parent phenotypes are statistically distinguishable . 144
B.2 Example visualisations of four cases of bivariate trait expression 147
B.3 Phylogeny of all species used in this study 148
B.4 $d_{univariate}$ across all traits in the datasets (n = 1046 traits) 149
B.5 Summary of cross median dominance metrics with each cross contributing a single value . . 150
B.6 Expected patterns based on sampling error alone (from simulations). 151
B.7 There are no differences (all $P > 0.25$) in any dominance metrics between intraspecific and interspecific crosses 152
B.8 No association between any dominance metrics and genetic distance between the parents . . 153
B.9 No differences in dominance between plants and animals. 154
B.10 Parental phenotypic variation (CV) has significant effects on dominance in hybrids 155
B.11 The relationship between trait type and univariate dominance ($d_{univariate}$ 156
B.12 Relationship between mean pairwise parent-bias and mismatch dominance in systematic review and sunflower data 157
B.13 Distribution of pairwise trait correlations in the sunflower data (BC_1s only) 158
B.14 Distribution of partial regression coefficients in multiple regression analyses 159
B.15 Phenotype and fitness distribution for a trait pair with substantial mismatch consequences in BC_1 hybrid sunflowers 160
B.16 Photograph of sunflower experiment at the Lady Bird Johnson Wildflower Center, in Austin, Texas, USA 161

C.1 Simulation results to illustrate theoretical prediction 166
C.2 Diagnostics of the linear model testing the main hypothesis 167
C.3 Segregation variance in non-divergent traits is not predicted by divergence in those traits . . 169
C.4 All available evidence suggests that phenotypic divergence between parents is uncorrelated with their divergence time 170

D.1 Repeatability data for all measured traits 173
D.2 Summary of pairwise trait correlations in F_2 hybrids 174
D.3 Freshwater divergence from the anadromous ancestor 175
D.4 Phenotypic divergence between parents is positively associated with the number of traits that differ between them, as well as the number of possible pairwise mismatches 176
D.5 Examples of pairwise divergence:mismatch regressions 177
D.6 Visualization of pairwise mismatch in empirical data 178
D.7 Dominance patterns in F_2 hybrids 179
D.8 Variation in dominance for lateral plate count does not impact the mismatch-divergence relationship 180
D.9 The number of mismatched trait pairs ‘snowballs’ with the magnitude of phenotypic divergence between parents, but only in F_1 hybrids 181

E.1 Relationship between initial and final mass in the 2020 pond experiment 187
E.2 Relationship between the number of days in pond and final mass in the 2020 pond experiment 188
E.3 Fitness components with the two parent populations plotted separately 189
E.4 Relationship between mass and standard length for fish collected from ponds 190
E.5 Relationship between standard length and fecundity in previously published stickleback pond experiments 191

F.1 Heterozygosity does not differ between F_2 and F_3 hybrids 192
F.2 Heterozygosity does not differ between studies involving Paxton Lake benthic $\times$ limnetic hybrids 193
F.3 Estimates of excess heterozygosity for individuals and loci across 'replicates'. 194
F.4 Test of main hypothesis using 'observed' heterozygosity rather than excess heterozygosity . 195
F.5 de Finetti ternary diagrams for genotyped *individuals* in each of the bi-parental crosses . . . 196
F.6 de Finetti ternary diagrams for genotyped *loci* in each of the bi-parental crosses 197
F.7 No relationship between individual mean heterozygosity and growth (standard length) in the aquarium-raised biparental benthic-limnetic F_2 hybrids 198
F.8 No relationship between estimated directional selection at each locus and its mean excess heterozygosity in ponds 199

Acknowledgements

This thesis would not have come together without the friendship and support of many. It has been great being a part of the Biodiversity Research Centre. There are so many people to thank and although it may prove unwise I will not list names here for fear of leaving someone deserving out. I cherish the friendship of the skit makers (long live!), the coffee drinkers, the discussion group presenters, the B.E.E.R.S. and retreat goers, the lunch eaters, the KUBB tossers, and the rental accommodation sharers. My supervisory committee—Darren Irwin, Loren Rieseberg, Sally Otto, and Rick Taylor—were incredibly generous with their time and insightful. Thanks also to my examining committee members, Judith Mank and Jeannette Whitton. My supervisor, Dolph Schluter, provided invaluable guidance and support throughout my degree. Most of my thesis chapters were collaborations, and I am grateful to have had the opportunity to work with so many great scientists. Members of the Schluter lab, from undergraduates to post-docs, have provided consistent feedback on my science and support in field and lab. I have had the great pleasure to work with several collaborators on my thesis chapters and am immensly grateful to them all. Patrick Tamkee and Eric Lotto, the inSEAS core staff, provided outstanding logistical support without which my ichthyological pursuits would surely have failed. Thanks to NSERC, UBC, and the Zoology Department for financial support. My family encouraged me to pursue my interests wherever they made lead and gave me unconditional support all the way through. The companionship of my partner, Mackenzie, has been indescribably valuable throughout my time in Vancouver and will no doubt continue to be into the future.

Chapter 1

General Introduction

1.1 Speciation and the role of hybrid fitness

Speciation—the evolution of reproductive isolating barriers between diverging lineages (Coyne and Orr, 2004)—is the process largely responsible for maintaining the wondrous diversity of life. If not for these evolved isolating barriers, discontinuities could never persist between lineages. Consider the cedars and firs that grow on and near the UBC campus. Although plenty of cedar pollen might find its way onto the ovules of the firs, for various reasons no hybrid seeds are produced from this pollen and the two species remain distinct. With such extreme barriers it is easy to see how these lineages remain distinct: there is simply no possible way for cedar genes to introgress into fir, or vice-versa.

Clearly, complete gametic incompatibility is sufficient for speciation. Importantly, however, it is not necessary because plenty of species stably coexist in sympatry despite complete inter-fertility. For example, the threespine stickleback (*Gasterosteus aculeatus* L.) species pairs persist in lakes in coastal British Columbia, Canada, even though they produce perfectly vigorous hybrids in the lab (McPhail, 1984, 1992; Hatfield and Schluter, 1999). Hybrid zones, where hybrids are produced and exist in a continuum from one species to another (Barton and Hewitt, 1985), form between thousands of populations that nevertheless continue to persist as largely distinct lineages. Sympatric species that produce viable and fertile hybrids in benign (e.g., laboratory) environments can be maintained if hybrids cannot survive and reproduce under the prevailing ecological conditions due to a maladapted phenotype—an isolating barrier referred to as 'extrinsic postzygotic isolation'. Recent theory suggests that postzygotic barriers can be more effective at preventing gene flow than assortative mating in hybrid zones (i.e., pre-mating isolation; Irwin 2020). In many cases where species do not hybridize as a result of assortative mating, this is due to the process of reinforcement, whereby assortative mating evolves as a direct result of hybrids being unfit (McKinnon and Rundle, 2002; Hopkins, 2013). Thus, even when pre-mating isolation is apparent it might be an adaptive product of selection against hybridization.

Ecological (natural or sexual) selection against hybrids can be a sufficiently strong barrier to maintain the distinctiveness of sympatric reproductively isolated lineages in nature. Yet, we still lack a general understanding of how important extrinsic post-zygotic isolation is for speciation and how it evolves. If hybrids tended to perform just as well as parents in the field, how many fewer species would there be? Part of the reason for this is that we do not as a field know much about the mechanisms through which extrinsic post-zygotic isolation operates. The goal of this thesis is to generate hypotheses, document patterns, and test mechanisms of ecologically-mediated hybrid fitness in the field. In this brief introduction, I summarise progress to date and key gaps, then conclude with an overview of what follows in the thesis.

1.2 Evidence and mechanisms of extrinsic post-zygotic isolation

Much of what is known about post-zygotic isolation, especially its genetic basis, pertains to what are referred to as 'intrinsic' post-zygotic isolating barriers. Intrinsic incompatibilities are defined as those that affect the viability and/or fertility of individual hybrids possessing them, and is typically measured in the lab. The theory of hybrid incompatibilities, arrived at independently by Bateson (1909), Dobzhansky (1937), and Muller (1942), solved an important problem in speciation research. That is: how do genetic variants that kill or sterilize intermediate forms evolve? Such 'underdominant' alleles cannot readily evolve because intermediate steps are inviable. The key intellectual advance was that the authors realized that this process was much easier when two (or more) loci are involved and when the incompatibility is caused by their interaction (Fig. 1.1B). With the theory of hybrid incompatibilities in tow, it was no longer a mystery how evolution could proceed in a manner that rendered some hybrids inviable or sterile.

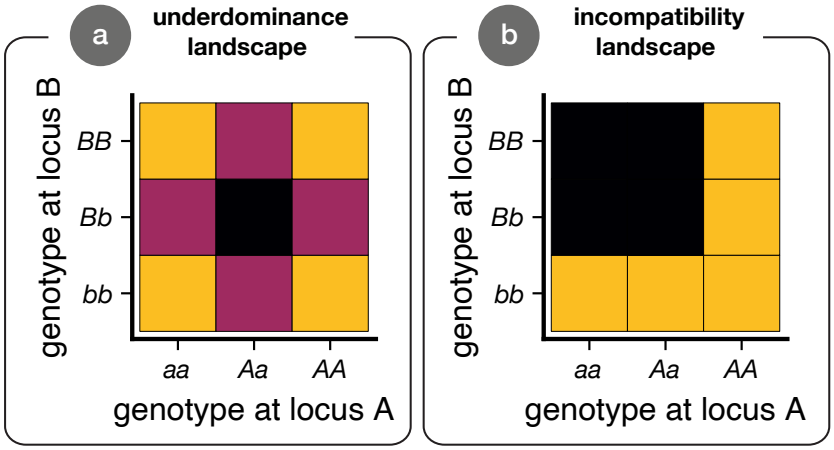


Figure 1.1: **Visual overview of underdominance and hybrid incompatibilities on 'intrinsic' fitness landscapes.** Both panels are fitness landscapes for loci (yellow is high fitness, purple is intermediate, and black is low). Panel (a) shows underdominance, wherein heterozygosity is disadvantageous. In this case, fitness can be predicted from heterozygosity alone. Panel (b) shows a case of hybrid incompatibilities, where fitness is affected by the interaction of alleles at two loci. In this case, the (big) *B* allele is incompatible with the (small) *a* allele.

In the late 1990s and early 2000s, the study of 'ecological speciation' came to prominence (Schluter, 1996, 2000, 2001; Rundle, 2002; Nosil, 2012; Ostevik et al., 2012). In large part due to this burgeoning field, evolutionary biologists became convinced that adaptation to prevailing ecological conditions could cause divergent lineages to produce hybrids with low fitness due to their poor fit under these conditions (Coyne and Orr, 2004; Langerhans and Riesch, 2013). All throughout this work, substantial progress was being made in generating and testing predictions about the genetic basis of intrinsic post-zygotic isolation (Orr, 1995; Presgraves, 2003, 2010; Matute et al., 2010; Moyle and Nakazato, 2010; Corbett-Detig et al., 2013; Wang et al., 2015), but little in the way of mechanistic connection was made between this work and those studying ecological speciation (also called 'speciation-by-adaptation'). That is, although hybrid incompatibilities have come to predominate the theory of intrinsic isolating barriers, they have not historically been considered important for extrinsic isolation.

In recent years, however, researchers have begun to identify how interactions among divergent genes—hybrid incompatibilities—can affect fitness in an ecological context. Typically, this work has been done looking at the interactions of divergent phenotypes. Vinšálková and Gvoždík (2007) found that newt hybrids had maternally inherited temperature preferences but were morphologically intermediate, potentially rendering them unfit in their preferred thermal niche. Matsubayashi et al. (2010) reviewed the literature and found that hybrids between several divergent phytophagous insect species often have better performance on one parent species' native host plant but have a genetic predisposition to be attracted to the other species', potentially reducing their fitness or that of their offspring. Arnegard et al. (2014) examined the relationship between trait values and fitness in F_2 hybrids between divergent stickleback species and found that individuals with mismatched jaw-traits performed worse than fish whose traits were relatively matched. Keagy et al. (2016) made similar crosses to Arnegard et al. (2014) and found that trait interactions governed male attractiveness to females. Cooper et al. (2018) found that natural *Drosophila* hybrids were attracted to xeric habitats like one of their parents but had low desiccation tolerance, similar to the other. Other studies have documented that more fitness conferred via more ecologically-relevant traits, like the ability to locate food (Turissini et al., 2017), declines with the genetic distance between parents, similar to 'intrinsic' postzygotic barriers in other groups (Coyne and Orr, 1989, 1997). Collectively, these studies imply that interactions among divergent loci, acting through trait 'mismatch' wherein hybrids express maladaptive combinations of traits, can be a critical determinant of the fitness of inter-population hybrids in natural environments.

Although the same general process—epistasis among alleles of different ancestry—can in principle underlie both intrinsic and extrinsic post-zygotic isolation, there are several key differences that might differentiate intrinsic and extrinsic incompatibilities. For brevity, I will list only two here. First, the fitness landscape of individual incompatibility loci likely has much shallower topography when these incompatibilities act via ecological selection (i.e., the fitness of the most fit ancestry combination is not much higher than that of the least fit). This is because individual extrinsic incompatibilities likely reduce fitness (i.e., survival probability and fecundity) by smaller magnitudes than intrinsic incompatibilities and thus, the 'problem' of fitness-valley crossing (Dobzhansky, 1937; Osmond and Otto, 2015) might cease to be a problem at all. Second, fitness landscapes of extrinsic incompatibilities are expected to be extremely context-dependent. Whether or not a pair of alleles is incompatible could readily vary across years, seasons, and community contexts. These differences render it difficult to evaluate the extent to which our theoretical and empirical knowledge of intrinsic post-zygotic isolating barriers translates to the study of extrinsic isolation. Much work remains to assess the prevalence, phenotypic mechanisms, and effect sizes of ecological hybrid incompatibilities. This thesis aims to take several steps forward on these fronts.

1.3 Key areas requiring research and their motivation

As discussed above, recent studies have shown that divergent alleles can interact to determine hybrid fitness in ecologically-relevant contexts. However, it does not tell us much about the importance of these processes for the larger questions we care about such as their role in speciation. Said another way, if not for ecologically-mediated hybrid incompatibilities, how much more slowly would lineages diversify? Ulti-

mately, as evolutionary biologists studying speciation, our goal is to arrive at generalities by documenting the processes that drive speciation and clarifying their relative importance. The limited available evidence makes ecological hybrid incompatibilities appear as a critically understudied and important determinant of hybrid fitness. Only by identifying generalities about the evolutionary ecology of post-zygotic isolation can we truly appraise their role in speciation.

Although a thesis structured around the topic of 'ecologically-mediated hybrid incompatibilities' might seem a narrow one at first glance, I hope to convince the reader by the end that it is more than broad enough to sustain several careers' worth of exciting research. Many more questions exist than do answers, and this thesis work just scratches the surface of the work remaining on this topic. We do not know how readily adaptation gives rise to ecological incompatibilities nor do we understand the processes that might expedite or slow this process. Research is required to clarify how common and severe trait mismatch is in hybrids, and data that directly link individual-level trait mismatch estimates to fitness are entirely lacking. We do not know if mismatch in hybrids changes deterministically as populations diverge, though theory predicts it will. Theory on hybrid ecological incompatibilities has also seldom been tested, and ample work remains to be done to see if we can predict the fitness effects of incompatibilities in the field and whether we can detect their genetic signature. Addressing these pressing topics will establish a baseline for future work in the area. In the section that follows, I outline how my thesis attempts to accomplish this.

1.4 Overview of the thesis

This thesis builds on a range of theoretical and empirical studies to advance the state of knowledge of the mechanisms through which adaptation affects the fitness of hybrids. My aim was to generate novel hypotheses, establish empirical patterns, and test theoretical predictions. The study of hybridization is wonderful because the problem is amenable to study from multiple perspectives. I use numerical and analytical theory, data synthesis, and original experiments to address my research questions. The chapters are outlined below, and this general introduction closes with similarly general remarks about the thesis document.

1.4.1 Chapter 2

In Chapter 2, I use theoretical models to evaluate how adaptation from standing genetic variation affects progress toward speciation. This area is a topic of pressing interest because it is increasingly clear that adaptation occurs from both the re-assortment of existing genetic variation and through the use of new (i.e., *de novo*) mutations. This chapter uses simulations where pairs of populations adapt from either new mutation alone, or new mutation and ancestral standing variation, and asks how this difference affects the phenotypes and fitness of interpopulation hybrids that form after secondary contact. In the process of this investigation, my co-authors and I also derived new analytical and theoretical predictions about how the probability of parallel genetic evolution changes as the environments to which populations are adapting become increasingly different.

The chapter has several key results. First, standing variation slows progress to speciation under parallel selection because populations are more likely to fix the same alleles during adaptation and hybrids have

phenotypes resembling parents and thus high fitness. Second, standing variation makes speciation faster via divergent selection. This is not because it makes adaptation happen more quickly (though it does do this) nor is it because of parallel genetic evolution (populations fix different alleles). Rather, populations tend to fix more pleiotropic alleles when adapting from standing variation and as a result exhibit maladaptive transgressive phenotypes. Another key result of the study was that the probability of parallel genetic evolution rapidly declines to zero as the phenotypic optima diverge. Critically, this decline happens more rapidly in more complex organisms (i.e., those with more traits or higher 'dimensionality'). This work generates a number of novel testable predictions, and provides insight into how major features of the genetics of adaptation affect progress to speciation via environment-specific selection against hybrids.

1.4.2 Chapter 3

In Chapter 3, I aimed to address three key questions. First, how are traits generally inherited in F_1 hybrids between divergent natural populations? Second, can we make predictions about F_1 trait values relative to parents from features of a cross, such as its taxon or the genetic divergence that separates parents? And does individual-level variation in trait mismatch have a predictable link to fitness in the field? To address the first two questions, I synthesized data from nearly 200 studies that measured the traits of F_1 hybrids and both parents in the same environment. For the third question, I re-analyzed an existing dataset of recombinant hybrid sunflowers transplanted into the field.

I found that individual traits, as measured in F_1 hybrids, are considerably more dominant than additive (i.e., individual traits are much more similar to one parent than to the other; $h = 0.2$ or 0.8). I also found that dominance is inconsistent among traits, and this causes hybrids to generally resemble one parent a bit more than the other. In addition, dominance causes F_1 hybrids to be rather 'mismatched' for divergent traits, where in the case of two traits F_1 hybrids resemble one parent for one of the traits but resemble the other parent for the second trait. I also found that no features of a cross (e.g., taxon, genetic distance between parents) reliably predicted any metric of dominance. In the sunflowers, I found that multivariate bias towards parents (i.e., resembling one parent more than the other or 'parent-bias') improved fitness while mismatch between traits tended to reduce fitness. Importantly, the effect of mismatch for *reducing* fitness was larger in magnitude (i.e., steeper slope) than that of parent-bias for *improving* it. This chapter reveals that trait mismatch in early-generation hybrids is more a rule than an exception and that this mismatch can, in at least one system, substantially reduce their fitness in the field.

1.4.3 Chapter 4

Chapter 4 uses data from the literature to test whether adaptive divergence proceeds via the fixation of alleles with deleterious pleiotropic side-effects. Pleiotropy is a central concept in many evolutionary models, from being the major explanation for aging (Williams, 1957) to being a central assumption of Fisher's (1930) geometric model. In spite of its central importance, little is known about the degree to which mutations affect few or many traits—that is, we know little about the 'universality' of pleiotropy. Determining the universality of pleiotropy is difficult because measuring it is typically painstaking—researchers typically must induce a single mutation and then measure how many traits it affects. Rather than undertaking such an experiment, I

leveraged a prediction of theoretical models that depends on there being universal pleiotropy and is readily testable using existing data. Specifically, if adaptation uses pleiotropic mutations, compensatory mutations are expected to fix that counteract the deleterious pleiotropy. Following hybridization, these pleiotropic and compensatory mutations should segregate in recombinant hybrids, resulting in hybrids with substantial transgressive phenotypic variation in traits for which the parents have the same phenotype. Importantly, the amount of segregation in these non-divergent traits is expected to be positively associated with the magnitude of phenotypic divergence among populations for divergent traits.

I assembled a small dataset of 15 crosses to test this prediction. I found strong support for the theoretical prediction that was robust to virtually all possible alternative analysis decisions. Tests of alternative explanations, such as that the pattern could be caused by drift, were not supported. This project provides positive, albeit strictly correlational, support for a key prediction of models that rely on the assumption of universal pleiotropy. Universal pleiotropy is a key assumption of Chapter 2 and many other models of adaptation and has major implications for the genotype–phenotype map. Thus, models that rely on this assumption are increasingly supported by empirical evidence, and we can conclude that varying single alleles will typically have off-target effects. Experiments that follow-up on this result to establish its generality would be timely.

1.4.4 Chapter 5

In Chapter 5, I use threespine stickleback fish (*Gasterosteus aculeatus* L.) to test the hypothesis that trait mismatch is associated with the amount of divergence between parents of a cross. In the threespine stickleback, marine populations colonized a range of post-glacial freshwater lakes to which they have adapted over the previous 10,000+ years. The freshwater lakes vary greatly from one another, with some being massive and containing both piscivorous species and species that compete for food with stickleback, and others being small and shallow with no other fish species except stickleback. The former, being deep and species-rich waters, resemble the ancestral marine habitat of the open ocean and stickleback in these lakes tend to retain many of the ancestral characteristics and zooplanktivorous niche. More derived stickleback inhabit lakes in the latter category—those that are small and species-poor—and are specialized to feed on benthic invertebrates. Lakes that are intermediate between these extremes abound and are home to relatively intermediate stickleback.

Theory predicts that the amount of segregating phenotypic variation in recombinant hybrids should increase as populations diverge phenotypically. This should, in turn, lead to hybrids with increasingly novel combinations of traits. I used a 'space-for-time' design to test this prediction in stickleback, because the zooplanktivore–benthivore divergence axis is an effective and meaningful axis of divergence that is positively associated with genomic divergence. I crossed fish from a single marine (anadromous) stickleback population with twelve derived freshwater populations of increasing divergence from the marine form. I then measured traits of lab-raised parents and hybrids (F_1 and F_2) and used these data to quantify the magnitude of hybrid mismatch. I found that, as predicted, hybrid mismatch was greater in 'wider' crosses. However, the details were surprising. First, although I expected the pattern only to occur in F_2 hybrids, mismatch increased with divergence in F_1s as well because dominance varied considerably among traits. Further analysis clearly indicates that the cause of mismatch in F_1s is dominance and the cause of mismatch in F_2s is segregation

variance. Together, this chapter indicates that hybrid mismatch evolves predictably in stickleback, which might be an important mechanism linking divergent adaptation with reproductive isolation.

1.4.5 Chapter 6

In the past three decades, it has become increasingly clear that adaptation in response to divergent natural selection readily leads to speciation. Much less is known, however, about the efficacy of parallel natural selection for driving speciation. In Chapter 6, I tested the hypothesis that parallel phenotypic evolution, if underpinned by non-parallelism at the genetic level, can lead to the accumulation of hybrid incompatibilities. Such incompatibilities would manifest as novel phenotypes in hybrids that render them poorly adapted to the phenotypic optimum of their (phenotypically similar) parent populations.

To test whether parallel evolution leads to hybrid incompatibilities, I used the stickleback benthic-limnetic species pairs, which are among the best models of parallel phenotypic evolution yet discovered in fishes. I made between-lake within-species crosses between all three of the extant species pairs for both benthics and limnetics, and compared growth rates of parents and F_1 & F_2 hybrids in experimental ponds and in aquaria. I hypothesized that because benthics are more derived, they would have evolved more incompatibilities than limnetics. I tracked the growth of nearly 4,000 individual fish in ponds and 2,000 fish in aquaria, and found both expected and unexpected results.

Surprisingly, patterns for benthics and limnetics were similar in terms of the rank-order fitness of pure species and hybrid classes. Differences among cross types were also larger in magnitude among limnetics, refuting the hypothesis that patterns would be more exaggerated in benthics. Across both species, I found evidence for heterosis in the ponds, with F_1 hybrids growing faster and surviving more frequently than both pure species and F_2 hybrids. F_2 hybrids often performed similarly or worse than parents, however, and a simple toy model demonstrates that this pattern can only be caused by the existence of hybrid incompatibilities between divergent alleles at different loci. I found no evidence for hybrid incompatibility in the aquarium-raised fish, whereas heterosis was apparent for limnetics in aquaria. This suggests that hybrid incompatibilities are only expressed in the field, whereas heterosis is somewhat intrinsic in limnetics and extrinsic in benthics. Finally, although these main effects were significant, patterns differed meaningfully across inter-lake crosses, which confirms a primary role for stochastic processes during speciation by parallel natural selection. Nevertheless, the fact that the main effects were so similar for both benthics and limnetics suggests that general rules governing the fitness of hybrids might become apparent through the noise if enough comparisons are made.

1.4.6 Chapter 7

In Chapter 7, the final data chapter, I use the benthic-limnetic stickleback species pairs and allopatric anadromous-freshwater populations to test whether the genetic signature of hybrid incompatibilities is stronger in the field than in the lab. I can expect this prediction to be supported if incompatibilities affect traits like the ability to find or capture food or avoid predators. In F_2 hybrid populations, theory predicts that the net strength of selection against hybrid incompatibilities can be measured as directional selection for heterozygosity. I compared patterns of excess heterozygosity in benthic-limnetic F_2 hybrids from both Priest and

Paxton Lakes between populations raised in the lab vs. experimental ponds. I did the same for fish from Cranby Lake crossed with anadromous fish from the Little Campbell River. If incompatibilities are only expressed under ecologically-relevant circumstances, they should only be detected in the ponds.

I found that in aquaria, F_2s met Hardy-Weinberg expectations and had no significant excess heterozygosity. In the ponds, however, surviving fish almost invariably had significant excess heterozygosity which I hypothesize results from mortality-selection against individuals with many weakly-selected ecological incompatibilities. This work illustrates that hybrid incompatibilities can be environment dependent and highlights a simple summary statistic that can be used to test this. Environment-specific heterosis resulting from dominance might also contribute to this pattern, especially in the allopatric marine-freshwater crosses. More importantly, it implies that ecological hybrid incompatibilities might evolve before 'intrinsic' incompatibilities and could be the genetic basis of much ecology-based reproductive isolation in nature.

1.5 Remarks

In this introduction, I hope to have convinced the reader that the evolutionary ecology of hybridization is a rich topic with opportunity for theoretical & empirical generalities waiting to be established and predictions ripe for testing in the field and lab. That is, I hope the work presented herein raises more questions than it answers. The six chapters that follow comprise the bulk of the thesis, and I conclude the document by summarizing its findings, discussing what we know now that we didn't before, and by highlighting key areas where progress should be made to solidify our understanding about the general importance and mechanisms of extrinsic post-zygotic isolating barriers for speciation.

Chapter 2

Parallel genetic evolution and speciation from standing variation[1]

2.1 Introduction

In recent years, two general features of evolution by natural selection have become increasingly established. First, adaptation often proceeds largely via the reassortment of ancestral standing variation rather than via complete reliance on *de novo* mutations (Barrett and Schluter, 2008). And second, variation in the direction of natural selection acting on pairs of populations is best represented by a quantitative continuum ranging from parallel selection—favouring identical phenotypes—to divergent selection—favouring distinct phenotypes— rather than falling into discrete 'parallel' or 'divergent' bins (Bolnick et al. 2018). It is unclear, however, how the extent of parallel genetic evolution—use of the same alleles during adaptation—changes with the difference in the direction of selection experienced by a pair of populations. It is also unclear whether adaptation from standing variation has implications for speciation that are distinct from those when adaptation is from new mutation alone, and whether its effect changes along the continuum from parallel to divergent natural selection. Here, I investigate genetic parallelism and speciation under adaptation from standing variation across this selection continuum.

Adaptation facilitates progress towards speciation when populations evolve reproductive isolating barriers as a by-product. One reason these reproductive isolating barriers might arise is because genetic differences between populations have maladaptive consequences when combined in hybrids (i.e., postzygotic isolation), thereby reducing gene flow upon secondary contact. When a pair of populations adapts in response to divergent natural selection, hybrids might have an intermediate phenotype that is unfit in either parental environment (Schluter 2000). When a pair of populations are subject to parallel selection, they may diverge genetically by chance (Mani and Clarke 1990; Schluter 2009) and hybrids might have novel transgressive phenotypes that are poorly suited to the common parental habitat (Barton 1989, 2001). Hybrid unfitness is therefore determined by two factors: (1) additive gene action causing hybrids to 'fall between the peaks' (Rundle and Whitlock 2001), and (2) cryptic genetic divergence that is released following hybridization and causes some hybrids to possess maladaptive transgressive phenotypes which vary in directions orthogonal to the axis of parental divergence (Arnegard et al. 2014; Keagy et al. 2016). How adaptation from standing variation affects progress toward speciation-by-selection (Langerhans and Riesch 2013) is largely unexplored theoretically.

[1] A version of this chapter has been published as Thompson, K.A., Osmond, M.M., and Schluter, D. 2019. Parallel genetic evolution and speciation from standing variation. *Evolution Letters*. 3(2): 129–141.

Adaptation from standing variation is common and underlies some of the most spectacular adaptive radiations found in nature (Brawand et al. 2015). Genomic studies often implicate standing variation as the major source of genetic parallelism in replicate populations colonizing similar environments (Jones et al., 2012*a*,*b*; Roesti et al., 2014; Lee and Coop, 2017; Haenel et al., 2019) and adapting to novel stressors (Reid et al., 2016; Alves et al., 2019). Previous research has shown that the correlation between selection coefficients of a given allele in each of two populations inhabiting different environments is expected to increase with the similarity in the direction of selection (equation 6 in Martin and Lenormand 2015). I therefore expected the extent of parallel genetic evolution for two populations to decline from a maximum to a minimum value as the angle between the directions of selection between them (θ) increases from completely parallel ($\theta = 0°$) to completely divergent ($\theta = 180°$). My specific goal was to characterize the pattern of decline in parallelism. I also hypothesized that adaptation from standing variation would reduce the evolution of reproductive isolation under parallel selection because parental populations would fix more of the same alleles and therefore evolve fewer incompatibilities (Schluter, 2009). Under divergent selection, I hypothesized that populations would fix alternative alleles regardless of whether they were selected from standing variation or new mutation. Therefore, I expected standing variation to have little effect on speciation by divergent selection compared to adaptation from new mutation alone.

I conducted a theoretical investigation into parallel genetic evolution and speciation from standing variation across the continuum from parallel to divergent natural selection. I primarily used individual-based simulations and include some simple analytical arguments to gain intuition. I compared results from simulations where adaptation proceeds simultaneously via the sorting of ancestral standing genetic variation and *de novo* mutation to simulations where adaptation proceeds via *de novo* mutation alone. My results provide insight into the circumstances under which I should expect high vs. low genetic parallelism and also suggest that standing variation has substantial implications for speciation that depend on the difference in the direction of natural selection between populations.

2.2 Methods

I used computer simulations to investigate genetic parallelism and progress toward speciation—via ecologically-dependent postzygotic reproductive isolation—from standing variation across the continuum from parallel to divergent natural selection. My simulations consider pairs of populations and multivariate phenotypes determined by multiple additive loci. In each of my simulations, a single ancestral population founds two identical populations that each adapt in their respective environments without gene flow (i.e., allopatry; see Fig. 2.1A). After adaptation, populations interbreed to form recombinant hybrids. This general colonization history—a single population splitting into two populations that independently adapt to their respective novel environments—is modelled around the process of adaptation as it can occur in nature, for example in postglacial fishes (Bell and Foster 1994) and in birds or plants isolated within glacial refugia (e.g., (Weir and Schluter 2004; Pettengill and Moeller 2012). In many such cases, ecologically-dependent postzygotic isolation is thought to be essential for maintaining reproductive isolation (Nosil 2012). See Table 2.1 for descriptions of all parameters and values used in simulations.

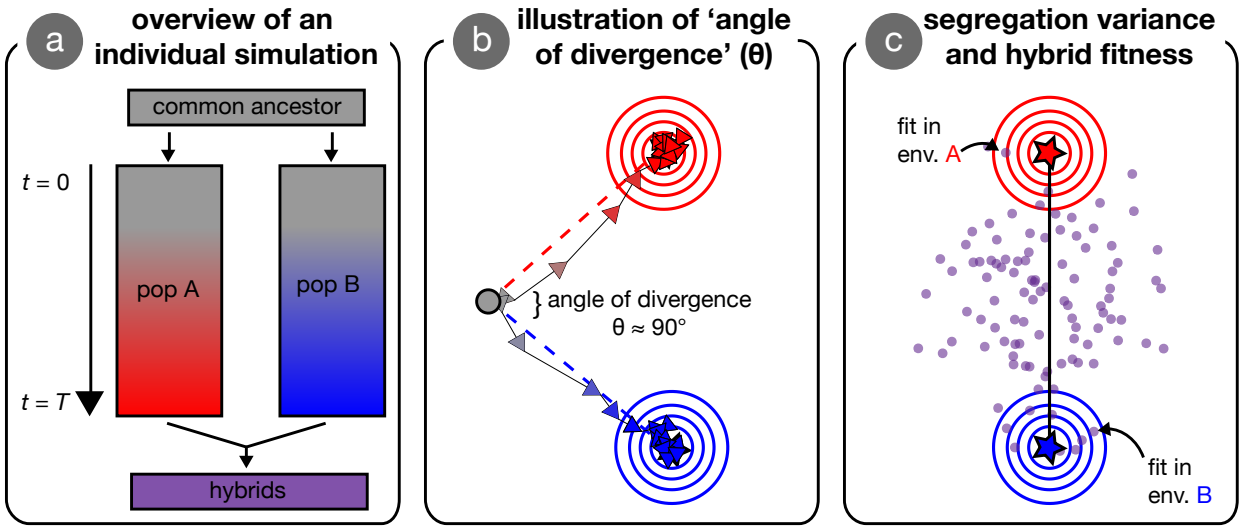


Figure 2.1: **Visual overview of simulations and concepts.** Panel (a) provides an overview of an individual simulation run. An ancestral population founds two initially-identical parental populations, that evolve independently for T generations in their respective environments. After T generations of adaptation, these parental populations interbreed to form hybrids. Panel (b) illustrates the process of adaptation in my simulations, wherein two populations (red and blue arrows connect the mean phenotype every 200 generations) independently adapt to specified optima (red and blue stars; behind arrows in [b] but visible in [c]). Concentric circles represent fitness contours around the two optima. The ancestor state is indicated by the grey dot, with the angle of divergence, θ, shown between the two axes of selection (red and blue dashed lines; angle shown is approximately 90°). Panel (c) illustrates the segregation variance in a group of hybrids. Individual hybrids (purple points) that are near an optimum have high fitness when measured in that environment. The black line is the line connecting parental optima—variance along this line can increase mean hybrid fitness whereas variance orthogonal to this line is deleterious.

2.2.1 Genotype to phenotype

The phenotype of a haploid individual is represented by an m-dimensional vector, $z = [z_1, z_2, \ldots, z_m]$, with m being the number of uncorrelated 'traits' or phenotypic 'dimensions' (for further discussion of dimensionality, see Orr 2000 & Tenaillon 2014). Each trait value, z_i, is determined by the summed effects of alleles at all underlying loci (i.e., mutations act additively to determine the phenotype), which are initially fixed for alleles with an effect of 0 on all m traits. I primarily present results from simulations with five phenotypic dimensions (m = 5) in the main text. Results for alternative parameter combinations can be found in the supplementary figures (Figs. A.1–A.7).

2.2.2 Life-cycle

I model a Wright-Fisher population (Fisher 1930; Wright 1931) with haploid selection. Fitness is a Gaussian function that depends on the Euclidean distance between an individual's phenotype and the phenotypic optimum, ||**z** - **o**||, and the strength of selection, σ (e.g., Lande 1979):

$$W = \exp(-\sigma||\mathbf{z} - \mathbf{o}||^2/2) \tag{2.1}$$

Table 2.1: **Description of parameters and parameter values in parental populations for simulations presented in the main text.**

Parameter	Value
α, mutation size SD in each dimension	0.1
d, distance between ancestral and parental phenotypic optima	1
N, number of haploid individuals	1000
m, number of traits, or 'dimensionality'	5
n, initial number of segregating loci	0 (DNM); 100 (DNM & SGV)*
μ, probability an individual acquires a new mutation	0.001
σ, strength of selection	1
θ, angle of divergence (°)	$0° \leq \theta \leq 180°$

*DNM: *de novo* mutation; SGV: standing genetic variation

(My qualitative conclusions are robust to alternative assumptions about fitness functions [see Fig. A.8)]. After selection, *N* haploid parents are randomly sampled with replacement from a multinomial distribution with probabilities proportional to their fitness, *W*. Parents then randomly mate and produce two haploid offspring per pair, with free recombination between all loci. With probability μ, an offspring gains a mutation; I assume an effectively infinite number of loci such that all mutations arise at a previously unmutated locus ('infinite-sites' *sensu* Kimura 1969). Mutational effects are drawn from a multivariate normal distribution ('continuum-of-alleles' *sensu* Kimura 1965), with a mean of 0 and an SD of α in all *m* traits and no correlations among traits (i.e., universal pleiotropy).

Generating ancestral standing genetic variation

To generate ancestral standing variation, I conducted burn-in simulations of a large ancestral population (N_{anc} = 10,000) under stabilizing selection (σ_{anc} = 0.01) at the origin (σ_{anc} = [0, 0, ... 0]) for 100,000 generations. All other parameters in the ancestor (e.g., mutation rate) were identical to those of parental populations (Table 2.1). This parameter combination facilitates the accumulation of appreciable standing variation (see Fig. A.9), but my general conclusions hold if the ancestor is under much stronger selection (σ_{anc} = 1) that puts it into the multivariate 'House-of-Cards' regime (Turelli 1985; see Fig. A.10).

Ancestral populations reached mutation-selection-drift balance such that the rate of acquisition of new mutations was balanced by the rate of loss of mutations that arose in earlier generations (Fig. A.9A). Both the mean frequency of derived alleles and the phenotypic (genotypic) variance were stable (Fig. A.9B), as has been found in other models of phenotypes under stabilizing selection (e.g., Barton 1989). Segregating derived alleles were all at unique loci by assumption—that is, each polymorphic locus has exactly two alleles and each derived allele can be traced back to a single mutation event. In addition, segregating derived alleles were at low frequency in the ancestral population (see Fig. A.9D for the site frequency spectrum). High derived allele frequencies and fixation are sometimes reached by drift when mutations have nearly-neutral selective coefficients and by positive selection when mutations compensate for deleterious alleles that have risen to high frequency by drift (Hartl, 2002; Orr, 2005).

2.2.3 Adaptation to a new environment

In simulations with standing genetic variation, a parental population was established by first randomly choosing n polymorphic loci in the ancestor (see Fig. A.11 for effect of n on genetic parallelism and segregation variance). Each parental individual received the mutant (i.e., 'derived') allele at each of these n loci with a probability equal to the allele's frequency in the ancestor. Loci fixed in the ancestral population were also fixed in the parental population but were not considered when quantifying parallelism. This admittedly artificial sampling procedure allowed us more control over the amount of standing genetic variation across simulations with different parameter values. Further control was achieved by making the second parental population initially identical to the first, so that each possessed the exact same collection of genotypes and there were therefore no founder effects. Populations adapted from only new (i.e., *de novo*) mutation when $n = 0$. Within each parameter combination, I began each replicate simulation from a unique realization of the ancestor (i.e., distinct burn-in). After initialization, parental populations adapted to their respective phenotypic optima without inter-population gene flow (Fig. 2.1B), and adaptation proceeded via natural selection on ancestral standing variation (if $n > 0$) and new mutation simultaneously.

Two properties of the new phenotypic optima are key. The first is the Euclidean distance between each optimum and the origin, d (assumed the same for both parental populations for simulations presented in main text). More distant optima yield a greater amount of genetic and phenotypic change. In the main text I set $d = 1$, which is equivalent to 10 times the SD of mutation effect size (α). The second key feature of the new optima is the angle of divergence, θ, separating vectors that originate at the origin and each pass through one of the parental optima (dashed lines in Fig. 2.1B). Angle is used to quantify the difference in the direction of selection from parallel ($\theta = 0°$) to divergent ($\theta = 180°$) and is explicitly invoked in most empirical metrics that quantify phenotypic parallelism (Bolnick et al., 2018). The value of θ is what determines the mean phenotypic differences that evolve between parental populations in my simulations (because d is held constant).

I ended the adaptation phase of simulations after T generations ($T = 2000$ in the main text), at which time all populations had reached their phenotypic optima (Fig. A.12A) and mutation-selection-drift balance (Fig. A.12B). An unavoidable and important effect of standing variation is that it quickens adaptation because populations do not have to wait for beneficial alleles to arise (Barrett et al., 2008). In my model and others like it (e.g., Barton 2001; Chevin et al. 2014), reproductive isolation evolves rapidly during the initial stages of adaptation. After populations reach their respective phenotypic optima, genetic divergence accumulates slowly at a rate proportional to the mutation rate (Barton, 1989, 2001; Chevin et al., 2014). Therefore, my results reflect quasi-equilibrium conditions rather than transient states and are unaffected by standing variation's influence on the speed of adaptation.

2.2.4 Quantification of genetic parallelism and hybrid segregation variance & fitness

After the adaptation phase of simulations had ended, I calculated the proportion of alleles that fixed in both populations (i.e., genetic parallelism). To quantify parallel genetic evolution between parental populations, I first determined the number of alleles that fixed in each population (f_1 & f_2) and the number of alleles that

fixed in both populations ($f_{1,2}$). I then calculated my metric of 'genetic parallelism' as:

$$P_{\mathrm{g}} = \frac{1}{2}\left(\frac{f_1}{f_{1,2}} + \frac{f_2}{f_{1,2}}\right) \tag{2.2}$$

Values of 1 indicate complete genetic parallelism (i.e., all alleles that fixed were fixed in both populations) and values of 0 indicate complete genetic non-parallelism (i.e., no allele fixed in both populations). I use this metric because of its ease of interpretation and note that it is highly correlated with other metrics of genetic divergence between populations (e.g., F_{ST}; Fig. A.13). I present some results with this metric scaled to a forced minimum of 0 and a forced maximum of 1 in order to facilitate comparison of simulations conducted with different parameters.

To create inter-population hybrids, I then randomly sampled 100 individuals from each population without replacement. Each sampled individual was paired with an individual from the other population to form 100 unique inter-population mating pairs. Every inter-population mating pair then produced one recombinant haploid F1 hybrid for a total of 100 potentially unique hybrids.

After forming hybrids I quantified their phenotypic variation—the net segregation variance (Wright, 1968; Slatkin and Lande, 1994)—calculated here as the mean phenotypic variance across all m traits. I present analyses of individual axes of variance where relevant. Higher segregation variance results when parents are differentiated by a greater number of alternative alleles (holding effect size constant) or alleles of individually-larger effect (holding number of alleles constant) (Castle, 1921; Slatkin and Lande, 1994; Chevin et al., 2014). Segregation variance captures the phenotypic consequences of hybridization and has a direct impact on fitness whereas genetic (non)parallelism is only indirectly related to fitness. Phenotypic variance in parental populations (i.e., before hybridization) is near zero and does not differ between populations founded with vs. without standing variation nor does it depend on the initial distance to the optimum (d; Fig. A.12C). Such low variance is expected because my simulations have fixed and frequency-independent optima, no migration, and parameter values corresponding to strong selection and relatively weak mutation ('house-of-cards'; Turelli 1984, 1985).

An individual hybrid's fitness in a given parental environment was calculated from its phenotype in the same manner as the fitness of parental populations (Fig. 2.1C). I determined the fitness (Eqn. 2.1) of each hybrid in both parental environments and recorded its fitness as the larger of the two values. This can be imagined as, for example, giving the hybrid a choice of alternative host-plants (see Drès and Mallet 2002) where it always chooses the host on which it has higher performance. My fitness metric reflects what is traditionally recognized as 'extrinsic' postzygotic isolation (Coyne and Orr, 2004), and explicitly considers environment-specific epistasis for fitness (Bateson 1909; Dobzhansky 1937; Muller 1942; Chevin et al. 2014; Fraïsse et al. 2016; see also Arnegard et al. 2014; Schumer et al. 2014; Ono et al. 2017 for discussion of environment-specific hybrid incompatibilities). I consider my model to be one of 'extrinsic' rather than 'intrinsic' isolation because I do not consider traits such as gamete viability, which experiences environment-independent selection. Rather, traits in my model are best imagined to be more akin to ecologically-relevant traits like beak depth, under stabilizing selection with optima that depend on the environment. Because hybrids are recombinant, hybrid fitness reflects both the effects of displacement of the mean phenotype from the optimum (the 'lag' load) and what in diploids is known as hybrid breakdown (Burton et al., 2006). I

report hybrid fitness relative to the parents for each individual simulation, calculated as: [mean fitness of hybrids] / [mean fitness of parents].

2.3 Results

2.3.1 Genetic parallelism and phenotypic segregation variance

I first investigated how genetic parallelism between two populations—the fraction of alleles that fixed in both populations vs. were unique to a single population—changes with the angle of divergence (θ) when adaptation is from standing variation. Genetic parallelism is highest under completely parallel natural selection (θ = 0°) and rapidly decreases toward its minimum value as θ increases (dark green line and points in Fig. 2.2A; see black line for visual comparison of deviation from linearity). This rapid decrease in genetic parallelism also occurs when the phenotypic distance between parental optima is used as the independent variable instead of θ, although with my parameters non-linearity is only appreciable in higher dimensions (see Fig. A.14). There is considerable variation in genetic parallelism among simulation runs even when populations adapt to identical environments, which results from stochastic processes in each run. For example, alleles are lost due to drift, populations fix weakly deleterious alleles or different de novo mutations, and populations fix alternative alleles from the standing variation early in the simulations, which affects the selection coefficients of all other alleles in later generations (Chevin and Hospital, 2008). Genetic parallelism never decreases to zero even under completely divergent selection (θ = 180°), indicating that populations fix some deleterious alleles. My conclusion that genetic parallelism rapidly decreases with θ is generally robust to variation in population size and selection strength, except for when small populations are under weak selection (Fig. A.1), likely due to an overwhelming effect of drift (see Fig. A.15 for divergence between populations due to drift alone at various population sizes).

The changes in segregation variance generally mirror patterns of genetic parallelism (Fig. 2.2B). With standing variation, segregation variance is low under parallel selection and rapidly increases with θ. Chevin et al. (2014) found that segregation variance (proportional to their 'variance load') does not depend on θ, but in contrast to my model they did not permit genetic parallelism. When there is no standing variation, segregation variance is not affected by the angle of divergence (light green line and points in Fig. 2.2C; linear model slope $\pm$ 1 SE: $4.9 \times 10^{-7} \pm 5.8 \times 10^{-6}$, in agreement with the findings of Chevin et al. 2014; their Fig. 2). At large angles, segregation variance is greater when populations adapt from standing variation than when they adapt from new mutation alone, and the magnitude of this difference increases with dimensionality (see Fig. A.16).

Genetic parallelism decreases with θ (and segregation variance increases) because the fraction of alleles that are beneficial in both parental populations declines as θ increases. For a given population, beneficial alleles bring populations closer to the middle of a hypersphere centred at the phenotypic optimum (the geometric model [Fisher 1930]; see cartoon inset of Fig. 2.3A). Considering two populations, each with its own hypersphere, a given allele is beneficial in both—and thus could fix in parallel via positive natural selection—if it brings a population's mean phenotype into the region where the two hyperspheres overlap (purple region in Fig. 2.3A inset). The size of this region of overlap decreases rapidly with θ (Fig. 2.3A;

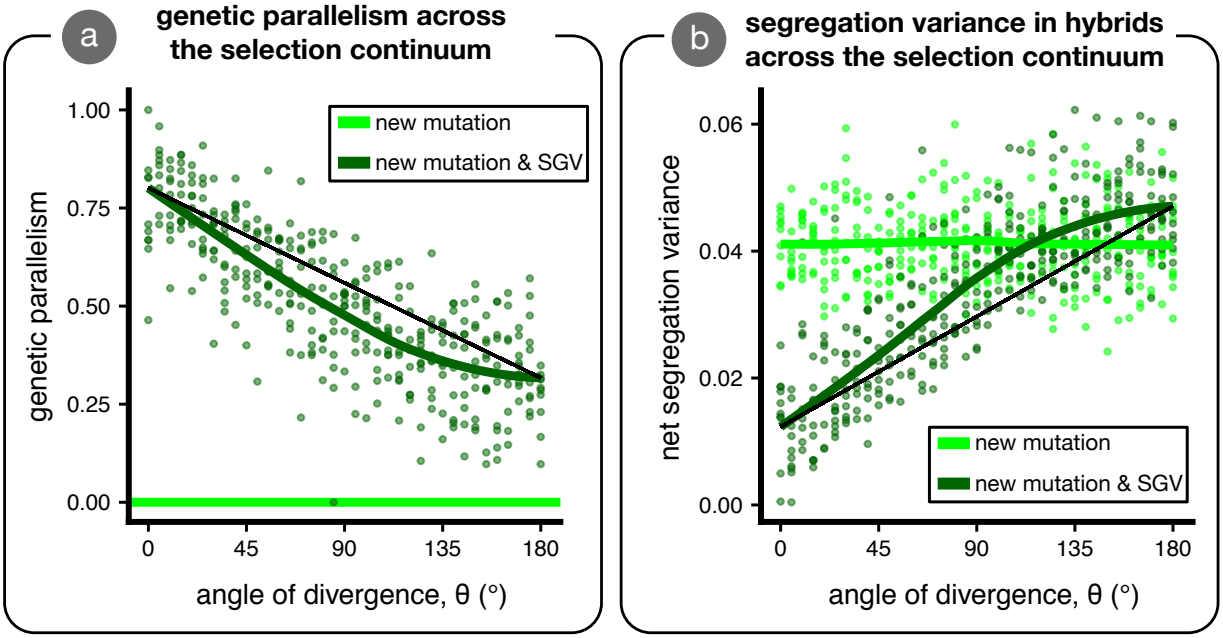


Figure 2.2: **Genetic parallelism and phenotypic segregation variance.** Parental populations adapted from either new mutation only (light green) or from a combination of new mutation and standing genetic variation (SGV) (dark green). Panel (a) shows the proportion of alleles that fixed in both populations vs. were unique to a single population (genetic parallelism; equation 2). The thin black line connects the fit at $\theta = 0°$ to the fit at $\theta = 180°$ and is shown only to facilitate visualization of the non-linearity. Panel (b) is similar to (a), except with the net segregation variance in hybrids as the dependent variable. Plotted are the results from 10 replicate simulations for each of 37 angles of divergence ($d = 1$). Green lines are loess fits.

see Appendix A.1 for mathematical details), and therefore so does the fraction of alleles present as standing variation that are beneficial in both populations. The rate of decrease of overlap is faster with greater dimensionality (compare solid line to dashed lines in Fig. 2.3A) but—perhaps surprisingly—does not depend on the distance to the optima (d; if $d_1 = d_2 = d$) and is not expected to change over the course of an 'adaptive walk' (*sensu* Orr 1998; see Appendix for detailed explanation). Briefly, this is because adaptation's effect is to shrink the radii of the hyperspheres (at roughly equivalent rates in the two populations if adaptation proceeds relatively deterministically). Thus, because the fraction of overlap (Eq. A1) does not depend on the radii of the hyperspheres, the fraction of overlap is expected to remain constant throughout adaptation. Simulations conducted for four different dimensionalities ($m = 2, 5, 10, 25$) qualitatively capture the predicted pattern of decreasing parallelism with increasing dimensionality (Fig. 2.3B), although drift, a limited supply of standing variation, and run-specific epistasis (etc.) contribute to differences between hypersphere overlap and genetic parallelism.

I also modelled an alternative case in which θ is held constant but populations differ in the distance to their respective optima (i.e., different vector 'lengths' rather than 'angles' *sensu* Bolnick et al., [2018]). Even if selection is completely parallel (i.e., $\theta = 0°$), if the distance between the ancestral phenotype and the phenotypic optimum of population 2 is twice that of the ancestor–optimum distance for population 1 (i.e., d_2

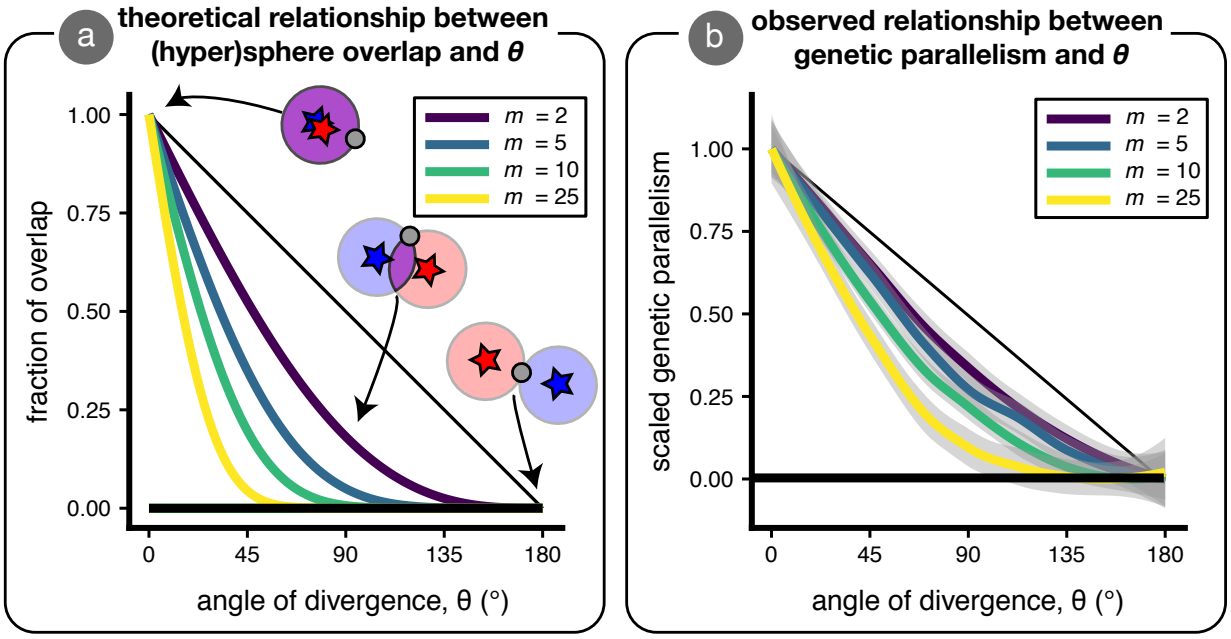


Figure 2.3: **The relationship between trait dimensionality (*m*) and genetic parallelism.** Panel (a) is an analytical result that depicts the relationship between θ and the fraction of overlap between two (hyper)spheres for four different dimensionalities (see equation A1). In the inset cartoon, mutations that bring the phenotype into the red and blue regions are initially beneficial only in the 'red' or 'blue' environments, while mutations that bring the phenotype into the purple region are beneficial in both environments. The horizontal black line is set at $y = 0$, where there is no overlap. Panel (b) is a proof-of-concept figure showing loess fits of simulation results with 95 % confidence intervals. Within a dimensionality, parallelism is scaled between 0 (minimum value of loess fit) and 1 (maximum value of loess fit). Simulations were conducted with strong natural selection ($\sigma = 10$) to minimize the effect of drift. (See Fig. A.19) for a similar result except with segregation variance on the y-axis.)

$= 2d_1$), less than 5 % of the alleles beneficial to population 2 are also beneficial to population 1 (for $m = 5$; see Fig. A.17). This result indicates that differences in vector lengths are important to consider—in addition to angles—for reducing the extent of genetic parallelism.

2.3.2 Hybrid fitness

In this section, I evaluate the effect of standing variation on hybrid fitness across the continuum from parallel to divergent natural selection. The most readily observable pattern is that the mean relative fitness of hybrids is lower under divergent selection than under parallel selection regardless of whether adaptation proceeds with standing variation (Fig. 2.4A). This pattern occurs because the hybrid mean phenotype is increasingly distant from either parental optimum as θ increases. In Figure 2.4A, I plot the fitness of the hybrid mean phenotype (representing the 'lag' load) as a thin black line.

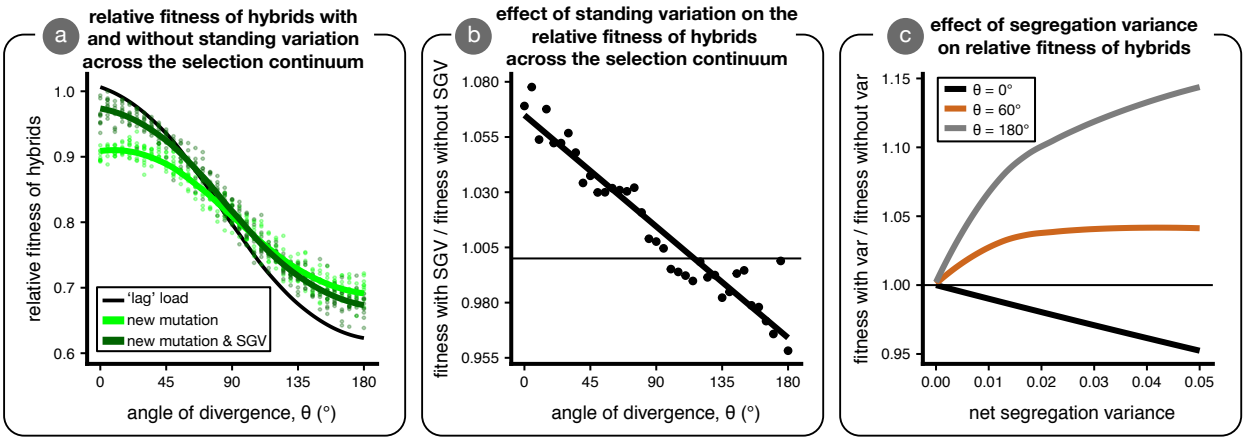


Figure 2.4: **The effect of standing variation on mean hybrid fitness.** Panel (a) shows the mean relative fitness of hybrids—as compared to parents—across environments in simulations initiated without (light green) and with (dark green) ancestral standing genetic variation. The thin black line represents the mean relative fitness of hybrids due only to the deviation of the observed mean phenotype from an optimum ('lag' load) and is close to 1 when the hybrid mean phenotype is on the optimum. Panel (b) shows the effect of standing variation on mean relative hybrid fitness (the ratio of values for dark / light green lines in panel [a]); the horizontal line shows where there is no effect of standing variation on relative mean hybrid fitness. Panel (c) is an analytical result that illustrates the relationship between segregation variance and mean hybrid fitness for three angles of divergence (black, $\theta = 0°$; brown, $\theta = 60°$; grey, $\theta = 180°$) when the hybrid phenotype is multivariate normal with a mean exactly in between the two parental optima and equal variance in all phenotypic dimensions (no covariance). Hybrid fitness is plotted for each angle relative to the case of no variance; the horizontal line indicates when segregation variance has no effect on hybrid fitness.

Compared to when adaptation is from new mutation, adaptation from standing variation improves mean hybrid fitness when parental populations adapt to similar optima but reduces hybrid fitness when parents undergo divergent adaptation (Fig. 2.4B). This pattern is caused by environment-specific effects of segregation variance on mean hybrid fitness (Fig. 2.4C, Fig. 2.5). When the hybrid phenotype distribution is centred at the phenotypic optimum—as it is under parallel selection ($\theta = 0°$)—segregation variance is universally deleterious. When parental populations adapt to identical optima from only new mutation, hybrids vary considerably around the parental optimum and thus have relatively low mean fitness. When populations have access to a common pool of standing variation, parallel genetic evolution leads to lower segregation variance around the optimum and therefore higher mean fitness under parallel selection compared to when populations adapt from only new mutation (Fig. 2.5A; see Fig. A.18 for similar results but for maximum hybrid fitness instead of mean).

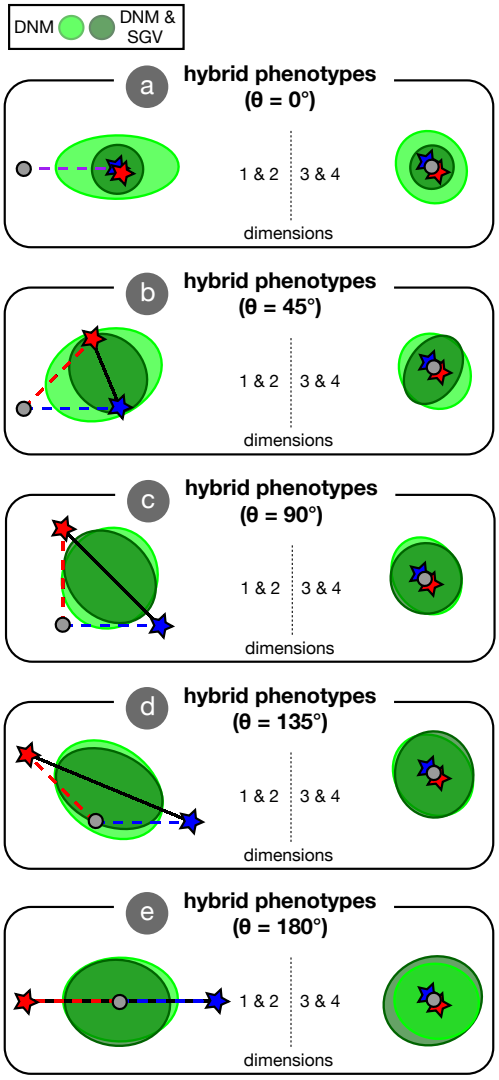


Figure 2.5: **The effect of standing variation on the distribution of hybrid phenotypes.** I plot ellipses containing 95 % of hybrid phenotypes for five angles of divergence (θ) evenly spaced along the continuum of (a) completely parallel ($\theta = 0°$) to (e) completely divergent ($\theta = 180°$) selection. Separate ellipses are shown for simulations where populations adapted from only new mutation (light green; DNM) or both new mutation and standing genetic variation (dark green; DNM & SGV). Each ellipse is fit to 1000 hybrids resulting from 10 replicate simulations. Parental optima are depicted as stars and the origin (ancestral optimum) is shown as a grey dot. The left side of each panel shows the first two trait dimensions—the only dimensions in which the optima might differ. The right side of each panel shows the third and fourth dimensions—both of which are under stabilizing selection for a phenotype identical to the ancestral phenotype. The axes of selection connect the origin and optima (dashed red and blue lines) and I also show the axis connecting parental optima as a solid black line. Ellipse plotting order is reversed on the right side of panel (e) to facilitate visualization.

At large angles of divergence, adaptation from standing variation reduces hybrid fitness compared to when adaptation is from only new mutation. The reasons for this are twofold. First, since I allow hybrids to 'choose' their environment (measuring their fitness in the parental environment they are better adapted to), at larger angles hybrids increasingly fall into a 'fitness valley'. In this case some variation along the

axis of divergence can be beneficial (see Fig. 2.4C). Second, since fitness in either environment is a Gaussian function, variation becomes beneficial when the mean is far from the optimum (by Jensen's inequality), even when considering fitness in only a single environment. This result is robust to variation in parameter values (see Figs. A.4–A.6), except when selection is very weak in small populations. There are appreciable differences in patterns of phenotypic variation in hybrids when their parents adapt with standing variation vs. when adaptation is from new mutation alone (Fig. 2.5). Only phenotypic variation along the axis connecting parental optima (black line connecting stars in Fig. 2.5) is beneficial, whereas variation along orthogonal axes is deleterious. When $\theta = 180°$, hybrids have reduced variation along the axis connecting parental optima and slightly more variation along all other axes (see Fig. 2.5E). Thus, maladaptive segregation variance—resulting from cryptic genetic differences between parental populations revealed only after hybridization—reduces hybrid fitness under large angles of divergence.

Why does adaptation from standing variation alter patterns of phenotypic segregation variance in hybrids? As discussed above, adaptation from standing genetic variation reduces segregation variance under parallel selection because parents fix the same alleles that therefore do not segregate in hybrids. Populations adapting from standing variation also fix a greater number (Fig. 2.6A) of smaller effect alleles (Fig. 2.6B) than populations evolving without standing variation. Fixation of smaller-effect alleles likely occurs under adaptation from standing variation because stabilizing selection in the ancestor effectively removes large-effect alleles from the standing variation (Fig. A.9) and because weakly beneficial alleles have a higher probability of fixation when present in standing variation compared to if they arose *de novo* (Orr and Betancourt, 2001; Hermisson and Pennings, 2005; Matuszewski et al., 2015). This latter effect seemed to allow alleles with more deleterious pleiotropic effects to fix during adaptation from standing variation than when adaptation was from new mutation alone (Fig. 2.6C). That is, populations initiated with standing genetic variation used alleles with proportionally larger pleiotropic side-effects. I quantified pleiotropy by taking the ratio of the mean effect size of fixed alleles along the axis of selection in parents (red or blue dashed lines in Fig. 2.5) vs. the mean effect size across all orthogonal axes, termed the 'efficiency index'. Values of 1 (horizontal line in Fig. 2.6C) imply that an allele had an equivalent effect along the axis of selection as on orthogonal axes. Increasingly positive values reflect alleles that take a population to the optimum more 'efficiently' (i.e., directly along the dashed blue or red line in Fig. 2.5). Together, these results indicate that adaptive walks from standing variation in my simulations involved more—slightly smaller—steps and are more 'meandering' than adaptive walks from new mutation alone, which use fewer—slightly larger—and more direct steps (but see Ralph and Coop 2015). These differences in the properties of alleles fixed in simulations initiated with vs. without standing variation contribute to the patterns of phenotypic segregation variance that ultimately determine the fitness of hybrids.

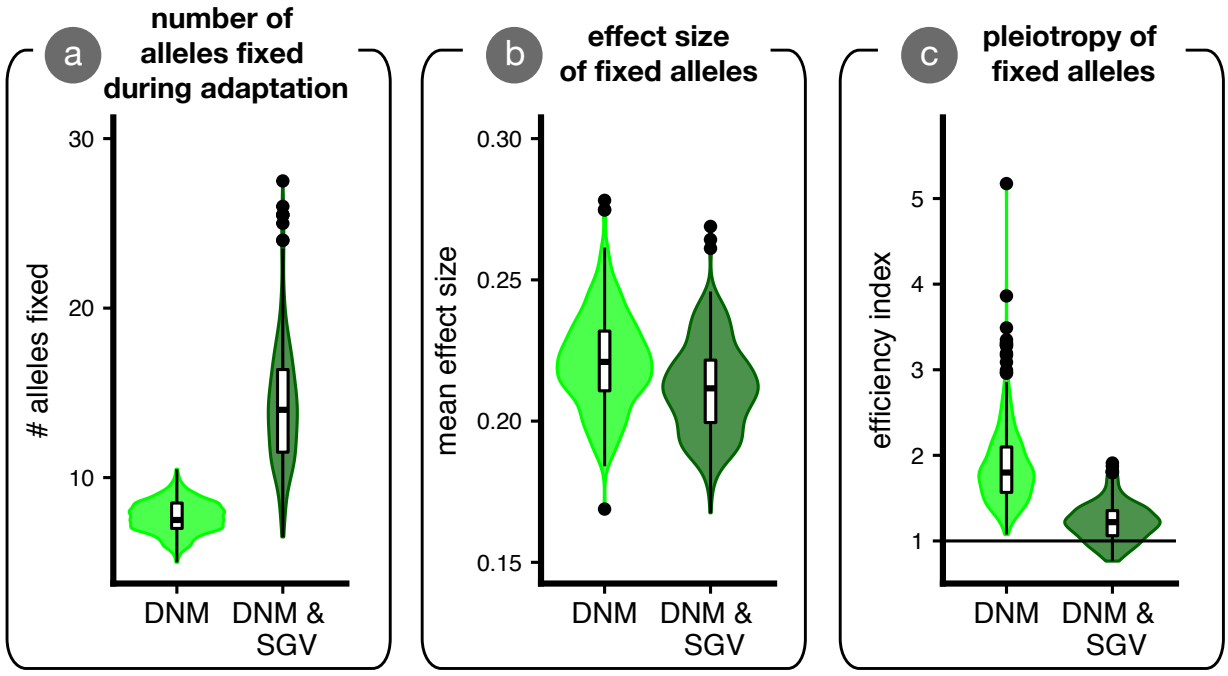


Figure 2.6: **Properties of alleles fixed during adaptation.** I show results from simulations where parental populations adapted from only de novo mutation (light green; DNM) vs. adaptation from standing variation and new mutation (dark green; DNM & SGV). Each replicate simulation contributed one data point to the plot. Panel (a) shows the average number of alleles fixed during adaptation. Panel (b) shows the average effect size (Euclidean length of mutation vector) of alleles fixed during adaptation. Panel (c) shows the allele 'efficiency index', which plots the ratio of a fixed mutations' effect size in the direction of selection vs. orthogonal directions. Values of 1 (horizontal line) are equally balanced in these directions, and mutations are more 'efficient' (i.e., they point more directly at the optimum) as this index increases. Statistical tests confirm all differences as highly significant (not shown).

2.4 Discussion

In this study I investigated parallel genetic evolution and progress toward speciation under adaptation from standing variation. I characterized how the extent of genetic parallelism from standing variation changes with the angle of divergence between parental optima, then illustrated how adaptation from standing variation affects hybrid fitness under various forms of natural selection. Here, I highlight my key findings, predictions for empirical systems, and suggestions for future work.

2.4.1 Key predictions and possible tests

The first principal finding of my study is that the degree of genetic parallelism rapidly declines as the angle of divergence increases from parallel toward divergent, especially when a large number of traits affect fitness. Practically, this means that the extent of genetic parallelism should decline quickly with phenotypic divergence. It is possible to test this prediction in natural or experimental populations using techniques such as 'Phenotypic Change Vector Analysis', which estimates important parameters such as the angle between

the vectors and/or the difference in their magnitudes (Bolnick et al., 2018). (Of course, phenotypic measurements are imperfect and typically non-comprehensive, and accordingly estimates of inter-population divergence are necessarily made with some error.) Natural systems exhibiting repeated instances of easily-quantified phenotypic divergence (Oke et al., 2017; Stuart et al., 2017) are particularly amenable to this approach. Given that phenotypic and genetic parallelism are not linearly related (Fig. A.14), I suggest that analytical predictions about the extent of genetic parallelism ought to be considered when generating predictions for empirical systems. I also note that studies quantifying genetic parallelism (Jones et al., 2012*b*) typically do not quantify non-parallel changes. In order to test the predictions about genetic parallelism, it will be necessary for future studies to measure both the number of parallel genetic changes (numerator) and the total number of genetic changes (denominator) in pairs of populations being compared (Alves et al., 2019).

My second principal finding is that—relative to when adaptation is from only *de novo* mutation—adaptation from standing genetic variation improves the mean fitness of hybrids under parallel natural selection, has little effect at intermediate angles of divergence, and reduces mean hybrid fitness under completely divergent selection. Practically, this indicates that adaptation from standing variation works against 'mutation-order' speciation and facilitates 'ecological' speciation (Schluter, 2009; Schluter and Conte, 2009). This hypothesis could be tested most readily in experimental systems where the amount of ancestral standing variation can be easily manipulated, and where interpopulation hybrids can easily be generated to have their fitness measured in parental environments. It would also be worthwhile to empirically test whether alleles fixed from standing variation are indeed more pleiotropic than alleles fixed from *de novo* mutation, as predicted by my simulations.

I emphasize that the mechanism through which adaptation from standing variation affects hybrid fitness (relative to adaptation from *de novo* mutation) differs between simulations where populations adapted under parallel vs. divergent selection. Under parallel selection, standing variation's effect on hybrid fitness is caused largely by parallel genetic evolution and therefore adaptation from standing variation is most likely to have an effect if populations adapting in parallel are founded with the same standing variation. Under divergent selection, standing variation's effect on hybrid fitness is not caused by genetic parallelism but rather by cryptic genetic differences—cryptic because they don't reveal themselves until after hybridization—that evolve between parental populations. Therefore, my predictions about the effect of adaptation from standing variation on hybrid fitness under divergent selection should hold regardless of whether populations have the same or different initial standing variation. A simple prediction—testable theoretically and empirically—resulting from my study is that founder effects should have a greater effect on hybrid fitness under parallel selection than under divergent selection.

2.4.2 Alternative sources of standing variation

My model addresses the case of adaptation from a pool of standing genetic variation at mutation-selection-drift balance. This framework does not address cases of adaptation where standing variation is generated from other sources. For example, in threespine stickleback, the marine ancestral form maintains standing variation for freshwater-adapted alleles in a balance between migration of alleles from freshwater populations

and negative selection in the sea (the 'transporter' hypothesis; Schluter and Conte 2009; Nelson and Cresko 2018. In this case, the pool of standing variation is enriched for alleles that have already swept to high frequencies in freshwater populations—that is, they are 'pre-tested' by selection. Scenarios such as this are especially likely to lead to genetic parallelism (Schluter and Conte, 2009). The extent to which adaptation from standing variation proceeds via the sorting of naïve alleles (as in my model) vs pre-tested alleles (as in the transporter model) is unresolved.

2.4.3 Possible extensions

Some of my conclusions will change under alternative assumptions. Some assumptions—for example a lack of recurrent *de novo* mutation or gene flow—reduce the extent of genetic parallelism (Nosil and Flaxman, 2011; Anderson and Harmon, 2014; Ralph and Coop, 2015; Barghi et al., 2019). I also assumed universal pleiotropy, and future work examining the effect of modularity on my results—especially on changes in parallelism with the angle of divergence—would be valuable. In addition, I considered only haploid selection, had only additive effects of alleles on phenotypes, and assumed that the sole fitness optima available to hybrids are those to which the two parents are adapted. My analytical results also ignore variation in the probability that particular mutations arise and fix (or are present as standing variation). Extending my approach to integrate the distribution of fitness effects of new mutations (Eyre-Walker and Keightley, 2007), the correlation of selection coefficients across environments (Kassen, 2014; Martin and Lenormand, 2015), and existing theory on the probability of genetic parallelism from standing variation (MacPherson and Nuismer, 2017) will be valuable.

I also note that the only reproductive isolating barrier I considered was environment-specific postzygotic isolation. Postzygotic isolation can also be environment-independent, and such 'intrinsic' isolating barriers are correlated with genetic divergence between populations (Orr 1995; Matute et al. 2010; Moyle and Nakazato 2010; Wang et al. 2015). Therefore, my measure of genetic parallelism might be interpreted as being inversely proportional to the strength of intrinsic barriers. I also did not consider prezygotic barriers such as assortative mating (Gavrilets, 2004). Accordingly, my results might be most relevant for empirical systems where ecology-based postzygotic isolation has a primary role in the origin of species.

2.4.4 Concluding remarks

In this study I characterized patterns of genetic parallelism and progress toward speciation from standing variation in pairs of populations with quantitative differences in the direction of selection between them. My findings generate new hypotheses for empirical studies on genetic parallelism and speciation. As evolutionary biologists develop increasingly powerful tools for detecting parallel genetic adaptation in nature, it will be important to keep in mind that genetic parallelism could be less common than I might intuit from patterns of selection and phenotypic similarity. I have also shown that adaptation from standing variation is expected to weaken the strength of isolating barriers that evolve between populations subject to parallel natural selection. By contrast, adaptation from standing variation can facilitate the process of speciation via divergent natural selection (i.e., 'ecological' speciation), suggesting that adaptation from standing variation might have a role in adaptive radiation beyond simply increasing the rate of adaptation.

Chapter 3

Patterns, predictors, and consequences of dominance in hybrids[2]

3.1 Introduction

When divergent populations occur in sympatry, they might mate and form hybrids (Mallet 2005). If those hybrids are viable and fertile, whether they survive and reproduce depends on their ability to persist under prevailing ecological conditions. Because selection against hybrids limits gene flow between parents (Harrison 1993), understanding the mechanisms underlying hybrid performance in the field is key to understanding postzygotic isolation (Barton and Hewitt 1985; Gompert et al. 2017). Quantifying general patterns of phenotype expression in hybrids would clarify mechanisms of natural and sexual selection that act against hybrids. For example, if hybrids resemble one parent they could thrive in that parent's niche and readily back-cross (Mallet 1986). Alternatively, if hybrids are phenotypically intermediate for all traits, or possess mismatched trait combinations due to dominance in opposing directions (i.e., they resemble parent 1 for trait x, but parent 2 for trait y), they might be unable to survive and reproduce in the available niche space (Hatfield and Schluter 1999; Matsubayashi et al. 2010; Arnegard et al. 2014; Cooper et al. 2018). Currently, little is known about general patterns of trait expression in hybrids.

Previous synthetic studies investigating hybrid phenotypes have conflicting conclusions. Some authors suggest that hybrid intermediacy is the rule (Hubbs 1940, 1955) whereas others find that hybrids are better described as mosaics of parental and intermediate characters (Rieseberg and Ellstrand 1993). Such previous studies typically lacked a quantitative framework and/or focused on a single taxon (e.g., fish or plants), limiting our ability to arrive at general conclusions. In addition, previous studies of hybrid phenotype expression tend to use data from domesticated taxa, wherein dominance is often elevated compared to natural populations (Crnokrak and Roff 1995; Fisher 1931). Here, I use a geometric approach to quantify patterns of hybrid phenotypes across a broad range of wild (or recently-wild) plant and animal taxa in a way that is comparable across studies. By quantifying the 'parent-bias' across each pair of traits I determine the extent to which hybrids are intermediate or tend to resemble one parent more than the other. And by quantifying the 'mismatch' (also termed 'opposing dominance' [Matsubayashi et al. 2010; Nosil 2012]) I can determine the extent to which hybrids have mismatched combinations of divergent parental traits.

In this chapter, I systematically document patterns of phenotype expression in hybrids, investigate the possible predictors of these patterns, and use experimental data to explore the fitness consequences of trait

[2] A version of this chapter has been published as Thompson, K.A., Urquhart-Cronish, M., Whitney, K.D., Rieseberg, L.H., and Schluter, D. Patterns, predictors, and consequences of dominance in hybrids. *The American Naturalist* 197(3) (awaiting pagination).

interactions in the field. I first summarize the results of 198 studies that compared the phenotypes of hybrids and parents in a common environment and test whether features of a cross—such as the genetic distance between, or taxon of, the parents—are associated with dominance (section I). I then use data from an experimental planting of recombinant hybrid sunflowers to evaluate whether patterns of pairwise parent-bias and mismatch predict fitness in hypothesized directions (section II). My results provide insight into the mechanisms that might commonly underlie selection against hybrids in nature.

3.2 I: Patterns & predictors of dominance

3.2.1 Methods

In this section, I provide a brief summary of my methodology for collecting and analysing data on hybrid trait expression from the literature, and then describe the patterns evident in the data. A detailed explanation of all methods is given in the Supplementary Methods.

Systematic review of dominance patterns in F_1 hybrids

I conducted a systematic literature search and identified 198 studies from which I could collect data of at least one divergent phenotypic trait measured in two parent taxa and their F_1 hybrids in a common environment. I included studies that conducted crosses between wild-collected parental populations or laboratory populations with ≤ 10 generations of captivity. Crosses in the dataset are both intraspecific (43%) and interspecific (57%). Data from wild hybrids (i.e., not from controlled crosses) were only included if hybrids were genotyped to confirm hybrid status and generation. I aimed to include only traits with environment-dependent effects on fitness—traits plausibly under divergent selection between populations (stabilizing selection within populations), rather than directional selection in the same direction in both populations. Said another way, I attempted only to include traits that could be characterized as 'non-fitness' traits [Merilä and Sheldon 1999] or 'ordinary' traits [Orr and Betancourt 2001]). For example, traits such as 'embryo viability' are almost certainly under directional selection and were not included in my database. I excluded likely fitness components because developmental difficulties resulting from hybrid incompatibilities, or heterosis resulting from outbreeding, often affects such traits in hybrids (Coyne and Orr 2004). This choice to exclude fitness traits likely renders my analysis on dominance more conservative since hybrid breakdown or heterosis would manifest as a transgressive phenotype (see Stelkens and Seehausen 2009). By contrast, traits such as 'limb length' might have particular values best suited to some environments and genetic backgrounds—it is implausible that such traits would always be selected to a maximum or minimum value. Data from backcross (BC_1 only) and F_2 hybrids were collected when available, but were used in a previous publication to test a theoretical prediction about pleiotropy (Thompson 2020). The studies in my analysis spanned a range of taxa but included mostly vascular plants (approx. 34 %), vertebrates (approx. 30 %), and arthropods (approx. 30 %), with the few remaining studies using annelids (<1 %), echinoderms (<1 %), red algae (Rhodophyta; <1 %), and molluscs (approx. 2 %).

I restricted my dataset to putatively divergently-selected traits. I retained all traits for which the parents were > 1 phenotypic standard deviation (SD) apart, which was the case for 71.7 % of measured traits (using

Table 3.1: Hypothetical examples of possible F_1 trait values, and corresponding values for cross mean $d_{univariate}$, $d_{parent\text{-}bias}$, and $d_{mismatch}$ (parent phenotypes are scaled to [0, 0] and [1, 1]).

F_1 phenotype	mean $d_{univariate}$	$d_{parent\text{-}bias}$	$d_{mismatch}$
[0.5, 0.5]	0	0	0
[0, 1]	1	0	1
[1, 1]	1	1	0
[0.25, 0.75]	0.5	0	0.5
[0.5, 1]	0.5	0.5	0.5
[1.25, 1]	1.25	1.25	0.25

the smaller of the two parental SDs). I also retained traits for which the parents were < 1 SD apart but had statistically distinguishable phenotypes (*t*-test $P < 0.05$), which accounted for an additional 9.4 % of measured traits. The data for the remaining 18.9 % of traits for which I collected data were discarded (see Fig. B.1). In total, data used for the analysis of dominance patterns comes from 233 unique crosses.

After filtering traits, I converted all trait data that were published with a transformation applied (e.g., ln(x), $\sqrt{x}$) to their original measurement scale because expectations are not the same on a log or square-root scale as for raw units. This choice to analyse all traits in their original measurement units influences the dominance patterns because traits that are intermediate on the raw scale might be dominant on a log-scale, or vice-versa, but results greater comparability among traits and studies. I then put all traits in all studies on a common scale where one (arbitrarily determined) parent had a value of 0 for all traits and the other had a value of 1 (see Fig. 3.1). Under an expectation of additivity, an F_1 hybrid would have a trait value of 0.5 for all traits. Because I do not make any assumptions about which trait value is ancestral or derived, I cannot distinguish between dominance and recessivity. For example, a trait's degree of dominance is the same whether the hybrid trait value is 0.2 or 0.8. Importantly, however, hybrids having two traits with values 0.2 and 0.8 have an arithmetic mean phenotype of 0.5 but this hybrid is mismatched rather than intermediate. This failure of simple averaging highlights the need for geometry-based dominance metrics.

Quantifying dominance in F_1 hybrids

I quantified three metrics of dominance. Within a cross, each dominance metric was scaled such that values of 0 indicate no dominance, values of 1 indicate the maximum dominance without transgressing the parental trait range, and values greater than 1 result from transgression (see Table 3.1 for hypothetical hybrid phenotypes and corresponding dominance values for all three metrics).

The first dominance metric is 'univariate' dominance ($d_{univariate}$), which considers traits individually. $d_{univariate}$ measures the deviation of trait values from the additive expectation of 0.5, regardless of direction. For a single trait, this was calculated as:

$$d_{univariate} = 2(|z_i - 0.5|), \tag{3.1}$$

where z_i is the scaled mean phenotype of trait i. A $d_{univariate}$ value of 0 results when a trait is exactly

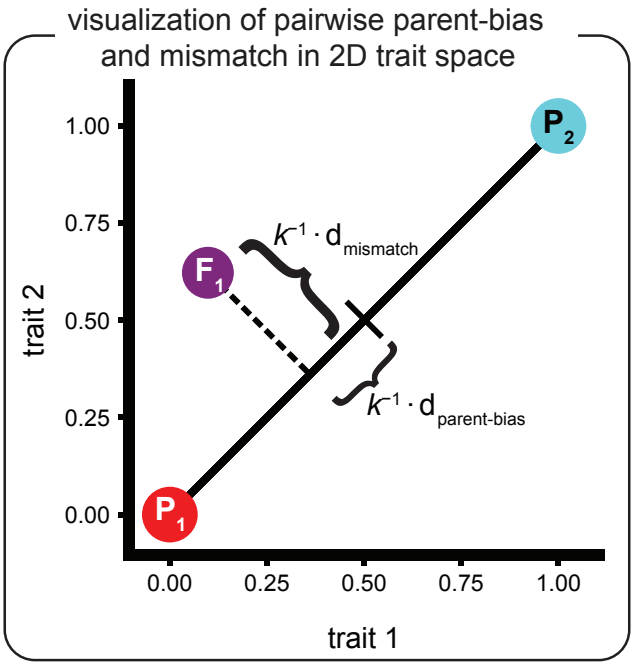


Figure 3.1: **Visual overview of how two-dimensional dominance metrics were calculated.** When studies contained two or more divergent traits, I calculated pairwise parent-bias ($d_{parent\text{-}bias}$) and mismatch ($d_{mismatch}$) of the hybrid phenotype (F_1) with respect to the line connecting the two parent phenotypes (P_1 & P_2 [note that which parent is called P_1 or P_2 is arbitrary). This procedure was repeated for every pair of traits. The scaling factor, k, renders the maximum value observed without transgression (i.e., $d_{mismatch}$ when F_1 trait values are [0, 1]; or pairwise $d_{parent\text{-}bias}$ when F_1 trait values are [0, 0]) equal to 1. For two traits, $k = \sqrt{2}$. Dominance values > 1 can result when traits are transgressive. In this hypothetical example, $d_{parent\text{-}bias}$ is approximately 0.25 and $d_{mismatch}$ is approximately 0.5; see Table 3.1 for other possible F_1 two-dimensional hybrid phenotypes and their corresponding dominance values.

intermediate ($z_i = 0.5$; the mean of the parental trait values, 0 and 1); a $d_{univariate}$ value of 0.5 results when the F_1's mean trait value is halfway between intermediate and that of one parent (i.e., $z_i = 0.25$); and a $d_{univariate}$ value of 1 results when the F_1 hybrid mean equals that of one of the parents (i.e., $z_i = 0$ or $z_i = 1$). Transgressive traits have $d_{univariate}$ values > 1. I averaged $d_{univariate}$ values across traits within each cross to obtain estimates of cross mean $d_{univariate}$.

The remaining two dominance metrics consider pairs of traits at a time and are calculated in two dimensions (see Fig. 3.1 for general overview and Fig. B.2 for examples from the dataset). I consider all *pairs* of traits, instead of all traits together, to increase the comparability of dominance values among studies measuring different numbers of traits. For crosses where three or more divergent traits were measured, I calculated two-dimensional dominance metrics for each trait pair and then took the mean of all pairwise estimates as the value for that cross.

The second metric of dominance is pairwise parent-bias ($d_{parent\text{-}bias}$), which captures deviation from bivariate intermediacy in the direction of either parent. Imagine a cross between two plant species, one of which has flowers that are narrow (mean width = z_1) and red (colour = z_2), representing the bivariate phenotype of [0, 0], and the other species has wide yellow flowers, represented by a bivariate phenotype

of [1, 1]). If their F_1 hybrid's standardized phenotype is [0, 1] (i.e., narrow yellow flowers), then the mean $d_{univariate}$ = 1 but pairwise $d_{parent\text{-}bias}$ = 0. A $d_{parent\text{-}bias}$ of 0 would also result if the F_1 hybrid was exactly intermediate between the parents (i.e., [0.5, 0.5]). $d_{parent\text{-}bias}$ has a minimum value of zero when dominance is equally strong in the direction of both parents and increases indefinitely as dominance increases in a manner that is biased toward one parent. For each pair of traits, I first determined the scalar projection, *b*, of the hybrid phenotype onto the line connecting parents (solid line in Fig. 3.1). This projection is calculated as:

$$b = \frac{z_1 + z_2}{k}, \tag{3.2}$$

where z_1 and z_2 are the hybrid values for trait 1 and 2. I then calculated pairwise parent-bias as:

$$d_{parent\text{-}bias} = k \cdot \left| \frac{k}{2} - b \right|, \tag{3.3}$$

where b is the scalar projection from eqn. 3.2, k is a scaling factor ($k = \sqrt{2}$) used to give a hybrid a phenotype with parental values for both traits (i.e., [0, 0] or [1, 1]) a $d_{parent\text{-}bias}$ value of 1. $d_{parent\text{-}bias}$ cannot exceed $d_{univariate}$.

The third and final metric of dominance is pairwise mismatch ($d_{mismatch}$), which captures the perpendicular distance between the mean hybrid phenotype and the line connecting parental mean phenotypes (Fig. 3.1). $d_{mismatch}$ has a minimum value of zero when the hybrid phenotype is on the line connecting parents (i.e., when both hybrid traits in the pair are equally displaced towards the same parent) and increases indefinitely as the variance in dominance among traits increases. Returning to the earlier example of a cross between plants with divergent floral traits, $d_{mismatch}$ values of 0 would characterize hybrids with phenotypes that are varying degrees of intermediate (e.g., [0.5, 0.5] or [0.75, 0.75]) or recover parental phenotypes [0, 0] or [1, 1]. A $d_{mismatch}$ value of 1 results when dominance is complete but in opposite directions [0, 1] or [1, 0], which corresponds to narrow yellow flowers or wide red flowers. For each pair of traits, I calculated mismatch as:

$$d_{mismatch} = k \cdot \sqrt{z_1^2 + z_2^2 - b^2}, \tag{3.4}$$

where, z_1 and z_2 are as in eqn. 3.1, and b and k are as in eqn. 3.3.

Evaluating patterns caused by sampling error

The above metrics of dominance, applied to data, are a product of both biology—net dominance effects of genes—and measurement and/or sampling error. Such error around an intermediate phenotype would appear as dominance because I calculate dominance as the difference between the observed mean phenotype and the mid-parent value. In addition, a given amount of error is more likely to result in high (e.g., $d > 1$) dominance estimates when the parents involved in a cross are phenotypically similar than when they are more divergent. It is therefore important to quantify the magnitude of dominance observed due to sampling error alone.

To quantify patterns of dominance due to sampling error, I generated 1000 simulated datasets that had an identical structure to the raw data but where hybrid mean phenotypes were replaced with means calculated from a simulated distribution. Specifically, I generated random normal vectors (using the `rnorm` function) for each trait measured in F_1 hybrids with a length equivalent to the number of hybrids measured by authors and the original trait SD, but an expected mean that was exactly intermediate between the parents (i.e., `rnorm(n =` n_{F_1}`, mean =` $\mu_{(P_1,P_2)}$`, sd =` SD_{F_1}`)`). I then took the mean of each random vector and replaced the observed hybrid mean with the simulated mean. The simulated mean can differ from strict intermediacy due only to sampling error. I calculated each of the three dominance metrics for each cross in all simulated datasets and compared the distribution of estimates to what I observed in the original data.

Testing possible predictors of dominance in F_1 hybrids

I explored several possible predictors of dominance motivated by previous results and theoretical predictions. For example, previous studies have determined that genetic distance between cross parents affects the frequency with which hybrid traits transgress the parent range (Stelkens and Seehausen 2009), a pattern that should be captured by my $d_{univariate}$ metric.

To determine if genetic distance affects my dominance metrics, I computed genetic distance using gene sequence data and tested whether it was associated with any metric of dominance. To maximise the number of crosses for which I could estimate genetic distance, I used cytochrome b for animals and the internal transcribed spacer I and II for plants. Because the species in my dataset can hybridize, it is possible that I might underestimate genetic divergence if there is hybridization and introgression in nature—this problem might be especially pronounced for the mitochondrial cytochrome b. I could not obtain sufficient nuclear data for animals, so the genetic distance data should be interpreted with this limitation in mind because mtDNA often seems to introgress more readily than nuclear genes (e.g., Bachtrog et al. 2006; Wang et al. 2020). Genetic distance was calculable for less than one quarter of all crosses, and only 3 intraspecific crosses, so I also compared dominance metrics between intraspecific and interspecific crosses—the underlying assumption being that genetic distance between parents is lower in the former compared to the latter.

Various taxon-specific reviews have arrived at different conclusions about the extent of dominance observed in hybrids. For example, Rieseberg and Ellstrand (1993) considered plants only and concluded that dominance is common in hybrids whereas Hubbs (1955) worked on fish and concluded that dominance in hybrids is rare. To test whether there might be variation in dominance between taxa, I built a phylogeny encompassing nearly all crosses in my dataset (Fig. B.3) and tested for phylogenetic signal in dominance metrics. I also tested whether there are differences in dominance between predefined taxonomic groups such as plants and animals.

Finally, I tested for parent-of-origin effects. If parent-of-origin effects are common and have some systematic basis, then hybrid trait values might, for example, tend to resemble the maternal parent more than the paternal; this is testable in the present dataset because many crosses (n = 96) were conducted in both directions.

3.2.2 Results

Patterns of dominance in F_1 hybrids

I used data gathered from the literature to generate estimates of dominance in F_1 hybrids. I first consider each trait individually, and calculate mean univariate dominance ($d_{univariate} \pm 1$ SE) for each unique cross in the dataset. Considering all cross mean $d_{univariate}$ estimates together, the mean $d_{univariate}$ for traits measured in F_1 hybrids was 0.79 ± 0.078 (Fig. 3.2A [see Fig. B.4 for the same figure with the x-axis extended]; median = 0.55), which suggests that the average trait is not intermediate but rather more than halfway between intermediate and parental. In approximately 20 % of crosses (and 20 % of individual traits), the mean $d_{univariate}$ was > 1, indicating transgression.

In addition to $d_{univariate}$, I calculated two complementary two-dimensional dominance metrics to investigate whether hybrids tend to be biased toward one parent over the other ($d_{parent\text{-}bias}$) or have mismatched combinations of divergent traits ($d_{mismatch}$). These metrics are different from $d_{univariate}$ because high single-trait dominance could either be in the same direction for both traits (leading to more parent-bias) or in opposite directions (leading to more mismatch). I find that the mean pairwise $d_{parent\text{-}bias}$ among crosses was 0.68 ± 0.01 (Fig. 3.2B; median = 0.44), implying that, for a given pair of traits, hybrids on average resemble one parent $\geq 68\%$ more than the other. The mean pairwise $d_{mismatch}$ was 0.60 ± 0.10 (Fig. 3.2C; median = 0.31), implying that the average hybrid is about 60% as mismatched as is maximally possible without transgression for a given pair of traits. Mismatch did not differ between pairs of traits that were both in the same category and pairs of traits from different categories ($F_{1,102.47} = 0.0199$, $P = 0.88$).

I generated simulated datasets to estimate the magnitude of dominance I would expect from sampling error alone. I find that the simulation-based estimates of all three dominance metrics were approximately one-third as large as what is observed in the real data, with little variation among replicate simulations (see Fig. B.6). These simulation results indicate that the majority of my signal is biological rather than caused by sampling error.

Predictors of dominance in F_1 hybrids

I next investigated whether dominance patterns in F_1 hybrids are associated with genetic distance and phylogeny. I found no significant associations between any metric of dominance and any metric of genetic distance (see detailed results in Figs. B.7–B.9). In addition, there was no evidence for phylogenetic signal in any dominance metrics (all $\lambda < 1 \times 10^{-5}$, all $P = 1$), and no difference in any dominance metrics in comparisons of major clades (Fig. B.9). I found that dominance is lower when the parental populations have larger differences in their phenotype coefficients of variation and greater when parents are more variable, although each of these factors explains less than 1 % of the variance in $d_{univariate}$ (see Fig. B.10). Trait type generally did not affect dominance, although chemical traits (e.g., pheromones) seemed to have higher dominance and transgression than all other trait types (Fig. B.11). Some caution is warranted here, however, because chemical traits were the least well-represented category in the data.

Because many crosses were conducted reciprocally (i.e., hybrid crosses were conducted with each parent species serving as dam), I could evaluate parent-of-origin effects on trait values. I found that 25.6 % of traits

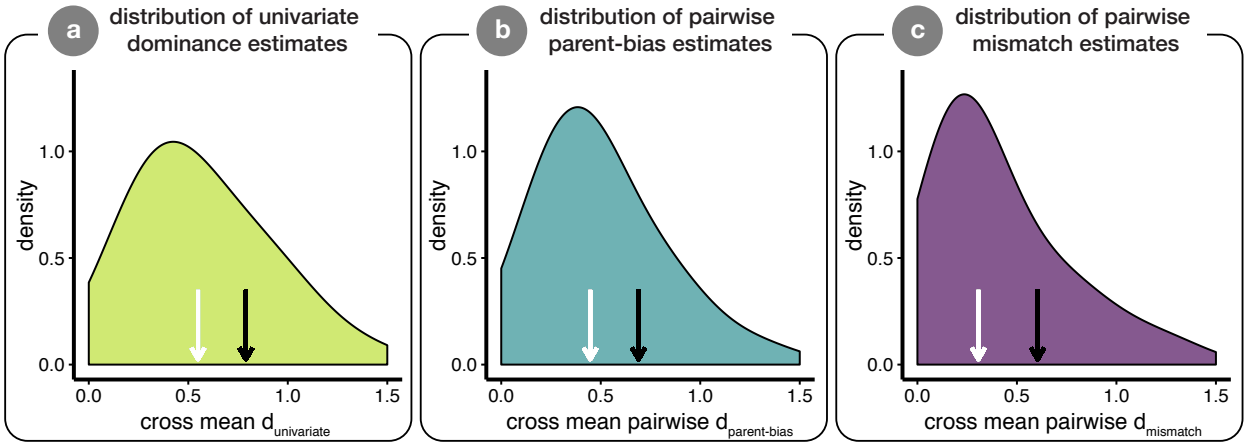


Figure 3.2: **Patterns of dominance in F_1 hybrids.** The density plots (y-axis standardized across panels) show the three main dominance metrics contained herein, with each cross contributing at most a single value per panel. For all three dominance metrics, values of 0 indicate no dominance, values of 1 indicate the maximum without transgression, and values > 1 reflect transgression. The x-axis is truncated at 1.5, but the means (black arrows) and medians (white arrows; values given in text) are calculated from the whole dataset (see Fig. B.5 for a summary of patterns when each cross contributes a median rather than mean value). Panel **a** shows univariate dominance ($d_{univariate}$; eqn. 3.1), panel **b** shows parent-bias (pairwise $d_{parent\text{-}bias}$; eqn. 3.3), and panel **c** shows mismatch (pairwise $d_{mismatch}$; eqn. 3.4). Panel **a** contains one value from all crosses ($n = 233$) while panels (**b**) and (**c**) only contain information from crosses wherein two or more traits were measured ($n = 165$).

differed significantly (at $P = 0.05$) between cross directions. The mean magnitude of phenotypic difference between cross directions was 0.65 SDs (units of smaller parental SD). Within each cross that was conducted in two directions, I calculated the fraction of traits that exhibited maternal bias and tested whether this fraction deviated significantly from 0.5. I found that traits of F_1 hybrids tend to resemble the maternal parent about 57 % of the time. ($t_{94} = 2.034$, $P = 0.0447$, 95 % CI = [0.502, 0.657]), suggesting that cytoplasmic or maternal effects are slightly more common than paternal effects.

3.3 II: Fitness consequences of parent-bias and mismatch in recombinant sunflowers

The above analyses were motivated by the hypothesis that, compared to a hybrid that is a perfect intermediate, hybrids resembling parents should fare relatively well and hybrids that exhibit trait mismatches should fare relatively poorly. However, it is not possible to test the fitness consequences of parent-bias and mismatch in the data synthesized from the literature because no studies in my dataset have both individual-level phenotype and lifetime fitness data collected in the field. In addition, studies of F_1 hybrids would have limited power to detect fitness effects of parent-bias or mismatch because there is little genetically-based phenotypic variance among F_1s within a cross. Comparisons across systems are undesirable because of methodological and biological variation among studies and systems. The optimal way to investigate the fitness effects of parent-bias and mismatch is to examine an experimental population of recombinant hybrids—wherein there is

quantitative among-individual variation in the degree of parent-bias, mismatch, and fitness—and then to use these resulting data to test whether dominance metrics are associated with fitness.

3.3.1 Methods

Study system & experimental design

To evaluate the fitness effects of parent-bias and mismatch, I leveraged data from a field experiment in annual sunflowers (*Helianthus*). The two parent species of the cross were *H. annuus* ssp. *annuus* (hereafter simply *H. annuus*) and *H. debilis* ssp. *cucumerifolius* (hereafter *H. debilis*). *Helianthus annuus* is an annual, self-incompatible, diploid that is weedy and widely distributed in its native North America. *Helianthus debilis*, by contrast, is a small sunflower endemic to central Texas. The two species are highly divergent in many traits (see Tables B.1 & B.2). Compared to *H. debilis*, *H. annuus* is much larger, has a slower life-cycle and greater longevity, has higher water-use efficiency, has thicker leaves and more leaf trichomes, has larger ligules and phyllaries, and has a different branching architecture (Whitney et al. 2006, 2010).

After four weeks of growth in a glasshouse, 503 *H. annuus* $\times$ *H. debilis* BC_1 hybrid seedlings were planted alongside individuals of both parental species in central Texas. Fitness (seed number) as well as 30 architectural, floral, ecophysiological, phenological, and herbivore resistance (e.g., trichome density) traits were measured. Of the 503 BC_1 individuals, I retained 475 in the analyses. Fifteen were excluded because of labelling mistakes and/or oversights resulting in missing trait data; I expect these exclusions were random with respect to trait and fitness values. An additional 13 plants died before some traits could be measured and were also excluded. Thus, any effects of dominance on fitness detected in my experiment reflect fertility selection rather than viability selection. Of course, dominance could also affect viability but I could not evaluate this relationship within the current study design. I applied the same trait selection and filtering criteria as in the systematic review and retained 19 traits (see Table B.1 for trait details). The data from this experiment have been previously published (Whitney et al. 2006, 2010).

Quantifying dominance metrics in the sunflowers

For each plant, I calculated pairwise $d_{\text{parent-bias}}$ and d_{mismatch} (eqns. 3.3 & 3.4) and then took the average across all trait pairs. Mean pairwise $\ln(d_{\text{parent-bias}})$ and $\ln(d_{\text{mismatch}})$ are positively correlated in this dataset (r = 0.81, $P < 0.001$; Fig. B.12) because the traits of many BC_1 individuals are transgressive and high single-trait dominance sets the upper limit of both parent-bias and mismatch. Therefore, I investigated their respective effects on fitness using multiple linear regressions of the form:

$$\ln(W_i) = \beta_0 + \beta_1 \cdot \ln(d_{\text{parent-bias}}) + \beta_2 \cdot \ln(d_{\text{mismatch}}), \tag{3.5}$$

where W_i is an individual's absolute fitness (number of seeds) in this case, and $d_{\text{parent-bias}}$ and d_{mismatch} are the mean individual dominance values averaged across each trait pair (residual error term omitted for clarity). I ln-transformed the fitness component and dominance metrics because residuals exhibited severe heteroskedasticity when the raw values were used, although the qualitative conclusions do not change if un-transformed data are analyzed. Diagnostics of the regression model indicated that, in spite of the correlation

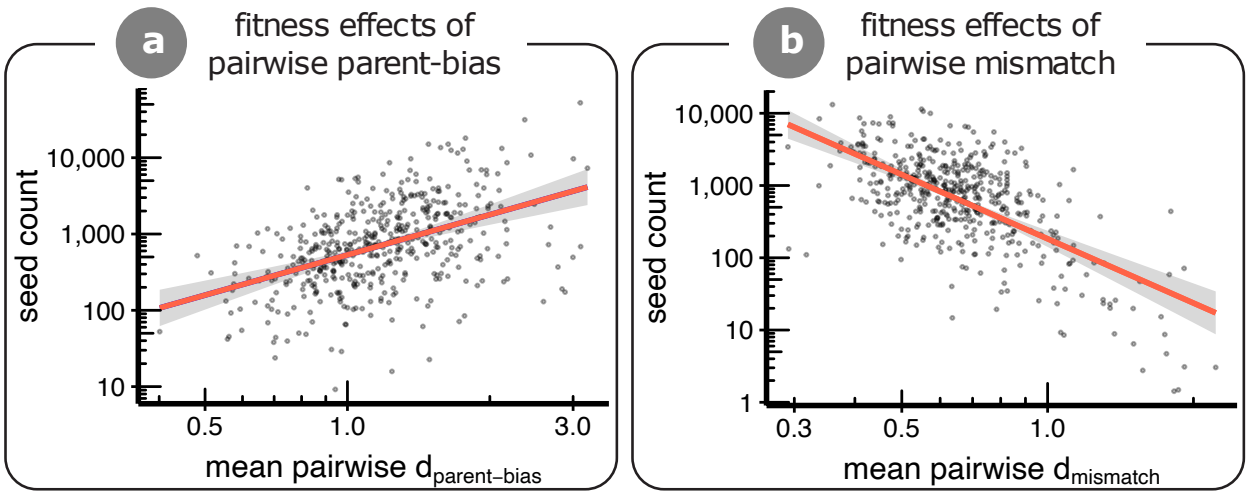


Figure 3.3: **Effect of parent-bias and mismatch on fitness in *H. annuus* × *H. debilis* BC_1 hybrid sunflowers growing in the field.** The points are partial residuals extracted from a multiple regression using `visreg` (Breheny and Burchett 2017). Each point represents one individual hybrid plant (n = 475). Both axes are $\log_{10}$ transformed. Panel **a** illustrates the effect of parent-bias and panel **b** illustrates the effect of mismatch.

between the predictors, my analysis does not suffer from multicollinearity (Variance Inflation Factor = 4.19; maximum Condition Index = 8.19). I also ran the same multiple regression for each trait pair separately and asked whether the sign of regression coefficients (β_1 and β_2) were consistent with those observed in the analysis of mean pairwise $d_{\text{parent-bias}}$ and d_{mismatch}.

3.3.2 Results

In the BC_1 sunflowers, $d_{\text{parent-bias}}$ was positively associated with seed count ($\hat{\beta}_1 = 1.75 \pm 0.26$ [SE], $F_{1,472} = 44.68$, $P = 6.56 \times 10^{-9}$; Fig. 3.3A) whereas d_{mismatch} had a negative association ($\hat{\beta}_2 = -2.95 \pm 0.16$, $F_{1,472} = 77.26$, $P < 2.80 \times 10^{-17}$; Fig. 3.3B). The multiple regression explained 20 % of the variation (i.e., r^2) in ln(seed count). Both main effects remained significant and in the same direction if an interaction term was specified in the model. In this dataset, the fitness consequences of a unit change in d_{mismatch} were larger than the fitness consequences of an equivalent unit change in pairwise $d_{\text{parent-bias}}$ ($|\hat{\beta}_1| \neq |\hat{\beta}_2|$; $F_{1,472} = 40.86$, $P = 3.94 \times 10^{-10}$). I note that pairwise trait correlations were typically quite low in these data (mean $|\rho| = 0.16$; Fig. B.13).

I also evaluated dominance-fitness relationships for each pair of traits separately. This analysis is heuristic because pairs of traits are not independent, but I present it to complement the above results. Considering only statistically significant coefficients, pairwise $d_{\text{parent-bias}}$ improved fitness for 67 % of trait pairs and d_{mismatch} reduced fitness for 81 % of trait pairs (Fig. B.14; see also Fig. B.15 for a graphical example of the trait pair with the most negative fitness consequences when mismatched). Both of these percentages are significant departures from 50 %, as determined by exact binomial tests (71 of 106 significant $d_{\text{parent-bias}}$ coefficients positive, $P = 6.1 \times 10^{-4}$; 72 of 89 significant d_{mismatch} coefficients negative, $P = 3.2 \times 10^{-9}$). Thus, the fitness consequences of pairwise $d_{\text{parent-bias}}$ and d_{mismatch} are consistent between analyses of an

individual's mean value averaged over all trait pairs and when considering pairs of traits individually.

I last evaluated whether the fitness effects of parent-bias and mismatch were driven by individuals with transgressive dominance values. I first removed all individuals from the dataset with mean pairwise $d_{mismatch}$ values > 1 and then conducted the same multiple regression analysis as above. I find that both main effects remain significant and in the same direction as above (results not shown but included in online `R` script; n = 430 plants). When conducting the analysis after removing individuals with transgressive mean pairwise $d_{parent\text{-}bias}$ values, the main effect terms were in the same direction as above but only mean pairwise $d_{mismatch}$ remained significant (n = 163 plants).

3.4 Discussion

In this chapter, I compiled data from studies that measured phenotypic traits in F_1 hybrids to characterise general patterns of hybrid trait expression. I then investigated whether the observed dominance could be predicted by genetic distance between the parents or phylogeny. Last, I tested whether parent-bias and mismatch were associated with fitness in a field experiment with recombinant hybrid sunflowers. The systematic review reveals that dominance is common: individual traits in F_1 hybrids are typically halfway between the parental midpoint and one parent's phenotype. This dominance of individual traits causes hybrids to resemble one parent more than the other and also to be mismatched. Neither genetic distance nor phylogeny predicted any metric of dominance, indicating that it will be difficult to make accurate predictions about the patterns of dominance for any individual cross. In the sunflower data, pairwise parent-bias improved fitness and mismatch reduced fitness. I discuss these results in the context of previous research on dominance and trait expression in hybrids, and highlight the implications for speciation research.

3.4.1 Genetic underpinnings of dominance and mismatch

Although dominance is commonly observed in F_1 hybrids, I do not know which trait values are derived vs. ancestral and therefore cannot relate my data to most theories on the evolution of dominance (e.g., Haldane's [1924; 1927] sieve). In any case, inter-population phenotypic divergence in most traits is likely underpinned by many quantitative trait loci (QTL) (Otto and Jones 2000) and my results hint at two general features of such QTL. First, high dominance in F_1s implies that, for many alleles underlying adaptation, the heterozygote phenotype is not simply the arithmetic mean of the alternative homozygote phenotypes. Such patterns have been documented in many QTL-mapping studies. For example, Miller et al. (2014) quantified dominance of QTL underlying marine-freshwater phenotypic divergence in threespine stickleback (*Gasterosteus aculeatus*) and found that the majority of QTL underlying marine-freshwater divergence in threespine stickleback had partial dominance effects. Second, the QTL underlying different traits seem to have unequal mean dominance coefficients—dominance for some traits is biased toward one parent and dominance in other traits is more intermediate or biased toward the other.

Any specific value of dominance is likely particular to the environment in which study organisms are measured. Dominance of individual loci has long been understood to depend on the environment (Hersh 1934), and substantial evidence suggests that hybrid phenotypes vary depending on prevailing environmen-

tal conditions (e.g., Demuth and Wade 2007). Although the results of each individual study are likely influenced by gene-by-environment interactions (G×Es), I can think of no reason why the overarching patterns documented here would change in any particular way if G×Es did not play a role.

In F_1 hybrids, transgressive trait expression might result from epistasis, although additive gene action seems more common in reviews of the topic (Rieseberg et al. 1999). Many traits in my dataset transgressed the parental range, but I caution that this does not necessarily hint at one underlying genetic architecture over another. Dominance values in an F_1 are the net effects of dominance at multiple individual loci plus additive and epistatic effects between loci, with transgressive effects at the tail of the distribution of possible outcomes. I therefore see my results in Fig. 3.2 as a documentation of pattern and do not speculate further about underlying causes.

3.4.2 Patterns & predictors of dominance

My results corroborate some previous findings but are inconsistent with others. Hubbs (1940) suggested that fishes show additive inheritance “as a very general rule" whereas Rieseberg and Ellstrand (1993) suggested plant hybrids are best characterised as being “a mosaic of both parental and intermediate morphological characters rather than just intermediate ones”. My quantitative analysis paints a picture more akin to mosaicism than strict intermediacy. In addition, I find no evidence for any major differences in dominance among taxonomic groups, which suggests the choice of study taxon does not bias estimates of dominance.

Stelkens and Seehausen (2009) found that the genetic distance between parents was positively correlated with transgression frequency—the tendency for traits to fall outside the range of parental values. I used an almost entirely independent dataset and found that genetic distance did not predict transgression or any other aspect of trait expression in hybrids. Perhaps the most likely cause of this discrepancy is that, in addition to ‘ordinary’ traits like those considered herein, Stelkens and Seehausen (2009) also considered more traditional ‘fitness’ traits. Transgression in such traits could reflect ‘intrinsic’ hybrid incompatibility (e.g., small body size due to poor condition or low seed production due to inviable ovules) and heterosis (e.g., larger body size or high seed count due to overcoming inbreeding depression in parents). Incompatibility increases with parental genetic divergence (Orr 1995; Moyle and Nakazato 2010; Matute et al. 2010; Wang et al. 2015), and heterosis seems to as well until inviability becomes substantial (Wei and Zhang 2018). Importantly, such inviability and heterosis would manifest as high $d_{univariate}$ or transgression using my approach. Because I consider traits that are putatively under stabilizing selection within populations, the mechanisms linking genetic distance with transgression in earlier studies do not apply to the present dataset.

Although I specifically excluded traits that are directly linked to fitness, it remains possible that hybrid incompatibilities underlie some patterns documented herein. For example, one study conducted crosses between wild *Drosophila melanogaster* and *D. simulans*. Male hybrids of this cross are (typically) inviable, and so David et al. (2002) only report data for females in parent species. If there is inviability that is undetected in some studies, this might influence estimates of dominance. A lack of relationship between genetic divergence and dominance, however, suggests that incompatibility is not likely the primary driving force of the observed patterns.

3.4.3 Fitness consequences of mismatch

My results clarify the potential for dominance to have a role in driving progress toward speciation. The data collated from the literature challenge the conjecture that reduced F_1 fitness is due only to phenotypic intermediacy and hybrids 'falling between parental niches' (Coyne and Orr 2004; Nosil 2012). Rather, F_1 hybrids often possess novel multivariate phenotypes that are mismatched for divergent traits. In nature, the phenotype of an organism is an integrated suite of traits that function together to influence performance and ultimately fitness (Brodie 1992; Arnold 1983). Because mismatch, caused by dominance in conflicting directions among traits, breaks up suites of integrated traits, mismatched hybrids might be poorly suited to any environment.

In the sunflower data, I found that pairwise parent-bias improved fitness and mismatch reduced fitness. Importantly, mismatch was more detrimental than parent-bias was beneficial. F_1 hybrids are likely closer in phenotype to one parent than the other, and yet at the same time have some traits resembling the less-similar parent which might render them unable to survive and reproduce in the similar-parent's niche or perhaps in any niche at all. At present, it is not clear how general this finding is. It would be valuable to conduct more field experiments with recombinant hybrids to arrive at generalities in the ways that parent-bias and mismatch affect fitness.

It is informative to examine the trait pairs that had the highest fitness consequences when mismatched (Fig. B.14). The most negative fitness effects resulted when development duration was mismatched with height of the uppermost branch (Fig. B.15). *Helianthus annuus* has a more prolonged phenology than *H. debilis*, taking about 28 days longer to initiate inflorescence formation in this experiment (Table S2). In addition, *H. annuus* is a tall plant with branches distributed throughout the main stem (uppermost branch height above ground: mean = 133 cm), whereas *H. debilis* is a much shorter plant with branches clustered at the base (uppermost branch: mean = 17 cm). Plants that mature slowly (like *H. annuus*) but are short and compact (like *H. debilis*) have lower fitness than rapid-developing compact plants (Fig. B.15). The parental phenotypes apparently reflect a trade-off, where the benefits of being a compact plant are compromised by a prolonged development.

Due to the segregation of divergent alleles, individual backcross and F_2 hybrids might be more mismatched on average than F_1s. Such increased mismatch would result in ecological hybrid breakdown—where recombinant hybrids have lower fitness than F_1s due to increased trait mismatches (Arnegard et al. 2014). In the present study, I was limited to comparing cross mean data. It would be valuable to compile data for hybrid crosses raised in a common environment where data for individual hybrids can be analyzed. In particular, quantifying how the magnitude of phenotypic mismatch observed in backcross and F_2 hybrids compares to F_1s would allow us to infer the likely strength of mismatch-based hybrid breakdown. Although F_2s are more variable than F_1s (East 1916), if divergent traits are linked in the genome (e.g., Westram et al. 2018) or are controlled by the same pleiotropic allele (e.g., Rennison et al. 2015), then segregation might not result in increased mismatch.

3.5 Conclusion

In this study I synthesized data from 198 studies to describe general patterns of phenotype expression in F_1 hybrids. Compared to previous studies with a similar goal, the distinguishing features of my analysis are that I used quantitative trait data rather than bins of 'parental' vs. 'intermediate', looked across several major clades, and examined divergently selected traits in wild organisms. For individual traits, reasonably high dominance is the rule rather than the exception. Previous studies have documented the phenomenon where dominance acts in opposite directions for different traits (Matsubayashi et al. 2010). I built on these previous studies by quantifying mismatch using simple geometry and demonstrating that mismatch affects the average hybrid to a fairly substantial degree.

Previous authors have qualitatively drawn a link between trait mismatches and hybrid fitness (e.g., Arnegard et al. 2014; Cooper et al. 2018), and I add to these earlier results by directly linking individual-level mismatch metrics to fitness in sunflowers. This result contributes to a growing literature on trait interactions in hybrids, and I suggest that future studies use my approach (or a complementary approach) to test the fitness consequences of mismatch directly. Such trait interactions are similar to Bateson-Dobzhansky-Muller hybrid incompatibilities (BDMIs) with fitness consequences mediated via ecology. Ecological BDMIs have the opportunity to affect many F_1 hybrids and could be a major mechanism of extrinsic post-zygotic isolation. Only field observations and experiments can provide the data that are necessary to test this hypothesis.

Chapter 4

Experimental hybridization studies suggest that pleiotropic alleles commonly underlie adaptive divergence between natural populations

4.1 Introduction

When populations adapt to their environment, they increase the frequency of (or fix) alleles that affect the phenotypes of traits under selection. The alleles that underlie adaptation can affect multiple traits at a time, a phenomenon known as pleiotropy (Stearns 2010). In recent years, evidence has accumulated, largely from evolutionary model systems, which suggests that pleiotropy is common (although it might only affect a small subset of an organisms' traits; Wagner et al. 2008; Wang et al. 2010; Wagner and Zhang 2011; Hill and Zhang 2012). If the pleiotropic effects of alleles are deleterious, compensatory mutations that counteract this deleterious pleiotropy can be favoured by natural selection (Phillips 1996; for empirical examples of compensatory mutation, see Adam et al. 1993; Poon and Chao 2005; Howe and Denver 2008; Merker et al. 2018). Although this model of adaptation via pleiotropy and compensation emerges in many theoretical models of adaptation (Orr 2000; Barton 2001), it is unclear whether such a process typically characterises adaptation in natural populations.

Predictions from theoretical models of divergent adaptation and hybridization can be tested to infer whether adaptation in natural populations typically involves pleiotropy and compensation. Barton (2001) conducted simulations of Fisher's (1930) geometric model of adaptation in a case where two populations with ten traits experienced divergent selection on a single trait while the other nine were subject to stabilizing selection. Following hybridization of the two populations, there was appreciable segregation variance in the nine traits under stabilizing selection. This segregation variance was caused by the recombinant hybrids inheriting alternative combinations of compensatory alleles. Importantly, the amount and/or average effect size of compensatory alleles should be positively correlated with the amount of phenotypic divergence between the parents. Thus, the theoretical prediction under adaptation via pleiotropy and compensation is: as the phenotypic divergence between pairs of populations increases, so should the amount of segregation variance in non-divergent traits observed in their hybrids. See Fig. 4.1 for a visual overview of this prediction and Fig. C.1 for the results of computer simulations illustrating the prediction more quantitatively. In this chapter, I test this theoretical prediction using data collated from experimental hybridization studies.

In doing so, I illustrate that alternative processes such as genetic drift are unlikely to underlie the observed patterns.

4.2 Methods

I conducted a systematic literature search with the goal of identifying studies that measured phenotypic traits and variances in two parent taxa and their (intra-specific or inter-specific) hybrids in a common environment. Most of the collected data are analyzed in a separate study investigating phenotypic dominance in F_1 hybrids (Thompson et al. 2021). To be selected for inclusion in the larger dataset, studies had to measure at least one non-fitness trait (i.e., 'ordinary' trait [Orr 2001]) in two parent taxa (different species or divergent populations of the same species) and their F_1 hybrids. In addition, parent taxa had to be fewer than 10 generations removed from the wild (details of the literature search are given in the supplementary methods, and the reasons for excluding each study is included in the main literature search data frame [see Data Accessibility]). In total, I (with help) screened over eleven thousand studies and collected data from 198. Of these 198 studies, all that met the following two additional criteria were included in the present analysis: (1) F_2 hybrids were measured and (2) the parents had significantly different phenotypes for at least one trait and were statistically indistinguishable for at least one other trait.

After obtaining studies for possible inclusion, I filtered and binned the data to generate summary statistics for analysis. Filtering and binning decisions were—by necessity—somewhat subjective, and I present the test of the main hypothesis for summary datasets generated under alternative filtering and binning criteria in Table S2. Since the conclusions are generally robust (highest $P = 0.0523$) to alternative data processing decisions, it seems unlikely that the study selection criteria bias my conclusions. In addition, further analysis with potentially low-power studies removed illustrate that the observed patterns are not caused by associations between sample size (number of individuals measured) and any variables (see also Table S1). I also note that methods are only briefly detailed here in the main text, but full detail with appropriate citations are given in the Appendix. All analyses were conducted in `R` v3.5.1 (R Core Team 2019) and all data underlying the article are deposited in the Dryad Digital Repository: https://doi.org/10.5061/dryad.qjq2bvqc3.

In the main text, I restrict my analysis to morphological traits—by far the most frequently measured trait type in the studies that met the above criteria—to maximize the degree to which traits and units were comparable. In total, I retained data from 15 crosses (14 studies) for the present analysis (Bradshaw et al. 1998; Bratteler et al. 2006; Hermann et al. 2015; Husemann et al. 2017; Jacquemyn et al. 2012; Koelling and Mauricio 2010; MacNair et al. 1989; McPhail 1992; Mione and Anderson 2017; Pritchard et al. 2013; Raeymaekers et al. 2009; Selz et al. 2014; Shore and Barrett 1990; Vallejo-Marín et al. 2017). Of these 14 studies, nine crossed vascular plants, four crossed fish (one study contained two crosses), and one crossed copepods. Eight crosses were inter-specific and seven were intra-specific.

For each study I divided traits into two groups: those that differed between the parents — which I assume was the result of divergent selection — and those that did not and were more likely (though not necessarily) subject to stabilizing selection. I classified traits as divergent if they were significantly different ($P < 0.05$) in a *t*-test. My conclusions are unchanged if parent divergence in phenotypic standard deviations is used

[divergent if parents are $>$ 1 SD apart] as a binning criterion. For each trait, I calculated the degree of phenotypic divergence in units of parental phenotypic SDs using the smaller of the two parental values. For each study, I then calculated phenotypic divergence for both groups of traits as the mean of ln-transformed divergence values.

For traits that were statistically indistinguishable between parents, I determined the segregation variance of each as:

$$\mathrm{var}(s) = \frac{4\mathrm{var}(\mathrm{F}_2)}{2\mathrm{var}(\mathrm{F}_1) + \mathrm{var}(\mathrm{P}_1) + \mathrm{var}(\mathrm{P}_2)} \tag{4.1}$$

Wright (1968). This quantity normalizes for the standing variation observed in each parent and F_1 hybrids and captures the variance due to the segregation of population-specific or species-specific alleles. For each cross, I took the mean of these values across all non-divergent traits after ln-transformation as an estimate of segregation variance.

My prediction was that if adaptation commonly proceeds via pleiotropic and compensatory alleles, there should be a positive relationship between parental divergence—for divergently selected traits—and segregation variance—for traits that do not differ between the parents. Visualization of linear models and statistical tests of heteroskedasticity clearly showed that the assumptions of parametric statistical analyses were violated (see Fig. C.2). I therefore tested all predictions using Spearman's rank-order correlations, which test if more divergent pairs of populations beget hybrids with more (or less) segregation variance as compared to lesser divergent parental taxa.

A similar pattern to what is predicted above could be the result of genetic drift and have nothing to do with divergent natural selection. Specifically, if more phenotypically divergent pairs also diverged longer ago than less phenotypically divergent pairs, they might have fixed a greater number of compensatory mutations for all of their traits (if such mutations fix at a steady rate over time). If this was the case, one would detect the predicted pattern even if the alleles underlying divergence were not pleiotropic. It is therefore important to rule out this role for time by testing whether phenotypic divergence of parents is correlated with their divergence time in the studies analyzed herein. I did this using three main approaches: (1) by comparing phenotypic divergence of intra-specific cross parents to that of inter-specific cross parents, (2) by evaluating the correlation between neutral gene sequence divergence and phenotypic divergence (units of base pairs), and (3) by evaluating the correlation between estimates of divergence time and phenotypic divergence (similar to [2] but in units of time based on fossil-calibrated phylogenies).

4.3 Results

I observed a positive correlation between the mean parental phenotypic divergence in statistically divergent traits and the mean segregation variance in statistically indistinguishable traits (Spearman's ρ = 0.800, P = 0.000581, n = 15) (Fig. 4.2). The magnitude of the phenotypic difference between parents for statistically indistinguishable traits was not significantly correlated with the segregation variance in those traits (Spearman's ρ = 0.446, P = 0.0972, n = 15) (Fig. C.3). The patterns were generally robust to data processing decisions (see Table S2), only slightly surpassing the significance threshold when I included physiological

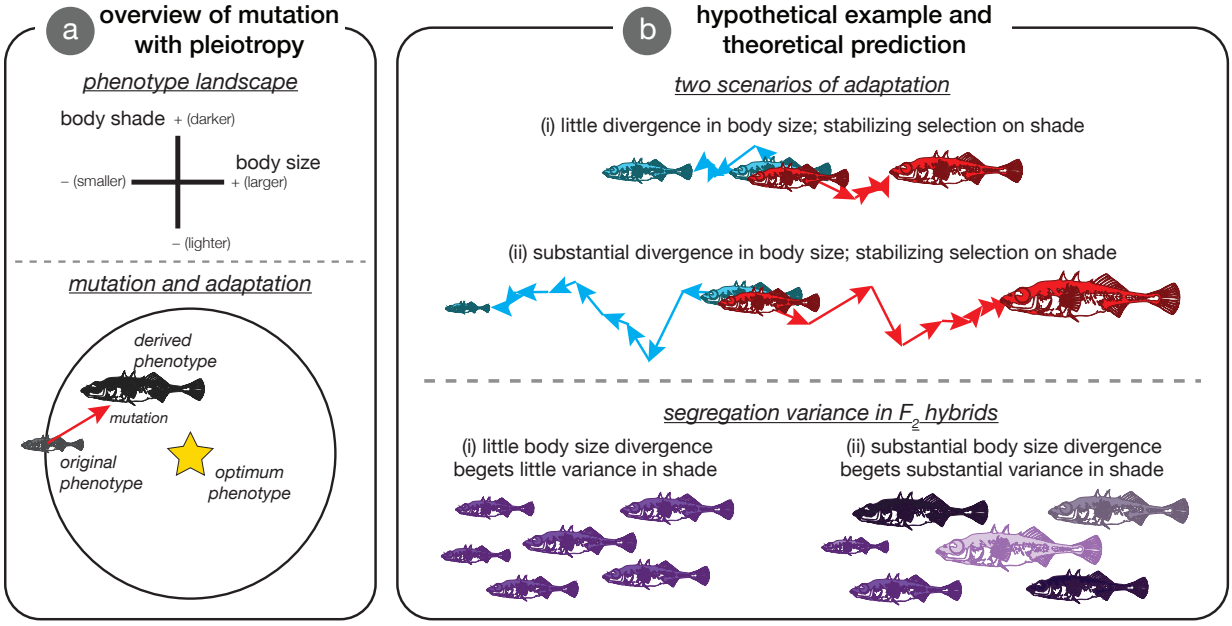


Figure 4.1: **Overview of adaptation with pleiotropic alleles and theoretical prediction.** Panel **A** shows a general overview of Fisher's geometric model, which relies on pleiotropic mutation. The upper section shows the phenotype landscape under consideration, wherein the x-axis is body size and the y-axis is body shade. The lower section illustrates the fixation of a pleiotropic allele during adaptation. The large circle defines the space wherein mutations are beneficial; mutations that point outside the circle are deleterious. The original phenotype is medium in size & shade, whereas the optimal phenotype is larger but the same shade. A mutation arises that greatly increases size and has a deleterious pleiotropic effect to darken shade. Since the mutation is beneficial (points inside the circle), it has a high probability of fixation in spite of the deleterious side-effect. Panel **B** illustrates the theoretical prediction in two diverging populations – red and blue – with the same initial phenotype for size and shade—colour here is just used to visually demarcate parent populations and hybrids (purple) and is not considered a trait. Arrows represent individual mutations as in panel (**a**). In each of two scenarios shade is under stabilizing selection in the two populations. Scenario 1 is a case where the two populations diverge little in body size and scenario 2 represents substantial divergence in body size. The lower section of the panel illustrates the outcome of hybridization. The key insight is that the segregation variance in shade is greater in scenario 2 than scenario 1. Body size segregates as well, but it would do so in a model without pleiotropy whereas shade would not necessarily. Darker recombinant hybrid individuals inherited mostly compensatory alleles that darken shade (i.e., point 'up') and lighter individuals inherited mostly compensatory alleles that darken shade (i.e., point 'down').

and chemical traits ($P = 0.052$) in the analysis.

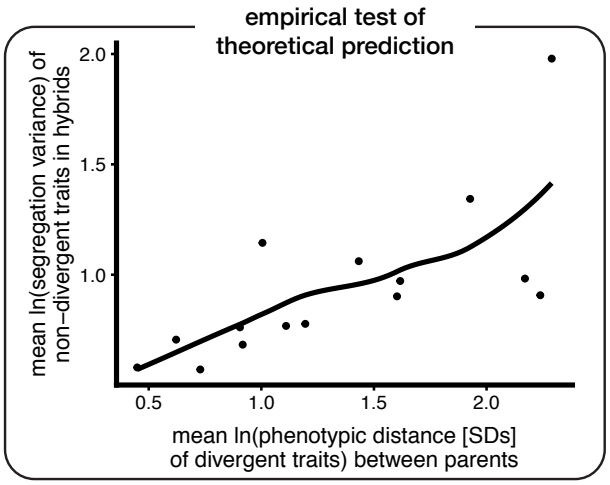


Figure 4.2: **Scatterplot depicting the relationship between phenotypic divergence in parents (statistically divergent traits) and segregation variance in hybrids (statistically indistinguishable traits).** Each point ($n = 15$) represents a unique cross between two populations or species. Points to the right on the x-axis represent crosses where the parent taxa exhibit a relatively large magnitude of phenotypic divergence for traits deemed 'divergent'. (Spearman's $\rho = 0.800$, $P = 0.000581$). The line is a loess fit.

Divergence time could correlate with phenotypic divergence between populations, which would render it difficult to disentangle the relative roles of time and phenotypic divergence in causing the pattern shown in Fig. 4.2. I found no evidence for a difference between intra-specific and inter-specific crosses in parental phenotypic divergence ($F_{1,13} = 0.013$, $P = 0.912$; Fig. C.4A & B). Additional analyses found no support for associations between any variable and genetic divergence (Fig. C.4C), divergence time (Fig. C.4D), or phylogeny (phylogenetic signal test, all $P > 0.5$).

4.4 Discussion

I leveraged data from experimental hybridization studies to conduct a correlative test of the hypothesis that divergent adaptation is associated with transgressive phenotypic variation in recombinant hybrids. This prediction holds if the genes underlying divergent adaptation are pleiotropic and does not if they are not (or if they are pleiotropic but have infinitely small individual effects [Barton et al. 2017]) (see Fig. C.1). Given the lack of effect of divergence time (or its correlates) on phenotypic divergence in the data, the consistency between the results presented here and the theoretical prediction provides indirect and correlative evidence that the genes used during adaptation are indeed pleiotropic and of appreciably large effect. The results might also hint at of the mode of adaptation for the taxa considered herein. For example, adaptation from standing variation causes greater transgressive segregation variance compared to adaptation from *de novo* mutation (Thompson et al. 2019), and thus the observed patterns could be a consequence of adaptive divergence from standing variation being commonplace (Barrett and Schluter 2008). Even if large-effect pleiotropic mutations arise, models with slowly moving fitness optima predict that only alleles with very-small effects

will be used during adaptation (Matuszewski et al. 2014). The analyses above suggest that optima in nature move quickly enough for alleles of non-trivial effect sizes to be incorporated.

My findings might initially appear to contradict the results of previous studies of transgressive segregation. For example, Stelkens and Seehausen (2009) and Stelkens et al. (2009) found that genetic distance, but not phenotypic distance, predicts transgressive segregation. Although this seems to contradict the pattern shown in Fig. 4.2, the predictions are not directly comparable because I binned traits into categories of divergent & non-divergent and compared parental divergence in the former to hybrid variance in the latter. By contrast, Stelkens' studies investigated the degree to which individual hybrids are transgressive for traits considered on their own or across all traits. Thus, my analyses test separate hypotheses. Rieseberg et al. (1999) also predicted that genetic divergence and transgressive segregation will be positively correlated when parents experience stabilizing selection at a common optimum. This prediction arises purely from substitutions fixed by drift and subsequent compensatory mutations. In the present dataset, genetic divergence is not correlated with transgressive segregation variance ($P = 0.801$; results not shown but analysis included in archived `R` script). It is likely that, in wild and outbred taxa, any effect of drift on transgressive segregation is obscured by the segregation of large-effect pleiotropic alleles and compensatory mutations fixed during adaptive divergence in other traits.

Experiments can be conducted to directly test the prediction considered herein. In an experimental evolution system where individuals and traits are easily measured, parental lines could be selected for divergence to varying degrees and then hybridized with a common ancestor. The traits that responded to divergent selection should be identified and measured, as should the traits that did not diverge and were putatively subject to stabilizing selection. The expectation is that—if mutations are universally pleiotropic—the amount of segregation variance in non-divergent traits should increase with the phenotypic distance of divergent traits. Because alleles fixed from standing variation are expected to be more pleiotropic than those fixed from *de novo* mutation (Thompson et al. 2019), the transgressive segregation variance should be greater if the population is able to use standing variance for adaptation compared to if it must rely on *de novo* mutation. If desired, one could attempt to identify the causal alleles directly using QTL mapping.

Segregation variance in non-divergent traits is expected to be deleterious and accordingly hybrid fitness should decline as the segregation variance increases. If segregation variance is observed for non-divergent traits, this directly implies that variance in the trait is deleterious—compensatory mutations would not be favoured if not for their ability to counteract *deleterious* pleiotropy. The problem with relying entirely on phenotypic measurements for empirical tests is that segregation variance could manifest in unmeasured traits and thus could easily be missed. It will therefore be useful, albeit difficult, to test predictions about fitness directly. If divergent experimental populations are hybridized, the fitness of F_1 and F_2 hybrids could be compared in a common environment. The clear prediction is that the loss in fitness of F_2 hybrids (due to segregating breakup of co-adapted compensatory alleles) compared to F_1s will be greater in more divergently selected lines. A difficulty arises when attributing this loss in fitness to segregation variance of non-divergent traits, because segregation variance in the divergent trait(s) will affect fitness in an environment-dependent manner (see Fig. 1 of Barton (2001)). For example, in an intermediate environment the F_2 would have lower fitness than the F_1 even without pleiotropy due to deleterious segregation variance of the selected

trait(s). However, if the F_2 has lower fitness than the F_1 in both the ancestral and derived environments, this implicates the segregation variance in non-divergent traits as the cause. Perhaps the best test would be to sequence the F_2s and look at selection on heterozygosity because the signature of selection against incompatible compensatory mutations in an F_2 is selection for heterozygosity (Simon et al. 2018). Thus, after measuring the fitness of F_2s in a particular environment, selection on the divergent trait(s) would manifest as selection favouring particular hybrid index and selection against segregating phenotypic variance in non-divergent traits would manifest as selection for heterozygosity. Such an experiment would be valuable for establishing a general link between adaptive divergence and reproductive isolation.

Although I illustrate a correspondence between theory and data, I did so using a correlational approach and with a small sample size of 15 crosses. I offer no conclusive proof that pleiotropic alleles and compensatory mutations are the cause of the observed pattern. There are other plausible mechanisms besides pleiotropy that could underlie segregation variance in non-divergent traits. For example parallel phenotypic evolution (if it has a non-parallel genetic basis [e.g., Ono et al. 2017]) can cause segregation variance in traits that do not differ between the parent taxa (Chevin et al. 2014; Thompson et al. 2019). For this mechanism to underlie the pattern shown in Fig. 4.2, there would have to be a correlation between parallel phenotypic evolution in some traits and divergent evolution in others — this seems unlikely. Although results presented herein are consistent with theory, empirical tests using experimental evolution would be a stronger and more direct test of the underlying mechanistic hypothesis. The ability of such studies to make a direct link to hybrid fitness is also very powerful. Such studies, paired with my indirect analysis across many taxa, would greatly strengthen my grasp on the generality of pleiotropy's role in adaptive evolution. At the very least, my analysis should serve to buttress the assessment that models fundamentally based on pleiotropy such as Fisher's (1930) geometric model are robust and useful abstractions of the evolutionary process.

Chapter 5

Adaptive divergence and the evolution of hybrid trait mismatch in threespine stickleback

5.1 Introduction

One of the central tenets of the 'ecological speciation' hypothesis is that adaptive phenotypic divergence leads to the evolution of reproductive isolation (Schluter 2000; Nosil 2012). Synthetic studies have found support for this link by illustrating that gene flow between populations decreases as their environments diverge, controlling for geographic distance (Shafer and Wolf 2013). Such patterns suggest that reproductive isolation between diverging lineages is indeed, to some extent, a function of phenotypic divergence. A critical determinant of reproductive isolation is the fitness of hybrids (Coyne and Orr 2004; Irwin 2020). 'Extrinsic postzygotic isolating barriers', which result from natural and/or sexual selection against otherwise viable and fertile hybrids, evolve before 'intrinsic' hybrid incompatibilities in many systems (Hatfield and Schluter, 1999). In these cases, the performance of hybrids under prevailing ecological conditions is likely of primary importance for determining gene flow between recently-diverged hybridizing lineages.

The phenotype of hybrids is the ultimate determinant of their performance (Arnold, 1983), and recent evidence suggests that hybrids are often 'mismatched' for divergent parental traits (Thompson et al., 2021). Mismatch refers to the case where hybrids express trait values in atypical combinations not seen in parents nor in hypothetical hybrids that are geometrically intermediate. Such mismatch results from two main mechanisms. First, differences between traits in the degree of dominance can lead to mismatch in hybrids. For example, a hybrid that is identical to one parent for some traits and identical to the other parent for other traits might have difficulty surviving and/or reproducing in nature if the trait combinations function poorly as a consequence (Matsubayashi et al., 2010). Less extreme mismatch can result when dominance is less polarized among different traits, for example when the magnitude of dominance is inconsistent among different traits that are dominant in the same direction (Thompson et al., 2021). Such dominance-caused mismatch can affect all hybrids from the first-generation (i.e., F_1) and beyond. Mismatch can also result from additive genetic variation that segregates in hybrids (i.e., segregation variance; Lande 1981; Slatkin and Lande 1994) from the first back-cross and second filial generations (i.e., BC_1 and F_2; Arnegard et al. 2014). In such recombinant hybrids, traits can become uncoupled and individual hybrids can express 'mismatched' traits due to their unique genetic composition. Importantly, dominance is often expressed in recombinant hybrids, which can mediate the mismatch caused by segregation variance.

Such trait mismatches could limit hybrid fitness in most conceivable environments because many combinations of trait values are unlikely to ever function well together. Since they cause reduced performance and fitness, mismatched traits in hybrids represent incompatibilities and are phenotypic analogs of classic Bateson-Dobzhansky-Muller incompatibilities (Bateson, 1909; Dobzhansky, 1937; Muller, 1942), but where the fitness effects are only expected to emerge under the appropriate environmental context (Arnegard et al., 2014). Several recent studies have provided evidence of such trait-trait incompatibilities in hybrids. In threespine stickleback fish, (*Gasterosteus aculeatus* L.), individual F_2 hybrids that had mismatched jaw traits tended to have reduced feeding performance compared to relatively intermediate individuals (Arnegard et al. 2014). In sunflowers (*Helianthus* spp.), individual backcross hybrids with a greater extent of mismatch across multiple pairs of traits had lower fitness than less mismatched individuals (Thompson et al. 2021). Other studies have drawn fitness inferences indirectly based on mismatched behavioural and morphological or physiological traits that characterize typical F_1 hybrids (e.g., being behaviourally attracted to one habitat but having a poorly suited physiology for that habitat; Vinšálková and Gvoždík 2007; Matsubayashi et al. 2010; Cooper et al. 2018). If trait mismatch in hybrids evolves in a manner that is predictable based on the phenotypic divergence between parents, this might represent a mechanism directly linking adaptive ecological divergence to reproductive isolation (Rundle, 2002; Nosil, 2012).

There are two main reasons why we might expect trait mismatch to increase with increasing adaptive phenotypic divergence between populations. The first is due to dominance-caused mismatch. A given amount of dominance relative to parents (e.g., 25 % more similar to parent A for trait 1, and 25 % more similar to parent B for trait 2) will generate a greater magnitude of mismatch as those two traits diverge further. The second mechanism linking mismatch to divergence is related to segregating phenotypic variation in recombinant (i.e., $F_{\geq 2}$ or backcross) hybrids. Theory predicts that the amount of phenotypic variation in the traits of recombinant hybrids will increase over the course of an adaptive walk (Slatkin and Lande 1994; Barton 2001; Chevin et al. 2014). This occurs because the number and/or effect size of QTL for a trait increases as populations diverge for that trait, and as a result a greater number of larger QTL are expected to segregate in F_2 and BC_1 hybrids in 'wider' crosses. Because of the number of crosses involved, few studies have explicitly tested this prediction with phenotype data (but see Thompson 2020; and see Edmands 1999 for a similar study with intrinsic fitness components). This increase in variance is expected to cause increased average mismatch because more extreme combinations of traits will appear in recombinant hybrids as divergence proceeds. Importantly, the effect of variance on mismatch depends on the underlying dominance patterns. If traits are additive and the hybrid mean phenotype is intermediate between parents, variance orthogonal to the axis of parental divergence will increase mismatch. If traits exhibit opposing dominance, where the hybrid mean phenotype is displaced from the axis of parental divergence, such orthogonal variance will have less of an effect on mismatch because variance will cause some individuals to be less mismatched than the mean and others to be more mismatched than the mean. As a result of these two mechanisms, mismatch in F_1s is primarily expected to be a result of dominance, whereas mismatch in recombinant hybrids might be due to one or both of dominance and segregation variance.

In this chapter, I use threespine stickleback fish to test the prediction that mismatch in hybrids increases with the magnitude of morphological divergence between parents. I leveraged the unique biology of stick-

leback, where populations have diverged to varying degrees from a common anadromous (i.e., spawning in rivers but otherwise residing in the sea) ancestor. Freshwater stickleback populations primarily differ along a limnetic (i.e., zooplanktivorous) to benthic (i.e., consuming large macro-invertebrates in vegetation or lake sediments) axis (Bell and Foster 1994). Although all have adapted to the freshwater habitat, the more limnetic freshwater populations tend to be more phenotypically similar to anadromous populations whereas the more benthic populations are relatively derived. Freshwater populations are recently (approx. 10 kya) and independently derived from an anadromous ancestor whose descendants remain abundant in the sea today and are readily crossed with derived forms. Because more benthic populations have undergone more phenotypic divergence from the anadromous ancestor, I hypothesize that their hybrids (in crosses with an extant anadromous population) will have more mismatch than those produced from less divergent populations. To test this hypothesis, I measure morphological traits in hybrids of 12 different ancestor-derived crosses, quantify mismatch, and investigate its causes. My results hint at a possible general mechanistic basis for the breakdown of (extrinsic) hybrid fitness during ecological speciation.

5.2 Methods

5.2.1 Study system

The threespine stickleback is a teleost fish species distributed throughout the coastal areas of the northern hemisphere (Bell and Foster 1994). Anadromous stickleback colonized an array of post-glacial lakes and have rapidly adapted to prevailing ecological conditions (Schluter 1996). Stickleback that live in lakes containing predators and other competitor fish species (e.g., prickly sculpin) remain similar to the anadromous population for many morphological traits (Ingram et al., 2012; Miller et al., 2019). By contrast, populations that have evolved in small lakes with few or no predators and competitors often have more derived phenotypes specialized for foraging on large benthic invertebrates.

Because adaptive divergence between anadromous and freshwater populations occurred so rapidly, populations can be readily crossed and typically have few if any 'intrinsic' incompatibilities (Hatfield and Schluter 1999; Rogers et al. 2012; Lackey and Boughman 2017). Extant anadromous populations, within a particular geographic location, are phenotypically similar to the ancestral populations that founded present-day freshwater populations (Morris et al. 2018). I leveraged this continuum of phenotypic divergence using crosses to test that hybrid mismatch will be greater when more benthic species are crossed with the anadromous ancestor than when the ancestor is crossed with more zooplanktivorous populations.

5.2.2 Fish collection and husbandry

Wild fish were collected in British Columbia, Canada, in April–June of 2017 and 2018. I sampled twelve freshwater populations from nine lakes (Fig. 5.1A; three lakes [Paxton, Priest, and Little Quarry] contain reproductively isolated benthic-limnetic 'species pairs' [McPhail 1992] and thus contributed two populations each). The anadromous population was collected from the Little Campbell River (Fig. 5.1A). Wild fish were caught using minnow traps or dip nets. I crossed six gravid anadromous females with six males from each freshwater population to generate six unique F_1 hybrid families per population, and also generated four to six

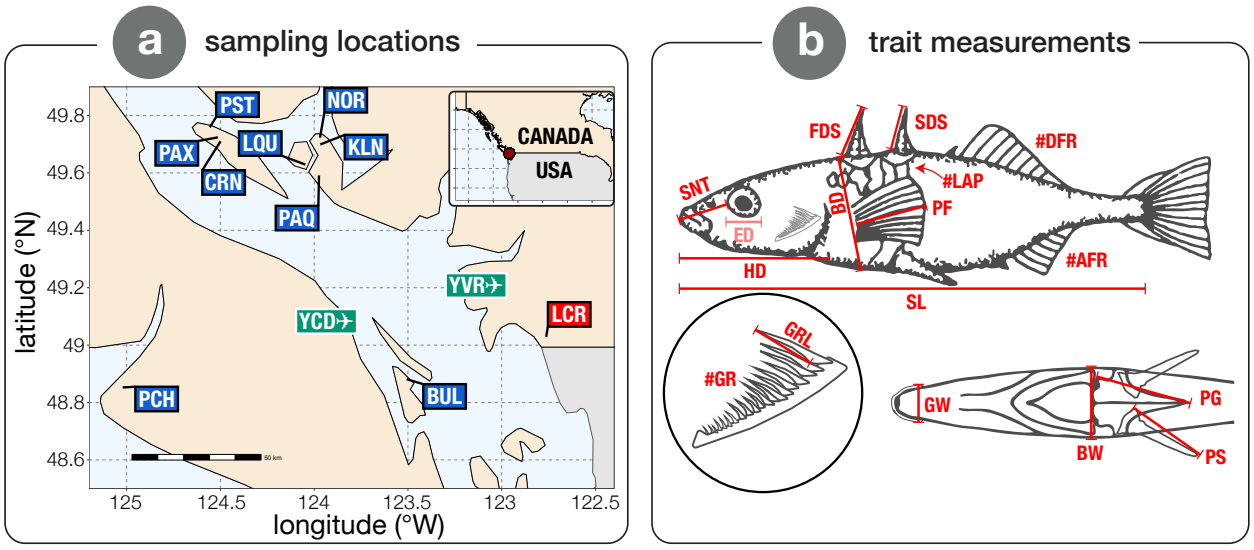


Figure 5.1: **Overview of sampling locations and trait measurements.** Panel (a) shows locations where the source populations were collected in British Columbia, Canada. Boxes show collection locations of the anadromous population (red box; LCR—Little Campbell River) and freshwater populations (blue boxes; left to right: PCH—Pachena Lake; PAX—Paxton Lake; CRN—Cranby Lake; PST—Priest Lake; LQU—Little Quarry Lake; PAQ—Paq (Lily) Lake; NOR—North Lake; KLN—Klein Lake; BUL—Bullock Lake). Green labels give airport codes for major cities (YCD—Nanaimo; YVR—Vancouver). Panel (b) shows the measurements of all 16 traits in the dataset and standard length. The upper section of the panel shows the lateral view (traits left to right: SNT—snout length; ED—eye diameter [the transparent shade of red indicates this trait was measured but not analyzed—see 'Repeatability' section of methods]; HD—head length; FDS—length of first dorsal spine; BD—body depth; SL—standard length; SDS—length of second dorsal spine; PF—pectoral fin length; #LAP—number of lateral armour plates; #DFR—number of dorsal fin rays; #AFR—number of anal fin rays). The bottom left section of the panel shows a zoomed in drawing of the upper arm of the outer gill raker arch (#GR—number of gill rakers; GRL—length of longest gill raker). The lower right section shows an anteroventral view of the body (GW—gape width; BW—body width; PG—length of pelvic girdle; PS—length of pelvic spine). The upper drawing was originally published by Bell and Foster (1994) and is re-used with permission from M. Bell.

non-hybrid (i.e., 'pure') families for each freshwater parental population and the anadromous ancestor. All offspring were raised in the lab under common conditions (see **Supplementary Methods**). Because hybrid crosses were made with the anadromous female, I cannot rule out cross-direction specific patterns. Crosses were conducted in only one direction to standardize cytoplasm among hybrid crosses and also because obtaining a sufficient number of wild gravid females for some populations was prohibitively difficult. When lab-raised fish reached reproductive maturity, F_1 hybrids from unrelated families were crossed to make three F_2 families within each cross population (with the exception of Paxton Lake benthics which, due to aquarium space constraints in 2018, had only two F_2 families from the same two F_1 parent families).

Fish from each family were lethally sampled when individuals in the tank reached a mean standard length of approximately 40 mm. Fish had not reached reproductive maturity at the time of sampling, and I therefore could not determine their sex. Fish were preserved in formalin, stained with alizarin red, and then stored permanently in 40% isopropyl alcohol. For F_1s, tanks were sub-sampled and remaining individuals were raised to produce F_2s. For F_2s, entire tanks were lethally sampled.

5.2.3 Phenotype measurements

I measured 16 traits and standard length on stained fish (Fig. 5.1B). For all traits, I measured at least 100 pure anadromous parents, and 30 pure freshwater parents, 30 F_1 hybrids, and 60 F_2 hybrids from each population and anadromous-freshwater cross (all lab-raised; see summary dataset [**to be archived on Dryad**] for trait means, standard deviations (SDs), and sample sizes for all populations). I used a dissecting microscope to count the number of dorsal fin rays, anal fin rays, lateral armour plates, and gill rakers. I also measured the length of the longest gill raker using an ocular micrometer. I photographed the left and ventral sides of each fish with a Nikon D300 camera and used ImageJ (Abramoff et al. 2004) to make linear measurements of body dimensions and bones (see Fig. 5.1 for more details on measurements). All measurements with the exception of eye diameter were highly repeatable ($r \geq 0.9$; see Fig D.1), and as a result all traits except eye diameter were used for subsequent analysis. A few ($n = 5$) fish had missing second dorsal spines, which caused them to be extreme outliers. These were likely broken off during processing, and we excluded these fish from our analyses.

I size-corrected all linear measurements by replacing raw measurements with the residuals from a log-log (ln-transformation) linear regression model on standard length conducted across the entire dataset. Log-transformation of linear measurements renders trait variances comparable across populations with different means. Some measurements are affected if fish are fixed with an expanded buccal cavity, so I further corrected for fixation position by assigning all fish a number (0, 1, or 2) depending on the extent to which the mouth was open and then performing a further correction as above using residuals for gape width, snout length, and head length. Trait measurements for missing spines (first dorsal spine or pelvic spine) or pelvic girdle were given a raw value of 0.1 mm. Unlike the second dorsal spine, variation in the presence of these traits is common and does not result in extreme outliers. Following size-correction, traits were standardized to a mean of 0 and a standard deviation of 1. I decided *a priori* to not use an approach of size-standardization (size correction by fitting a separate intercept for each population or group) because I controlled for much of the variation among populations by sampling them at a consistent mean size, and small within-group sample

sizes could lead to poorly estimated intercepts and thus less accurate size-corrected trait values.

5.2.4 Data analysis

I evaluated whether adaptive divergence between parent populations was associated with the phenotype observed in hybrids. I investigated quantitative patterns of trait mismatch, and then investigated trait dominance and variation as possible underlying causes of documented variation in mismatch.

Software

All data processing and model-fitting was done using R (R Core Team 2019) using the `tidyverse` (Wickham 2017). Mixed models were fit using `lme4` (Bates et al. 2014) and analysed using `lmerTest` (Kuznetsova et al. 2014 with the Kenward-Roger approximation for the denominator degrees of freedom (Kenward and Roger 1997). The 'map' function in `purrr` (Henry and Wickham 2019), and associated functions in `broom` (Robinson et al. 2020), were used to streamline code for iterating models over grouping variables. The `ggpubr` (Kassambara 2020) and `egg` (Auguie 2019) packages were used to create, customize, and annotate graphs. Partial residuals were plotted using `visreg` (Breheny and Burchett 2017). `for` loop code was streamlined with the functions in `magicfor` Makiyama 2016). I used the `emmeans` package (Lenth et al. 2020) and the 'cld' function in `multcomp` (Hothorn et al. 2008) to assist with post-hoc comparisons. The 'r2beta' function in `r2glmm` (Jaeger 2017) was used to calculate the partial r^2 coefficient for for the fixed effects following the method of Nakagawa and Schielzeth (2013). The functions in the 'correlation' package (Makowski et al. 2019) produced correlation matrices.

Quantifying phenotypic divergence

I quantified the magnitude of phenotypic divergence between pure anadromous and freshwater populations as my main predictor of mismatch. To do this, I simply calculated the Euclidean distance between each freshwater population's mean phenotype for all (mean and variance-standardized) traits and the anadromous mean phenotype for all traits. For pairwise analyses, I computed distances for the pair of traits being considered.

Trait 'mismatch'

Trait mismatch is a quantitative metric capturing the extent to which individual hybrids deviate from the line connecting parental mean phenotypes (Thompson et al., 2021). Mismatch is a response variable in my analyses and is measured for individual hybrids. I used two approaches to compute mismatch: one considering all traits at once in multivariate space and the other considering pairs of traits at a time. The former approach captures complexities that are missed when looking at pairs of traits but is less readily interpretable because of the high-dimensional data structure. Pairwise mismatch is oversimplified because mismatch is inherently multidimensional but is more intuitive biologically because patterns can be directly related back to traits and easily visualized. Both analyses include all traits regardless of how divergent the parents are. Genetic correlations between pairs of traits (as measured in F_2 hybrids) were low (median $|r_{\text{Pearson}}| = 0.2$), and most (87.4 %) were not statistically significant at $P = 0.05$ (Fig. D.2), and for this reason

I retain original traits and do not use dimensionality-reduction techniques such as principle components analysis.

Mismatch is the shortest (i.e., perpendicular) Euclidean distance between a hybrid's phenotype and the line that connects the two parental mean phenotypes (see Fig. 5.2A & 5.2B for visual overview). Mismatch was calculated as:

$$d_{\text{mismatch}} = \left\| (\vec{\text{F}_n} - \vec{\overline{\text{P}}_0}) - (\vec{\overline{\text{P}}_1} - \vec{\overline{\text{P}}_0}) \times \frac{(\vec{\text{F}_n} - \vec{\overline{\text{P}}_0}) \cdot (\vec{\overline{\text{P}}_1} - \vec{\overline{\text{P}}_0})}{\|\vec{\overline{\text{P}}_1} - \vec{\overline{\text{P}}_0}\|^2} \right\|, \tag{5.1}$$

where $\vec{\text{F}_n}$, $\vec{\overline{\text{P}}_0}$, and $\vec{\overline{\text{P}}_1}$ are the vectors of individual hybrid (F_n = F_1 or F_2), pooled mean anadromous, and pooled mean freshwater (of the focal population) scaled trait values, respectively. The magnitude of the resulting vector ($||d||$) is taken to get the scalar d_{mismatch}.

The goal of the mismatch analysis was to test the hypothesis that mismatch is more substantial in hybrids formed between more divergent parents. Because freshwater populations are phenotypically variable, hybrids might appear 'mismatched' if they have similar phenotypic variation to the freshwater parent. I therefore accounted for this by calculating the distance from the axis of parental divergence for the individuals of each freshwater parental population (eqn. 5.1). This quantity did not differ significantly among parent populations ($F_{11,34.6}$ = 0.23; P = 0.34), nor did it increase with the magnitude of phenotypic divergence between parents ($\hat{\beta} = 0.018$; $F_{1,43.9}$ = 0.23; P = 0.64). I therefore do not explicitly account for variation in the parental populations in my analyses. I do subtract the mean estimate of parent 'mismatch' from all hybrid mismatch values such that hybrids with no 'excess mismatch' relative to parents receive a value of 0. This simply changes the intercept and has no bearing on the main conclusions which concern slopes

I fit mixed models with mismatch (either multivariate or pairwise) as the response, and multivariate Euclidean distance between the parental populations, hybrid category (F_1 or F_2), and their two-way interaction as fixed effects. Family was a random effect. For pairwise mismatch metrics, regressions were repeated across all trait pairs. I evaluated the statistical significance of the regression models, as well as the distribution of regression coefficients across models to generated inferences about the relationship between adaptive divergence between parents and mismatch in hybrids.

Estimates of mismatch could be affected by measurement error, but measurement error is low in this dataset as determined by repeatability scores. Further, measurement error of ln-transformed trait values (i.e., absolute difference of the two ln-transformed trait measurements) is only correlated with divergence for one of fifteen traits (body depth), and freshwater population values for body depth are not significantly correlated with multivariate parent trait divergence (Spearman's rank-order correlation P = 0.1474). There is therefore no reason to think that methodological issues would generate a 'null' relationship between mismatch and divergence.

Mechanisms of mismatch

I examined how dominance and phenotypic variation contribute to mismatch. To determine how dominance affects mismatch, I calculated the mismatch of the mean hybrid phenotype for each population and hybrid

generation as the weighted (by sample size) mean across families. This generates a single estimate of mismatch for each population-generation combination, which is due only to dominance. I used a simple linear model to test whether the mismatch of the mean hybrid phenotype—which is not necessarily equivalent to the mean hybrid mismatch—was associated with the phenotypic distance between parents and whether this association differed between F_1 and F_2 hybrids.

Variance affects mismatch in a manner that depends on dominance, so I tested for the effects of variation on mismatch after accounting for dominance. Specifically, I calculated the mismatch of the mean hybrid within each family, and then subtracted each individual hybrid's observed mismatch value from this mean. If variance tends to increase mismatch, these values are expected to be large and positive. If variance causes some hybrids to be more mismatched and some to be less mismatched, these values will be centred around zero. After determining the effect of variance on mismatch for each individual hybrid, I tested whether this quantity was affected by the phenotypic divergence between parents in a mixed model where family was a random effect. Main effects were parent divergence and category (i.e., F_1 or F_2), and their interaction.

Patterns of phenotypic variation and dominance

My final series of analyses are meant to document patterns of dominance and trait variation among populations. Because dominance can cause mismatch if it is not consistent among traits (Thompson et al., 2021), I determined whether traits exhibited dominance using linear regressions that accounted for additive and dominant gene action. Throughout, I refer to dominance of the freshwater phenotype, such that traits are 'recessive' if they resemble the anadromous ancestor. For the additive term, pure anadromous, (F_1 and F_2) hybrids, and pure freshwater populations were assigned additive values of 0, 0.5, and 1, respectively (i.e., the proportion of their alleles that are 'freshwater'). For the dominance term, pure species, F_2 hybrids, and F_1 hybrids were assigned values of 0, 0.5, and 1, respectively (Lynch and Walsh, 1998) (i.e., the proportion of their genome that was heterozygous for divergent alleles). For these regressions, I only measure dominance on traits for which parents had different values (dominance is undefined for traits that don't differ). Specifically, I retained traits where the freshwater parent populations were either (or both) statistically different (*t*-test $P < 0.05$) or ≥ 1 SD apart (in units of anadromous standard deviations). I ran models where trait values were the response and additivity and dominance were (non-interacting) continuous predictors. I evaluated the statistical significance of the dominance coefficients as well as the direction of dominance (i.e., whether the anadromous or freshwater phenotype was dominant).

I next evaluated whether dominance differed among traits and populations. For these analyses, I standardized trait values of each individual so that hybrids with non-transgressive trait values fall between 0 and 1 (transgressive values are < 0 or > 1). Values of 0 indicate that F_1 hybrid trait values are the same as the anadromous parent (P_0; i.e., ancestral trait is dominant), values of 1 indicate that trait values are the same as the freshwater parent (P_{fresh}; i.e., derived trait is dominant), and values of 0.5 indicate the trait is additive. This scaled trait value, z_s, was calculated as:

$$z_s = \frac{F_n - \overline{P}_0}{\overline{P}_1 - \overline{P}_0}, \tag{5.2}$$

where F_n is individual hybrid's (either F_1 or F_2) trait value, and P_0 and P_1 represent the trait means of the anadromous and freshwater parents, respectively. The pooled means reflect the mean of families within a given cross type. To test whether dominance differed among traits and/or populations, I fit linear mixed models with 'z_s' values (eqn. 5.2) of individual fish as the response variable and family as a random effect. In models testing whether dominance coefficients varied among traits, 'trait' was the fixed effect and in models testing whether dominance differed among populations for a given trait, 'population' was the fixed effect. All *P*-values were Tukey-corrected. Because dominance varied among populations for some traits (see **Results**), I tested if dominance evolves predictably with the magnitude of phenotypic divergence between each derived population and their common ancestor for the trait in question (in units of anadromous standard deviations). In these models, family was a random effect.

I next examined patterns of phenotypic variation observed in hybrids to test hypotheses about variation being the mechanism underlying mismatch. For this analysis, I calculated the variance of each trait within each F_1 and F_2 hybrid family. I then took the mean variance across traits, and tested whether within-family mean trait variance evolved as a function of the magnitude of phenotypic divergence between parents. I fit a linear model with mean family variance as the response and parent divergence and category (and their interaction) as predictors.

5.3 Results

5.3.1 Patterns of phenotypic divergence among populations

I leveraged patterns of phenotypic divergence between anadromous and freshwater stickleback populations to quantify how mismatch evolves as adaptive ecological divergence proceeds. I found that freshwater populations differed substantially in their mean phenotypic distance to the mean anadromous phenotype (main effect of 'population': $F_{1,41.8} = 34.1$; $P = 1.0 \times 10^{-17}$; Fig. D.3). As expected, the benthic populations from the species pairs were among the most divergent from the anadromous ancestor, while two highly zooplanktivorous populations that co-exist with prickly sculpin were among the least diverged (Pachena Lake and North Lake). The least divergent population, Pachena Lake, was closer (in Euclidean phenotype space) to the anadromous population than it was to the most derived freshwater population, the Paxton Lake benthic species. The Paxton benthic population was 3.2× further from the anadromous population than Pachena Lake was. Thus, the stickleback populations considered here capture a clear quantitative continuum of phenotypic divergence from a common ancestral state. Unsurprisingly, the number of traits that differed between the freshwater and anadromous parents was positively correlated with my continuous quantitative predictor of divergence (Fig. D.4).

5.3.2 Evolution of trait mismatch

I found support for the prediction that hybrid trait mismatch increases with the magnitude of phenotypic divergence between parents. Considering all traits together, I found that multivariate mismatch in hybrids was positively associated with the magnitude of phenotypic divergence between parents ($\hat{\beta} = 0.085 \pm 0.029$ [SE], $F_{1,85.9} = 8.62$, $P = 0.0043$) (Fig. 5.2C). A separate model testing for an interaction between category

(i.e., F_1 or F_2) and parental phenotypic divergence found that it was non-significant ($\hat{\beta}$ = 0.035, $F_{1,75.9}$ = 0.31, P = 0.57). This result implies that for every unit of multivariate phenotypic divergence between parents, mismatch in hybrids increases by slightly less than one-tenth that amount. The lack of a significant interaction term implies that this result holds for both F_1 and F_2 hybrids.

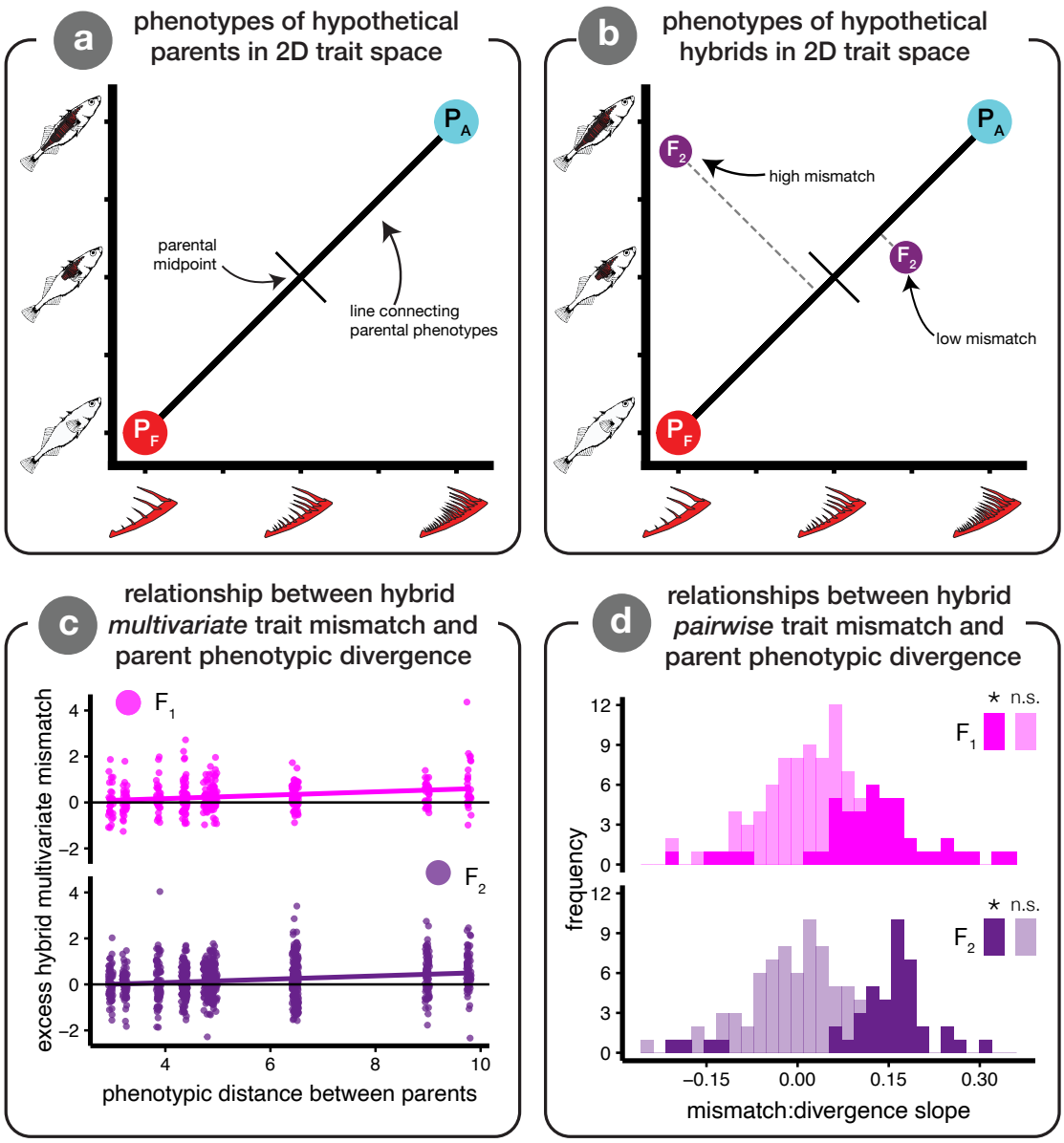


Figure 5.2: **Visual overview of mismatch calculation and observed mismatch results.** Panel (a) shows hypothetical parent phenotypes for the number of lateral armour plates and gill raker number in 2D trait space, and highlights the parental midpoint and the line connecting parents. Panel (b) shows two hypothetical hybrids in the same trait space with low or high values of mismatch (calculated as the length of the dashed grey lines). One hybrid is particularly mismatched because it has few gill rakers (a benthic freshwater-like value) and many lateral armour plates (an anadromous-like phenotype). The other hybrid has somewhat intermediate values for both traits and is, as a result, less mismatched. Panel (c) shows the statistically significant relationship between hybrid multivariate trait mismatch and parent phenotypic divergence in multivariate trait space for both F_1 and F_2 hybrids. Points are partial residuals from the model (after accounting for random effect of 'family') and are slightly jittered horizontally to aid visualization. The effect of phenotypic variation in parents (mean 'mismatch' across all parent populations) is subtracted from raw mismatch values to show the 'excess' mismatch in hybrids. This changes the height of the regression line (i.e., reduces the intercept) but does not affect the slope. Panel (d) is a histogram depicting the distribution of slopes for the pairwise mismatch analyses in F_1 (pink; upper) and F_2 (purple; lower) hybrids. These slopes are the how mismatch for a pair of traits changes with the magnitude of phenotypic divergence between parents for those two traits. All slopes are shown, with significant slopes (*) being shown in a darker shade than non-significant (n.s.) slopes. Bi-modality in the F_2s is caused by the segregation of genes with large phenotypic effect underlying lateral armour plates and the presence and size of the pelvic girdle.

I next examined the relationship between phenotypic divergence and hybrid mismatch for pairs of traits at a time. For these analyses, the predictor variable is the Euclidean distance between parent mean phenotypes for the pair of traits under consideration. I found that pairwise trait mismatch was significantly associated with parent trait divergence for 35 of 105 trait pairs in F_1s (33 %), of which 34 associations (i.e., slopes) were positive (97 %). In F_2 hybrids, 39 of 105 trait pairs (37 %) had significant mismatch-divergence associations and 38 were positive (97 %) (Fig. 5.2D). The mean absolute slope of significant relationships was approximately 0.14 in F_1 hybrids and 0.13 in F_2 hybrids, indicating that for every unit of divergence in phenotype space for a given pair of traits, mismatch between those two traits increases by a bit less than one-seventh that amount—slightly larger than the multivariate observation because this considers only significant trait pairs whereas the multivariate analysis considers all traits. Across all trait pairs, including non-significant mismatch-divergence relationships, pairwise slopes were comparable in magnitude to multivariate slopes (mean $|\hat{\beta}| \approx 0.075$). See examples of these pairwise regressions in Fig. D.5. For a depiction of pairwise mismatch for individual hybrids in 2D trait space with real data, see Fig. D.6.

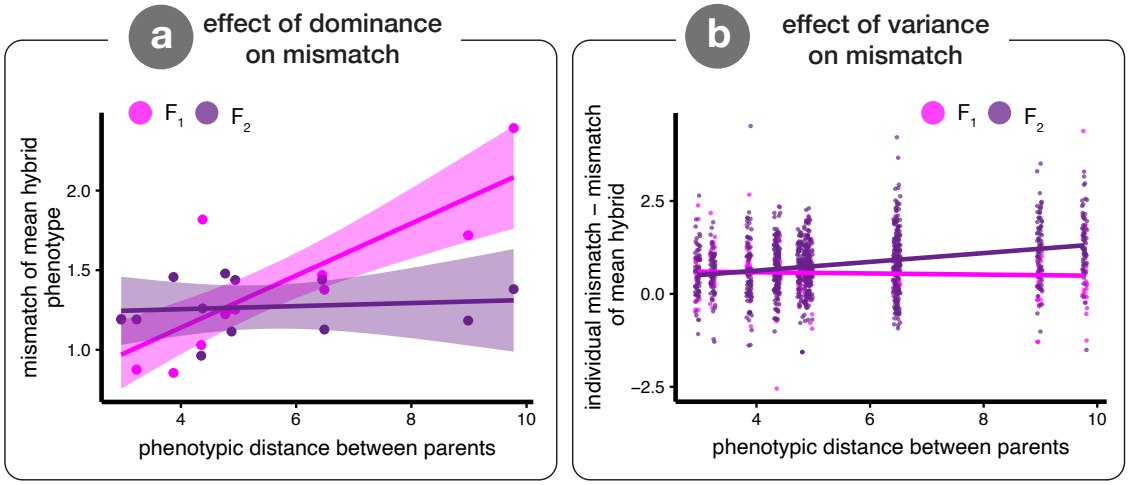


Figure 5.3: **Dominance is the primary driver of mismatch in F_1 hybrids, while variance is the primary driver of mismatch in F_2s.** Panel (A) depicts the mismatch of the mean hybrid phenotype, which is only caused by dominance. One point is shown per population and category (i.e., F_1 or F_2). The relationship is significant and positive in F_1s and non-significant in F_2s. Panel (B) depicts the effect of variance on mismatch, which is calculated as an individual's observed mismatch minus the mismatch of the mean hybrid. The relationship is significant and positive in F_2s and non-significant in F_1s.

5.3.3 Underlying causes of mismatch

Mismatch is caused by dominance and/or segregating genetic variation displacing the mean hybrid phenotype from multivariate intermediacy. In F_2 hybrids, mismatch can result from both segregation variance and dominance, whereas only dominance is expected to cause mismatch in F_1s. Because the effect of variance is dependent on underlying dominance patterns, I consider dominance first. I investigated the effects of dominance on mismatch by calculating the mismatch of the mean trait value for all traits within each population and hybrid generation, and testing whether this effect of dominance changed with the magnitude of pheno-

typic divergence between parents. I found that this relationship differed between F_1 and F_2 hybrids (parent divergence $\times$ category interaction term: $F_{1,20}$ = 11.47; P = 0.003). This interaction was caused by the fact that mismatch of the mean hybrid increased with the magnitude of phenotypic divergence between parents in F_1 hybrids ($\hat{\beta} = 0.16; t_{20} = 5.09; P = 0.0001$) but not in F_2s ($\hat{\beta} = 0.0096; t_{20} = 0.3; P = 0.76$). Patterns were qualitatively similar for pairwise trait analyses of dominance-effects (not shown). Thus, dominance is a major cause of mismatch in F_1 hybrids but not in F_2s (see results section 'Patterns of dominance in hybrids' for an analysis of dominance patterns).

Theory predicts that the phenotypic variation observed in hybrids will increase as their traits diverge and that this relationship will hold in F_2 hybrids but not in F_1s. The results of phenotypic variance support this prediction (Fig. 5.4). Across traits, mean variance increased with parent trait divergence in F_2 hybrids ($\hat{\beta} = 0.088 \pm 0.01$ [SE], t_{88} = 8.68; $P < 0.0001$) but not in F_1s ($\hat{\beta} = 0.0077 \pm 0.0077$, t_{88} = 1.0; P = 0.31). I tested whether variance causes mismatch by subtracting each individual's observed mismatch value from the mismatch of the mean hybrid of its family. I then regressed this variance-effect on mismatch against the magnitude of phenotypic divergence between parents. Similar to the dominance results, I found that this relationship differed between F_1 and F_2 hybrids (parent divergence $\times$ category interaction term: $F_{1,102.54}$ = 14.0; P = 0.0003; type-III SS). This interaction was caused by the fact that variance had an increasingly positive effect on mismatch in F_2 hybrids ($\hat{\beta} = 0.12$; $t_{38.3}$ = 4.93; $P < 0.0001$) but had no effect in F_1s ($\hat{\beta} = -0.013$; $t_{155.4}$ = -0.48; P = 0.63). Patterns were qualitatively similar for pairwise trait analyses of variance-effects (not shown). Thus, variance is a major cause of mismatch in F_2 hybrids but not in F_1s.

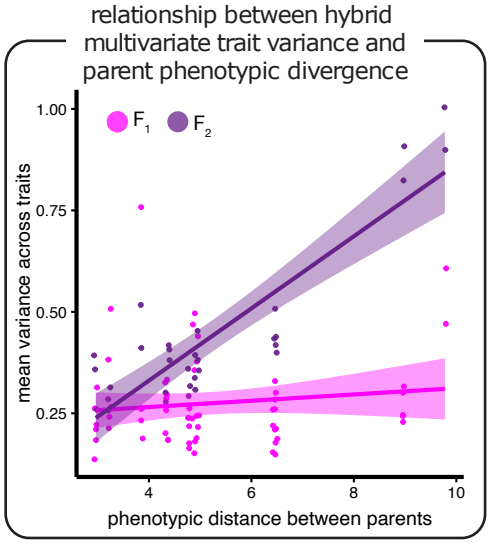


Figure 5.4: **F_2, but not F_1, hybrid phenotypic variation increases with the magnitude of phenotypic divergence between parents.** Points represent the mean of variances across all 15 traits within each independent family (minimum n = 5). F_1 and F_2 hybrids are distinguished by colour. Linear measurements are ln-transformed, so this result is not simply due to scaling means and variances (count data are raw but values are *lower* in more derived crosses for all meristic traits). Points are horizontally jittered to ease visualization.

5.3.4 Patterns of dominance in hybrids

Patterns of dominance differed among traits. Across all divergent traits in all populations, I found that 49 % of trait × population combinations (63 / 129) had significant dominance coefficients (i.e., approximately half of traits were not inherited additively). Of these traits with significant deviation from additivity, only 15.9 % were dominant (i.e., more similar to the freshwater parent than the anadromous parent; Fig. 5.5A; see Fig. D.7), while the remaining 84.1 % were recessive. As a result, dominance coefficients varied significantly among traits ('trait' main effect; $F_{14,106} = 2.37$, $P = 0.0066$), which is what results in the observed dominance-caused mismatch. For example, most F_1 hybrids have pectoral fins that are larger than the parental-midpoint, which is similar to the anadromous phenotype (Fig. 5.5A). Most F_1 hybrids also have heads that are longer than the parental-midpoint, however this is similar to the freshwater phenotype (note that this is on a log-scale). Importantly, as the magnitude of divergence between parents for two traits increases, so too will mismatch if the dominance coefficients of traits are held constant.

Dominance patterns were consistent among populations for most traits. For individual traits, dominance was typically statistically indistinguishable among all populations for a given hybrid class (e.g., mean 5.6% of pairwise differences significant among F_1 hybrids) and did not exhibit any association with phenotypic divergence. There was one notable exception, however: dominance of the lateral plate phenotype was predicted by parental phenotypic divergence. Specifically, hybrids whose freshwater parent had a relatively high number of plates (e.g., 7–8) often expressed the ancestral anadromous high-plated phenotype (i.e., trait was recessive) whereas hybrids whose freshwater parent was more derived (i.e., lower plate count [0–3 plates]) expressed an increasingly intermediate or derived freshwater phenotype (i.e., trait was dominant; Fig 5.5B). Although unexpected and biologically interesting, the evolution of dominance for lateral plate count does not affect the divergence-mismatch relationship (Fig. D.8).

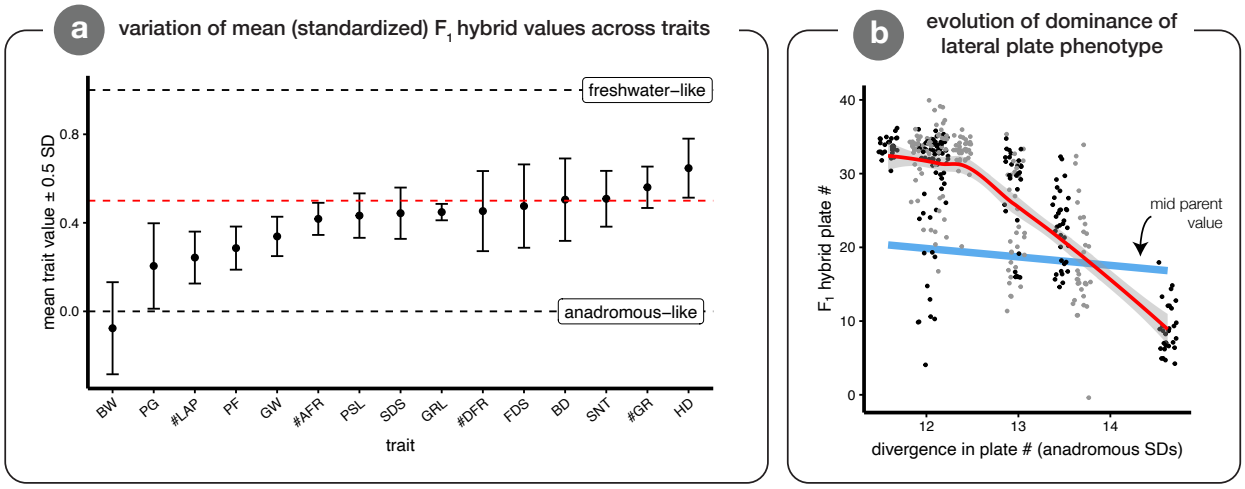


Figure 5.5: **Dominance of the freshwater phenotype in F_1 hybrids among traits and populations.** In both panels, phenotypes ([a] scaled and [b] raw), rather than dominance coefficients, are shown for ease of interpretation. Panel (a) depicts the estimated mean ($\pm\frac{1}{2}$ SD) *scaled* trait value—calculated across all populations—for all measured traits (N = 862 fish including F_1 hybrids and parents). The dashed lines at 0 and 1 represent the ancestral anadromous parent and derived freshwater parent trait values, respectively, and the red dashed line at 0.5 represents the expected value with no dominance. Panel (b) depicts the relationship between adaptive divergence in lateral plate number (in units of anadromous SDs for plate count) and F_1 phenotype (raw values) for lateral plate count. North Lake, which is fully plated, is not shown. Low values on the horizontal axis indicate that the parent population is less derived and larger values indicate that it is more derived. Black and grey dots delimit adjacent populations (as in chromosomes on a Manhattan plot). The red line is a loess-smooth fit to the data, and the blue line shows the parental midpoint. For a similar figure with F_2 hybrids, see Fig. D.7.

5.4 Discussion

In this chapter, I used experimental hybridization to investigate how hybrid phenotypes evolve as populations diverge to different extents. I was motivated by the fact that, although studies have documented a seemingly general relationship between ecological divergence and barriers to gene flow (Shafer and Wolf 2013), theoretical predictions of mechanisms that could link adaptive divergence to maladaptive hybrid phenotypes are untested. My results suggest that, at least in stickleback, more divergent parental populations tend to have hybrids with increasingly mismatched (multivariate and pairwise) phenotypes. Such a pattern might link adaptive ecological divergence with reduced gene flow via extrinsic hybrid fitness alongside other well-documented processes such as assortative mating (Rundle, 2002; Coyne and Orr, 2004; Jiang et al., 2013). This pattern manifested in both F_1 and F_2 hybrids, but for different reasons—it was caused by dominance in the F_1 and by phenotypic variation in the F_2. Below, I describe the possible mechanisms underlying these patterns, discuss their implications for speciation, and highlight critical next steps.

5.4.1 Causes of divergence-mismatch relationship

The key result of this study is that, as the magnitude of phenotypic divergence between ancestral anadromous and derived freshwater populations increases, so too does the magnitude of phenotypic mismatch observed in their hybrids. This was true when examining all traits together, and for approximately 40 % of all trait pairs. I specifically found that for every unit of divergence between parents, hybrid mismatch increases by approximately $\frac{1}{11}$ that amount in multivariate trait space. For trait pairs that showed a significant divergence-mismatch relationship, every unit divergence between the parents increased mismatch by approximately $\frac{1}{7}$ that amount. Although patterns were qualitatively similar among hybrid generations, further analyses reveals that they have different causes.

I found that mismatch was primarily caused by dominance in F_1 hybrids, and primarily caused by segregation variance in F_2s. In F_1s, traits exhibited substantial variation in dominance and had significant dominance more often than not. Although traits were dominant in F_2s, variation was more consistent among traits than in F_1s. In addition, traits tended to be more dominant in F_1s than F_2s. Because dominance typically did not change with the magnitude of divergence between parents (though see below for a discussion of lateral armour plates), increasing mismatch is therefore a natural consequence of phenotypic divergence.

By contrast, phenotypic variation was the primary cause of mismatch in F_2 hybrids and had little effect on mismatch in F_1s. Theory predicts that, if traits are inherited additively (such that F_1s were strictly intermediate), segregation variance in the F_2 variance would lead directly to mismatch and that this would increase with divergence. In support of theory (Slatkin and Lande, 1994; Barton, 2001; Chevin et al., 2014), I did find that total phenotypic variance increased with the magnitude of divergence between cross parents in F_2 hybrids. As a result, I observed the expected pattern of increasing mismatch with parent phenotypic divergence in F_2 hybrids. In F_1s, variance did not have an effect on mismatch, likely because F_1s were somewhat mismatched already due to dominance and some individuals became less mismatched as a result of variance while other individuals became more mismatched. Thus, my data support the theoretical prediction that segregation variance in recombinant hybrids causes trait mismatch, and that the size of this effect increases with the magnitude of divergence between cross parents.

5.4.2 Causes of recessivity

I found that most F_1 traits measured in the present study tended to be recessive (Fig. 5.5). The predominance of recessive alleles could be explained by processes that remove dominant alleles from the standing variation. In stickleback, alleles that allow populations to adapt to freshwater are maintained in the sea during the interglacial periods and are 'transported' back to freshwater when new post-glacial lakes are colonized (Schluter and Conte 2009). Alleles are thought to be maintained by migration-selection balance, but are likely under divergent selection between environments (Nelson et al., 2019). The alleles that freshwater populations use to adapt are therefore likely deleterious in the sea and must be maintained in the face of selection until they can be re-introduced into freshwater (Nelson et al., 2019). Theory and empirical studies indicate that, in a given environment, deleterious alleles that are recessive are maintained at higher frequencies than those that are relatively additive (Simmons and Crow, 1977; Charlesworth and Hughes, 1998; Willis, 1999; Robinson et al., 2018). Thus, a sort of 'transporter sieve' might act to reduce the frequency of dominant alleles in the

sea during inter-glacial periods, and hence influence the distribution of dominance coefficients of alleles in the standing variation. In stickleback, much of the (parallel) adaptation to freshwater habitats proceeds via the sorting of ancient standing variation (Jones et al., 2012*b*; Nelson and Cresko, 2018). Because the fixation probability of alleles in the standing variation increases with their initial frequency (MacPherson and Nuismer, 2017), if recessive alleles rise to higher frequency in the sea than additive or dominant ones, this might lead to their over-representation during adaptation. An alternative hypothesis is simply that most beneficial mutations are inherently recessive and their contributions to adaptation simply reflect the distribution of dominance among beneficial mutations.

5.4.3 Evolution of dominance for armour

Unlike most other traits, where dominance differed little among populations or was idiosyncratic with respect to parental divergence, dominance of the lateral armour plate phenotype exhibited predictable evolution. Although all freshwater populations (North Lake excepted) are low-plated (i.e., they have 0–9 plates whereas anadromous populations have around 30), populations on the higher-end of this distribution (e.g., Pachena Lake) gave rise to F_1 hybrids that largely expressed the marine phenotype. By contrast, the most derived populations (e.g., the Paxton Lake Benthic species) had values that were more similar to the freshwater parent. Plate reduction is known to be largely caused by a single large-effect variant at the *Eda* locus (Colosimo et al. 2005; Archambeault et al. 2020), which is fixed for the freshwater allele in all of the freshwater populations considered here (except for North Lake which is unusual in that it is fully plated). In addition, previous studies have shown that dominance modifiers influence plate number within populations (Colosimo et al., 2004) and that dominance for fitness at *Eda* can differ among natural populations (Schluter et al. 2021). My findings build on these earlier results by showing that the net effect of alleles that modify dominance of lateral armour plates is predictable based on the armour phenotype of the freshwater parent. It is unclear whether the alleles that modify the low-plated phenotype (i.e., reduce plate count from 7–9 plates to 0–2 plates) are themselves dominance modifiers or whether dominance modifiers are being selected for directly in the more derived populations.

Resolving the genetic basis of dominance and armour variation in these populations would shed light on whether the pattern—the predictable evolution of dominance—is incidental or adaptive. The former 'incidental' hypothesis reflects the view of Sewall Wright (Wright, 1934), who believed that dominance is largely a physiological process determined by the underlying biochemistry of adaptive substitutions. By contrast, Ronald Fisher (1931) held a differing view that observed dominance was not a product of biochemistry but rather the result of selection acting to modify the expression of a deleterious phenotypes in heterozygotes. Both phenomena are common (Mayo and Bürger, 1997), but it is unclear which is at work here. Because heterozygotes for *Eda* have likely been rare or absent for millennia in many of the freshwater populations, this suggests a physiological explanation rather than selection to reduce expression of armour plates in *Eda* heterozygotes (Otto and Bourguet, 1999). One promising way to move forward would be to identify the QTL that affect plate number differences among freshwater populations (e.g., by crossing Pachena Lake fish with Paxton Lake Benthics). In crosses with anadromous populations, it could be determined if these QTL modify dominance of *Eda* heterozygotes. Such a study would contribute novel data to a longstanding problem in

genetics.

5.4.4 Similarities to patterns of 'intrinsic' isolation

Trait mismatch represents an interaction between traits inherited from different parent lineages and affects the performance and fitness of hybrids. This is analogous to epistasis between genes that underlie traditional BDM incompatibilities. An important model in speciation genetics with some analogy to the results presented herein is the 'snowball' model of the accumulation of hybrid incompatibilities. This model, first put forward by Orr (1995), suggests that the number of hybrid incompatibilities should increase at least as quickly as the square of the number of substitutions separating species. This is because, for each additional substitution, the number of pairwise interactions—and thus pairwise incompatibilities—increases faster-than-linearly. Empirical work has found support for this snowball model (Moyle and Nakazato, 2010; Matute et al., 2010; Wang et al., 2013). My study can draw parallels to the snowball model. I determined the number of trait pairs that are significantly mismatched in hybrids (by testing whether it exceeds the 'mismatch' expected based on phenotypic variation in non-hybrid freshwater populations) and tested whether this number is correlated with the magnitude of phenotypic divergence between parents. In this analysis, I find evidence of a snowball (Fig. D.9). In F_1 hybrids, a quadratic model (with parent divergence as the dependent variable) is a substantially better predictor of the number of mismatched traits (AIC_{quad} = 54.4) than a linear model (AIC_{lin} = 69.1). For F_2s, the two models fit equally well (AIC_{lin} = 83.9; AIC_{quad} = 85.2). Because trait pairs are not independent this analysis should probably be viewed as heuristic. Nevertheless, trait mismatches could snowball in a similar manner to more 'intrinsic' BDMIs and this possibility warrants further investigation.

Because mismatch increases with parental divergence, this implies that extrinsic post-zygotic isolation evolves in a similar 'clock-like' manner to intrinsic post-zygotic isolation. Coyne and Orr (1989, 1997) were the first to demonstrate that reproductive isolation between populations evolves as a function of divergence time. In their case, they found that both pre-mating and intrinsic post-zygotic isolation increased with neutral genetic distance in *Drosophila*. This work spawned a small industry (Coughlan and Matute, 2020; Matute and Cooper, 2021) that found similar patterns in groups as diverse as orchids and fishes (Bolnick and Near, 2005; Scopece et al., 2013). The present chapter asks whether a component of extrinsic post-zygotic isolation evolves as divergence proceeds. In stickleback, genomic and phenotypic divergence are largely coincident (Miller et al., 2019; Wang, 2018), so my conclusions would likely be the same if I used genetic divergence as the predictor variable. To establish the generality of this pattern, it would be useful for other studies to determine if mismatch evolves predictably with genetic and/or phenotypic divergence in other systems.

5.4.5 Caveats and future directions

The biggest limitation of my study is its lack of a direct link between mismatch and fitness. In stickleback specifically, extrinsic selection against hybrids is well-known to occur in anadromous-freshwater stickleback hybrid zones (Jones et al. 2006; Vines et al. 2016), including the Little Campbell River (Hagen 1967) studied here. Clearly, we must begin to conduct comparative studies where mismatch can be linked to fitness directly. In hybrid zones, biologists can estimate the strength of selection against hybrids by, among other

methods, measuring cline width and back-crossing rates (Barton and Hewitt, 1985; Harrison, 1993). Natural stickleback hybrid zones abound in the present study region, and these areas of ongoing hybridization could be used to evaluate whether phenotypes measured in crosses (or gene expression) predicts the strength of selection against natural hybrids. Similar studies could be undertaken in other systems, and experimental arrays could be used to relate mismatch to back-crossing rates. Experimental evolution studies could also be used to robustly estimate the generality of the divergence-mismatch relationship and its effect on hybrid fitness. More work is necessary to solidify our general understanding of the fitness effects of mismatch.

Ultimately, the causes of speciation are multifarious and trait mismatch will be one of many causes of reproductive isolation. Empirical estimates of the relationship between ecological divergence and hybrid fitness (Edmands 1999; Funk et al. 2006; Shafer and Wolf 2013), or neutral divergence and hybrid fitness (Coughlan and Matute, 2020; Matute and Cooper, 2021), invariably find that these relationships are noisy. Because F_1 hybrid mismatch is prevalent in many systems (Thompson et al., 2021), increases in magnitude with divergence, and is expressed in all hybrids (i.e., mismatch is not expected to be highly asymmetric like lethal/sterilizing incompatibilities; Coyne and Orr 2004), it could be an immediate and powerful barrier to gene flow between many diverging lineages. As shown above, it might even 'snowball'. Nevertheless, so little is known about the extent, evolution, and fitness effects of trait mismatch that it is difficult at this stage to say whether it is a major or bit player in speciation generally. Future studies clarifying the importance of mismatch for speciation are sorely needed.

Chapter 6

Fitness consequences of hybridization after parallel evolution in threespine stickleback

6.1 Introduction

Ecological speciation occurs when reproductive isolation evolves between lineages as a consequence of their adaptation to divergent environments (Schluter, 2000; Nosil, 2012). This process is well-documented in nature and a substantial amount of evidence has accumulated over the past two decades to suggest that divergent natural selection has played an important role in many speciation events (Schluter, 2001; Langerhans and Riesch, 2013; Zhang et al., 2021). The alternative form of speciation by natural selection has come to be known as 'mutation-order' speciation, wherein parallel or uniform natural selection—favouring the same phenotypes in both lineages—drives the evolution of reproductive isolation (Schluter, 2009). Compared to speciation by divergent natural selection, speciation via parallel selection is relatively difficult to study due to the few cases of truly parallel evolution in nature (Ostevik et al., 2012), though recent work in Australian wildflowers has provided compelling support (Melo et al., 2019). As originally defined by Schluter (2009), mutation-order speciation results from the fixation of alternative alleles that would be advantageous in both populations but, due to chance processes, fix (or rise to high frequency) in only one of the two populations. When combined in hybrids, these alleles interact negatively and reduce the fitness of hybrids possessing both of them (Kvitek and Sherlock, 2011; Ono et al., 2017). Although originally proposed as a mechanism driving 'intrinsic' post-zygotic isolation, such as sterility or lethality, mutation-order processes can have important consequences for the fitness of otherwise viable and fertile hybrids. In many systems, strong 'extrinsic' barriers to gene flow can evolve before any intrinsic barriers, and my focus is on the evolution of such extrinsic post-zygotic isolation arising via mutation-order processes.

Theoretical models predict that parallel phenotypic evolution in allopatry, if underpinned by non-parallel evolution at the genetic level, can result in one or both of heterosis and hybrid breakdown (Barton, 2001; Johansen-Morris and Latta, 2006). One of the most valuable models in evolutionary genetics, Fisher's (1930) geometric model (Orr, 2005; Tenaillon, 2014), often finds heterosis in the F_1 and breakdown in the F_2 (Barton, 2001; Fraïsse et al., 2016; Simon et al., 2018; Yamaguchi and Otto, 2020). This heterosis results from the fact that F_1 hybrids are intermediate between parents for most traits, and therefore are typically closer to an optimum than one or both parents. Breakdown results from the segregation of divergent alleles during adaptation that greatly displace hybrids from the optimum (Chevin et al., 2014). Although the availability of a common pool of standing variation can reduce the degree of segregation variance, theoretical studies suggest that it typically does not fully preclude the evolution of some reproductive isolation (Thompson et al.,

2019). The relative magnitude of heterosis and breakdown vary widely depending on model parameters (Simon et al., 2018), with some parameters predicting breakdown even in the F_1 (Fraïsse et al., 2016). With respect to progress toward speciation via parallel selection, heterosis is expected to impede such progress while breakdown is expected to facilitate it. Given the range of possibilities, data from natural systems are needed to evaluate the fitness consequences of hybridization after parallel evolution in allopatry.

I examined the fitness consequences of hybridization after parallel evolution using the sympatric benthic and limnetic stickleback (*Gasterosteus aculeatus* L.) species pairs (Fig. 6.1A), which are native to three watersheds along the south coast of British Columbia, Canada (two additional species pairs, in Hadley Lake and Enos Lake, have recently become extinct—Hadley via complete removal of stickelback from the lake and Enos via collapse into a hybrid swarm). The benthic species is large-bodied and uses high suction to feed on large invertebrates in the substrate, while the limnetic species is relatively small-bodied and uses rapid jaw protrusion to feed on zooplankton in the open water (McPhail, 1984, 1992; Schluter and McPhail, 1992; Schluter, 1993; McGee et al., 2013, 2015). The species formed within the last 15,000 years after anadromous stickleback colonized newly-formed post-glacial lakes (Schluter, 1996), and genetic data indicate that each of the three extant species pairs is thought to have arisen independently (Taylor, 1999; Wang, 2018). Although sympatric, the species are thought to have arisen through a process of 'double invasion', wherein the original anadromous colonists became the benthic species and a second anadromous immigrant population eventually became the limnetic species. The benthic form is highly derived and is more genomically divergent from the anadromous ancestor than the limnetic species, which retains the ancestral zooplanktivorous niche (Jones et al., 2012*a*; Wang, 2018). Among fishes, the benthic-limnetic species pairs represent one of the clearest examples of parallel evolution when measured using approaches that geometrically quantify parallelism (i.e., phenotypic change vector analysis; Oke et al. 2017). Although evolution at the phenotypic level is highly parallel, QTL-mapping indicates that the benthic and limnetic forms in both Paxton and Priest lakes have only 40 % of QTL effects in common—suggesting that the majority of alleles used for adaptation are unique to one lake (no similar data are available for the Little Quarry Lake species pair). Given this phenotypic parallelism and genetic non-parallelism, conditions are right for heterosis and/or breakdown.

I made crosses among allopatric benthic populations and among allopatric limnetic populations (i.e., between-lake within-'species' crosses) and tracked the growth and survival of nearly 4,000 individually-tagged fish in semi-natural experimental ponds and over 2,000 fish in aquaria. I examine the fitness of parents, F_1 hybrids, and F_2 hybrids to evaluate patterns of heterosis and breakdown. Because benthics are more derived and theoretical models predict a positive association between the magnitude of phenotypic divergence (in this case between the benthic and anadromous) and hybrid breakdown (Barton, 2001; Fraïsse et al., 2016), I expected fitness differences among cross types to be more exagerated in benthics than in limnetics. My results provide insight into the efficacy of parallel natural selection for driving progress toward speciation.

6.2 Methods

Here I provide the methodological details necessary to fully understand the results. More detailed explanations regarding experimental procedures are given in Appendix E.

6.2.1 Experimental crosses

For simplicity of notation I refer to allopatric populations of the same form (i.e., benthic or limnetic) as the same 'species' even though they originated independently (Taylor and McPhail, 2000; Wang, 2018). I made all possible between-lake within-species crosses using the extant species pairs—Paxton, Priest, and Little Quarry Lakes. I generated pure (i.e., within-lake and within-species) individuals for each species and lake, and also made between-lake F_1 and F_2 hybrids. All crosses were between unrelated families, but some families and individuals were used in separate crosses—family structure was not accounted for in the analysis (see below for rationale). Parents of all crosses were born in the lab and were either one or two generations removed from the wild (i.e., all experimental fish were at minimum from the second laboratory generation). Crosses were made in March 2020, and fish were raised in 100L aquaria until mid-June 2020 when I began the pond experiment.

Figure 6.1: **Photograph of the study system.** The figure contains photographs of all three extant threespine stickleback species pairs. Each photograph was taken in 2017 and shows females of the benthic (top) and limnetic (bottom) species.

6.2.2 Pond experiment

The pond experiment took place in three experimental ponds on the University of British Columbia campus. The ponds were established between 2008 and 2010 and are 15 $\times$ 25 m in surface area. The 25 m length of the pond is made up of a 12.5 m gradually sloping shallow littoral zone with macroalgae and a 6 m deep open-water zone (see extended data Fig. 1 from Arnegard et al. 2014 for a detailed description and diagram). The diet of benthics and limnetics in the ponds closely matches their diet in natural lakes (Arnegard et al., 2014). Except for their use in previous experiments, the ponds are unmanipulated environments. Each pond contained fish (pure [i.e., non-hybrid] species, and F_1 & F_2 hybrids) with ancestry from two lakes (pond 4—Paxton & Little Quarry; pond 9—Priest & Little Quarry; pond 19—Priest & Paxton). For each pond,

tagging and release took place over the course of approximately two weeks (release windows—Pond 4: June 14–28; Pond 9: June 29–July 10; Pond 19: July 11–24).

Power analyses conducted in advance of the experiment were used to determine a sample size that would be necessary to detect a 2 % difference in standard length between treatment groups, parameterized on the standard length data of Arnegard et al. 2014. I determined that a comparison of approximately 100 fish in two treatment groups was sufficient for this purpose. I estimated that approximately 50 % of fish would perish during the experiment and so, for both benthics and limnetics, I aimed to introduce approximately 100 individuals of each pure species (i.e., 100 of both benthic parents and 100 of both limnetic parents), 200 F_1 hybrids (i.e., 200 benthic F_1s and 200 limnetic F_1s), and 200 F_2 hybrids into each pond. Total numbers of introduced fish are given in Table E.1.

Before introduction, each fish was weighed to the nearest 0.01 g and implanted with a sequential coded wire tag (CWT; Northwest Marine Technology, Anacortes, WA, USA) in the dorsal musculature on the right side of the body. CWTs can be recovered from surviving fish at the end of an experiment to identify an individual via dissection and the use of a microscope. Detailed methodology is given in Appendix E. The mean initial body mass ($\pm$ SD) was 0.56 $\pm$ 0.16 g for introduced benthics and 0.44 $\pm$ 0.112 g for limnetics, which corresponds to approx. 30–40 mm standard length for both species. After tagging, fish were returned to their original tank for a 48 hr recovery period, then moved to the ponds in coolers where their original tank water was diluted 50:50 with pond water. Fish were kept in these coolers overnight—with two aerating air stones, several plastic plants, and sections of PVC pipe—and released into the ponds the following morning. The few fish that perished before introduction had their tags extracted and read so they could be excluded from the analysis.

I retrieved surviving fish from each pond using minnow traps and by dip-netting beginning on 14 September 2020, after the experiment had run for approximately three months. Minnow traps were baited using old cheddar cheese wrapped in several layers of lightly perforated aluminum foil that fish could sense but not consume. After several days of trapping and netting, when fish returns slowed to fewer than five per evening of trapping, I added 2 L of 5 % rotenone to each pond. The rotenone caused remaining fish to swim to the surface of the pond, where they were easily collected with a dip net. After collection, fish were euthanized with an overdose of MS-222. I recorded the fresh mass of each fish, took a photograph, and then immediately stored the fish at -20° C in 15 mL centrifuge tubes containing unique paper labels. Coded wire tags were dissected out of frozen fish after lightly thawing them, then read under a microscope (Magniviewer, Northwest Marine Technology). Tags were matched with the original data to identify each fish. From this, I could calculate the survival and total growth of each individual released into ponds.

6.2.3 Lab experiment

At the same time as the pond experiment was ongoing, I conducted a similar experiment in the lab using siblings or relatives of fish released into ponds. The goal of this experiment was to provide a baseline estimate of 'intrinsic' survival and growth for fish of around the same age and starting size.

The laboratory growth rate experiment was conducted using 60 $\times$ 110 L aquaria in a common recirculating system. Most tanks contained both benthic and limnetic individuals, and all individuals within a species

were from the same family (benthics and limnetics are easily distinguished by eye). 20 g of fish were added to each tank to standardize initial mass. I recorded the mass of each fish at the experiment onset, but due to logistical constraints individual fish were not tagged. Immediately after I ended the pond experiment, I euthanized all surviving fish in aquaria with an overdose of MS-222 and recorded their mass (mean n days in aquaria = 74). I assume that the largest fish at introduction was the same individual as the largest fish at the end of the experiment (I do not know the day that a given fish died). I measured survival by comparing the number of fish of each species in tanks between the experiment's start and end. I estimate growth only in tanks where all fish survived to the end of the experiment.

6.2.4 Data analysis

Analyses proceeded via linear or generalized linear models depending on the fitness component and the hypothesis being tested. Response variables were either survival (binary; generalized linear model) or final mass (continuous; linear model). For experiment-wide differences between species and cross types within species, I fit models with pond as a random effect and initial mass as a fixed effect. For models of growth, I also included a 'duration' fixed effect which was the number of days between the introduction into ponds or aquaria and the final mass measurement. For survival models this 'duration' term was the minimum observation period, calculated as the difference between the day a fish was introduced into the pond and the first day of trapping for that pond at the end of the experiment. For models examining differences among ponds, I included pond as a fixed effect in the models and included a cross type $\times$ pond interaction term.

I did not include family (i.e., unique fertilized clutch) as a random effect, so I assume each fish constitutes a unique experimental unit (as in similar experiments, e.g., Martin and Wainwright 2013). Although families differ in survival and growth, family is completely confounded with species and cross type (by necessity) and therefore this fact alone is not useful for evaluating whether intrinsic differences among families, irrespective of their cross type, might underlie any of the main effects documented herein. To evaluate the causes of variation among families, I analyzed the initial mass data of 56 families that were split into 124 different tanks early in life. This analysis indicates that family—a biological variable—does not explain any additional variation in initial mass after accounting for 'tank'—a methodological variable (model comparison ANOVA; $P = 0.24$). Adding 'tank' into a model that contains only 'family' substantially improves the fit ($P < 0.0001$). This indicates that the specific tank a family was raised in underlies the 'family' effect. Such differences might result from many technical artifacts, such as there being slightly different densities, personalities, temperatures, and so on. I therefore conclude that the 'family' effect is noise rather than signal and feel the choice to exclude family is justified. During the pond experiment, fish across families were raised in the same broad environment (pond), which I do account for in my analysis. Accounting for initial mass in the main analysis remains useful, however, because it reduces the noise introduced by the existence of variation among rearing tanks.

Most models pool the two 'pure' (i.e., non-hybrid) species parents of a given cross for analysis. This was done because the point of comparison most relevant to heterosis and breakdown compares hybrids to the the mid-parent value. I note that important differences did exist between pure (i.e., non-hybrid) parental crosses, which are presented in the results.

All analysis was done in R version 4.0.3. Mixed models were fit with the `lme4` package (Bates et al., 2014) and analyzed with the sequential SS modifications to 'anova' in `lmerTest` (Kuznetsova et al., 2014). Main effects were evaluated using the Kenward-Roger approximation for the denominator degrees of freedom (Kenward and Roger, 1997). Post-hoc analyses were done using the 'emmeans' and 'pairs' functions in `emmeans` (Lenth et al., 2020), with Tukey HSD–corrected *P*-values. Regression model fits were visualized with `visreg` (Breheny and Burchett, 2017). Data processing used functions in the `tidyverse` (Wickham, 2017).

6.2.5 Estimating fitness

I wished to generate an estimate of fitness that considered variation in both survival and growth among cross types within a species. Estimates were made for each species (limnetic or benthic) and cross type (pure, F_1, or F_2) as measured across ponds. This is important because survival and growth cumulatively affect fitness—if one cross type had ½ the survival and ½ the fecundity of another, its fitness would be ¼. For simplicity, I simply multiply estimates of relative survival by estimates of relative size (which is proportional to overwinter survival and fecundity) to generate a single estimate of relative fitness. A more complicated analysis estimating fecundity and overwinter survival generates less conservative estimates of relative fitness (i.e., larger differences among classes) and is presented in Appendix E.

6.3 Results

6.3.1 Patterns of survival and growth in ponds

Broad differences between benthics and limnetics

I recovered and read CWTs from 59.6 % of fish that were introduced into ponds (*n* recaptured = 2213; *n* released = 3713). Across both benthics and limnetics, the initial size of the fish and the number of days a fish spent in the ponds predicted its final mass (Fig. E.1 & E.2). Recapture rates, which are assumed to reflect differences in survival, differed markedly between benthics and limnetics. I recaptured 78.2 % of benthics but only 40.9 % of limnetics ('species' main effect: $\chi^2 = 552.3$; $P < 0.0001$). Benthics grew in ponds more than limnetics, accumulating on average 0.98 ± 0.018 g [SE] more mass than individual limnetics ($F_{1,2191.1}$ = 2808.4; $P < 0.0001$). Recaptured benthics were on average $3.8\times$ their initial mass and recaptured limnetics were $2.4\times$ their initial mass.

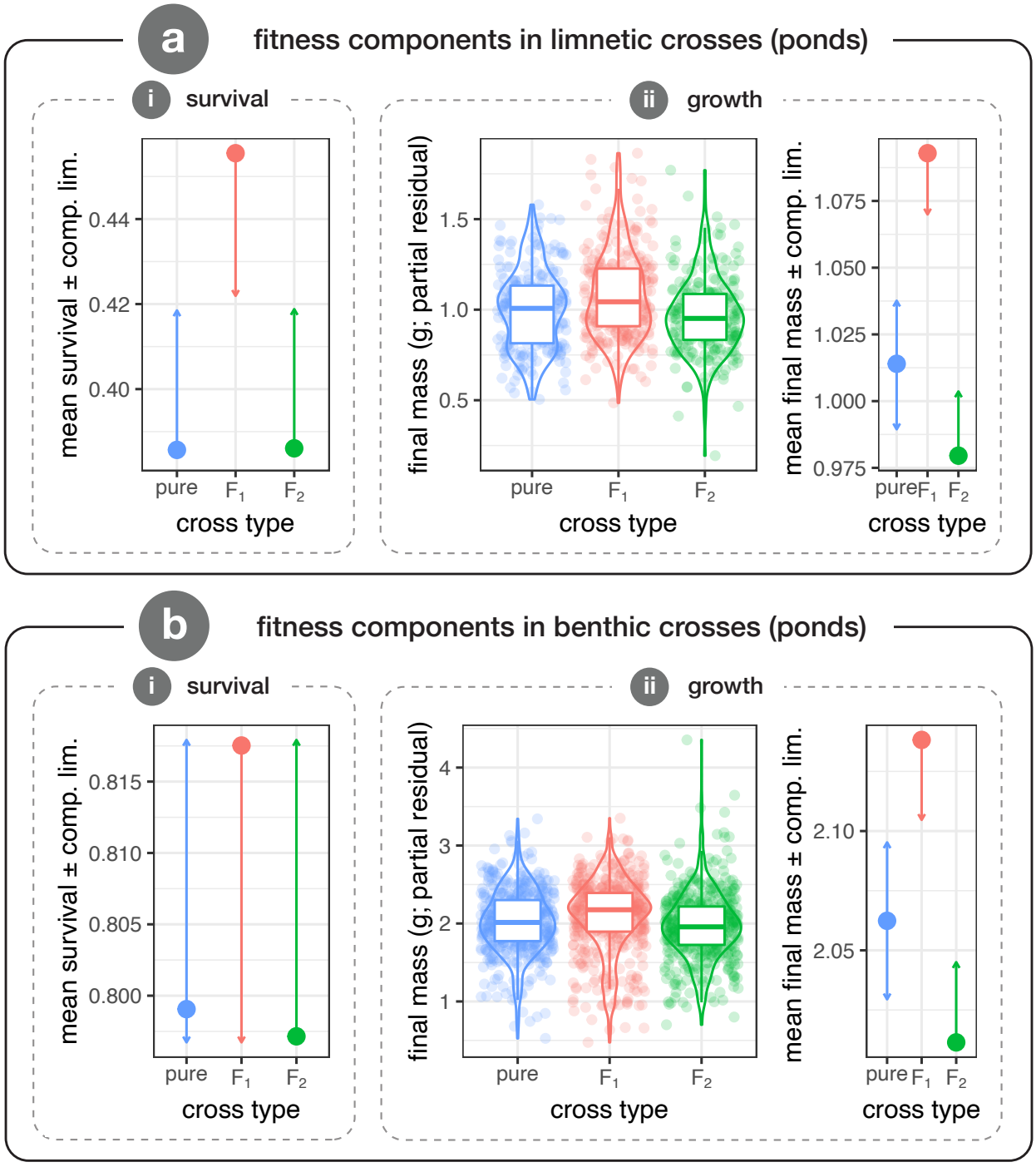


Figure 6.2: **Survival and growth of pure species and their F_1 & F_2 hybrids in the experimental ponds**. Data for limnetics are presented in panel (A), and data for benthics are shown in panel (B). All data are either estimated marginal means or partial residuals from models that included 'pond' as a random effect (see Methods). Mean survival ('i' panels) was estimated from generalized linear mixed models. Growth panels ('ii' panels) show individual-level variation (left) and estimated marginal means from linear mixed models. Arrows are comparison limits, with arrowheads indicating the direction of the comparison. Non-overlapping bars indicate a significant difference at $P = 0.05$ (Tukey's HSD).

Survival and growth differences among cross types in the experimental ponds

Survival and growth rates differed among cross types within species in the experimental ponds (Fig. 6.2). I first consider main effects across the three different inter-lake crosses—differences among crosses within species (including differences between the two pure parents) are considered below.

In limnetics, survival was higher in F_1s than non-hybrids (i.e., 'pure'; Tukey's HSD: odds ratio = 1.33 $\pm$ 0.155 [SE]; z = 2.46; P = 0.038) and F_2 hybrids (odds ratio = 1.33 $\pm$ 0.155; z = 2.45; P = 0.037), which did not differ from each other (odds ratio = 1.00 $\pm$ 0.12; z = 0.013; P = 0.99). A similar pattern was evident in the limnetic growth data, where F_1 hybrids grew more than both non-hybrids (β = 0.084 $\pm$ 0.020 [SE]; t_{745} = 4.107; P = 0.0001) and F_2 hybrids (β = 1.06 $\pm$ 0.029; t_{743} = ; $P < 0.0001$), which did not differ from each other (β = 0.027 $\pm$ 0.024; t_{710} = 0.95; P = 0.61).

Qualitative patterns for fitness components were similar for benthics as in limnetics, though the differences among cross types were smaller in magnitude. Survival in benthics was not significantly different between any cross types (all $P > 0.65$). Benthic F_1 hybrids grew more than both non-hybrids (β = 0.076 $\pm$ 0.028 [SE]; t_{1437} = 2.68; P = 0.0204) and F_2 hybrids (β = 0.13 $\pm$ 0.028; t_{1437} = 4.43; $P < 0.0001$), which did not differ from each other (β = 0.05 $\pm$ 0.028; t_{1437} = 1.80; P = 0.17).

As a result of the alignment between survival and growth, overall fitness differences among cross types matched expectations from the two fitness components and were more substantial in limnetics than in benthics. For simplicity, I simply estimate fitness as relative survival $\times$ relative growth, though less conservative estimates estimating overwinter survival and fecundity are presented in the Appendix. For both species, the F_1s had the highest fitness, followed by non-hybrids then by F_2 hybrids. For limnetics, the estimates of relative fitness were 1, 0.78, and 0.76, respectively. For benthics, these values were 1, 0.92, and 0.9. Thus, relative fitness of limnetic pure species and F_2 hybrids was less than ⅘ that of F_1 hybrids, whereas pure and F_2 hybrid benthics had approximately ⁹⁄₁₀ the fitness of F_1s.

Although there were strong main effects of cross type across the different ponds, patterns differed meaningfully among ponds (i.e., lake-of-origin pair). Because of the number of tests, I show the results of contrasts in Fig. 6.3 and refrain from extensive summary in text. Generally speaking, two of the three limnetic crosses broadly recapitulated the main effects, whereas there was no effect of cross type in the Little Quarry $\times$ Priest Lake cross. There were few differences among benthic cross types, except for the Paxton $\times$ Priest Lake cross where F_1s had much higher growth than both pure species and F_2 hybrids. In many of the crosses, there were substantial differences between cross parents which are shown in Fig. E.3. In most cases where heterosis is evident from a comparison of cross means, the F_1 mean exceeds both parents or is equivalent to the more fit parent.

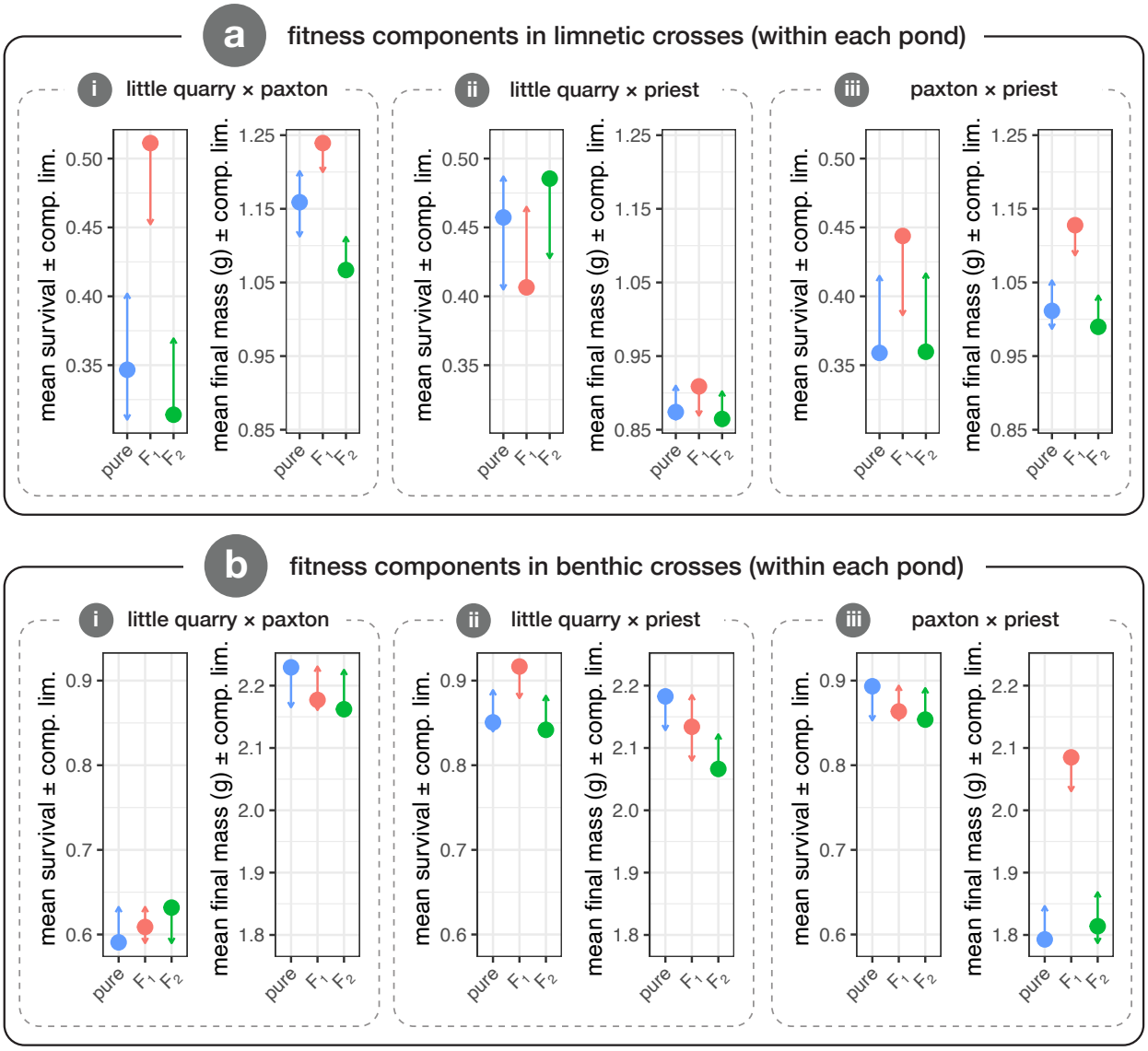


Figure 6.3: **Variation in survival and growth among cross types and crosses in the experimental ponds**. The upper panels show patterns in limnetics and the lower panels show benthics. All data are estimated marginal means ± comparison interval limits—non-overlapping intervals are significantly different at a Tukey-HSD $P < 0.05$. Arrows denote the direction of a contrast. Vertical axis limits for a given fitness component are fixed across panels for each species.

6.3.2 Causes of fitness differences among groups in ponds

Analysis of a toy model indicates that both heterosis and hybrid incompatibility underlie the rank-order of main effects observed in the pond experiment, which I qualitatively summarise as $F_1 >$ pure $\geq F_2$. For simplicity, the toy model assumes there is no dominance and that genotypic fitness landscapes are symmetrical. Heterosis is evident because F_1 hybrids exceed the fitness of pure species, and breakdown is evident because the F_2 comparison limits do not include the mid-point between pure species and F_1 hybrids. This is easily seen via toy models of simple two-locus fitness landscapes. Under a model of heterosis, fitness is a

positive function of heterozygosity (Fig. 6.4A, left) and as a result the rank order fitness is $F_1 > F_2 >$ pure. Under a model of Bateson-Dobzhansky-Muller incompatibilities (Fig. 6.4B), individuals that have alternative homozygous genotypes (i.e., AA at locus 1 and BB at locus 2) have the lowest fitness which results in rank-order fitness of F_1 = pure $> F_2$. Models combining features from both heterosis and BDMI models better match the empirical data. Models where the benefit of heterozygosity exceeds the cost of opposing homozygosity (i.e., heterosis $>$ incompatibility; Fig. 6.4C) result in $F_1 > F_2 \geq$ pure, and models with the opposite precedence (i.e., incompatibility $>$ heterosis) typically find $F_1 >$ pure $\geq F_2$. Thus, I conclude that BDMIs act at least as strongly on F_2s as heterosis in this system.

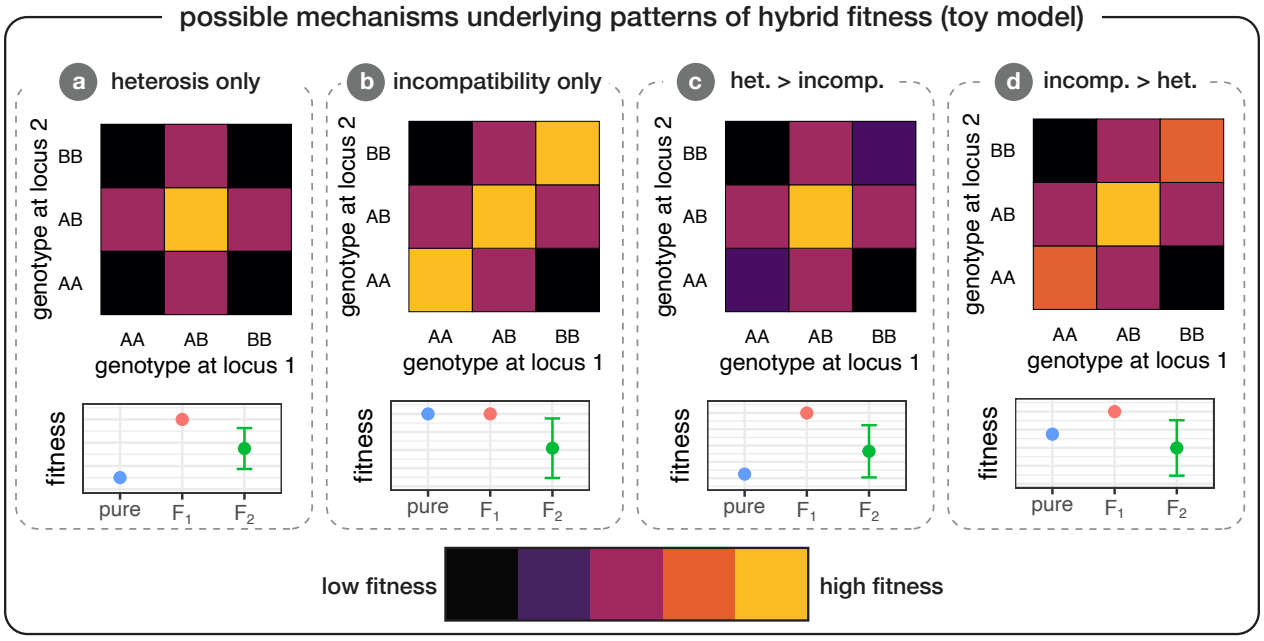


Figure 6.4: **Inferring the broad mechanisms underlying patterns of hybrid fitness.** Panel (A) illustrates four possible two-locus fitness landscapes (upper) and associated expected fitness values for parents and both F_1 and F_2 hybrids (lower). A legend for fitness is included at the bottom of the plot, with brighter colours indicating higher fitness. The final two landscapes consider cases with both heterosis and incompatibilities, but where either heterosis (iii) or incompatibilities (iv) are stronger. The lower panels show the expected fitness values (F_2s are mean $\pm$ 1SD) for each landscape.

6.3.3 Patterns of growth and survival in aquaria

As a point of comparison to the growth in ponds, I generate inferences about 'intrinsic' survival and growth of cross types using a separate set of fish reared in aquaria while the pond experiment was ongoing. The aquarium experiment ran for an approximately equal duration as the pond experiment and used fish from the same crosses (either siblings or relatives). As in the pond data, there were significant differences among limnetics for both survival and growth, and for benthics differences only manifested through growth.

In aquarium-raised limnetics, patterns were generally consistent for both fitness components where both hybrid classes were equally fit and both were more fit than non-hybrids. F_1 and F_2 hybrids had equivalent survival (odds ratio = 0.71 $\pm$ 0.18 [SE]; z = 1.40; P = 0.35) and growth (β = 0.013 $\pm$ 0.0215; t_{258} = 0.62;

$P = 0.81$). Non-hybrid individuals had lower survival than both F_1 (odds ratio = 1.86 $\pm$ 0.38; $z = 3.06$; $P = 0.0062$) and F_2 (odds ratio = 2.64 $\pm$ 0.64; $z = 3.98$; $P = 0.0002$) hybrids. Similar differences were observed for growth—F_1 hybrid limnetics grew significantly larger than parents ($\beta = 0.052 \pm 0.020$; $t_{258} = 2.63$; $P = 0.024$) and F_2s were similar though the difference was only marginally significant ($\beta = 0.066 \pm 0.028$; $t_{258} = 2.35$; $P = 0.051$). Note that the sample size is subtantially smaller for the growth analysis ($n = 264$ fish) than the survival analysis ($n = 1024$ fish) because the growth dataset was restricted to tanks where all fish survived. Thus, there is evidence for heterosis but not F_2 hybrid breakdown in the limnetic aquarium crosses.

In aquarium-raised benthics, I found general hybrid inferiority and no evidence for heterosis nor F_2 breakdown. Survival did not differ among cross types (all $P > 0.27$). For growth, however, I found that non-hybrids grew larger than both F_1 hybrids ($\beta = 0.22 \pm 0.029$; $t_{432} = 7.62$; $P < 0.0001$) and F_2 hybrids ($\beta = 0.18 \pm 0.030$; $t_{432} = 5.92$; $P < 0.0001$). F_1 and F_2 hybrids did not differ in growth ($\beta = 0.039 \pm 0.028$; $t_{432} = 1.39$; $P = 0.34$). The aquarium data thus provide no evidence for heterosis nor hybrid breakdown in benthics.

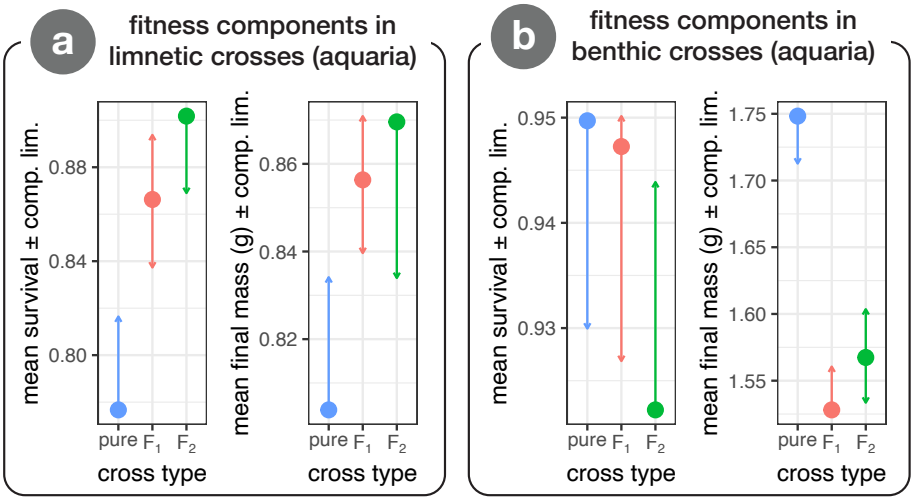


Figure 6.5: **Growth and survival in aquaria.** Panels show means and comparison interval limits for (A) limnetics and (B) benthics raised in recirculating aquaria. Survival data is from all fish, whereas growth data only uses data from tanks for which all individuals of a given species survived.

6.4 Discussion

In this study, I experimentally quantified the fitness consequences of hybridization after parallel phenotypic evolution in threespine stickleback. I considered hybridization between allopatric populations of the limnetic 'species' as well as between allopatric populations of the benthic 'species'. Fitness differences among the three cross types—pure species, F_1 hybrid, and F_2 hybrid—were qualitatively similar for both species, with the F_1 being superior to both pure species and F_2s, and the F_2 being less than or equal to parents. I hypothesized that fitness differences would be greater among benthic cross types than among limnetics, but found the opposite was true. The data suggest that both heterosis and hybrid incompatibilities govern the fitness of hybrids between these parallel species. Hybrid breakdown seems to be specific to the pond envi-

ronment because it did not occur in aquaria, whereas heterosis was apparent in aquarium-raised limnetics but not benthics. Although these main effects were significant, there were meaningful differences among unique crosses, which is consistent with the hypothesis that stochastic processes have a substantial and determinative role in governing the fitness outcomes of hybridization after parallel evolution. Nevertheless, the agreement among main effects for both species suggests that in spite of the stochastic processes operating for individual populations, there might still be general rules that become apparent when looking across pairs of populations. Below, I discuss the mechanisms that might underlie these patterns and highlight their implications more broadly for speciation via parallel natural selection.

6.4.1 Mechanisms underlying fitness differences between parents and hybrid generations in ponds

Although similar in rank-order, the differences between cross types were greater in magnitude among limnetics than among benthics. Because benthics are more derived than limnetics, and the two species have a quantitatively similar fraction of genomic non-parallelism for QTL (about 50 % of QTL unique to one lake; Conte et al. 2015), I predicted that patterns would be more pronounced in benthics. Greater group differences among limnetics might best be explained by differences in the selective regimes experienced by the two species. In the experimental ponds, limnetics were clearly under stronger selection than benthics, which is inferred through survival. If benthics experience relatively relaxed selection compared to limnetics, their fitness landscape might be viewed more as a plateau whereas that of the limnetics would be a relatively sharp peak. Thus, selection scrutinizing the phenotypic variation among limnetic crosses more closely than benthics might explain why predictions from patterns of genomic divergence were unmet.

Heterosis is the only mechanism that can readily explain the high relative fitness of F_1 hybrids compared to parents. Heterosis is often observed via extrinsic fitness in theoretical models where high dimensional organisms (i.e., having many traits) adapt to a common phenotypic optimum (Barton, 2001; Rosas et al., 2010; Fraïsse et al., 2016; Wei and Zhang, 2018; Fiévet et al., 2018; Dagilis et al., 2019; Vasseur et al., 2019). One possible cause of heterosis is inbreeding depression. Relative to limnetics, benthics have lower census population sizes (Schluter et al., 2017), heterozygosity (Jones et al., 2012*a*), and effective population sizes resulting from substantial recent bottlenecks (Wang, 2018). If benthics are inbred and/or cannot readily access beneficial alleles, they might benefit more from outbreeding than limnetics. The fact that heterosis was greater in limnetics, where inbreeding is likely relatively reduced, suggests that inbreeding is not the main cause of heterosis. Because morphological and niche-use divergence seem largely additive in this system (Miller et al., 2014; Arnegard et al., 2014), phenotypic dominance of advantageous traits is also an unlikely explanation for heterosis. Associative or pseudo-overdominance, wherein deleterious alleles hitchhike to high frequencies during adaptation (Ohta, 1971; Owens et al., 2019), might be the best explanation for the observed patterns of heterosis. In benthics, heterosis seems to be environment-specific, which is a phenomenon that has been noted before (Dominigues and Albornoz, 1987) and is well-known from cases of hybrid speciation (Rieseberg et al., 2007).

Because F_2 hybrids had equivalent or lower fitness than parents, this implicates Bateson-Dobzhansky-Muller incompatibilities. Quantification of hybrid breakdown can only occur after accounting for heterosis

(Barton, 2001), and a null expectation for F_2s is that their fitness will lie between the that of the parent mid-point and the heterotic F_1. The pond data clearly show that the F_2 hybrids of both species have lower fitness than this null expectation (i.e., their comparison limits do not include the pure–F_1 midpoint). In the aquarium-raised crosses, F_2 hybrids had equivalent fitness to F_1s in both species. This suggests that the hybrid incompatibilities are primarily subject to ecologically-mediated natural selection, rather than some form of endogenous conflict between parental genomes.

6.4.2 Variation in hybrid fitness among crosses

Although I focus my general inferences on the main effects, patterns of heterosis and hybrid breakdown varied considerably among populations. One unanticipated finding is that growth (but not survival) of the two non-hybrid (i.e., pure) crosses of a given species often differed within a pond. This implies that although the species are highly parallel with respect to some measured morphological characters, parallelism is imperfect with respect to other traits.

Qualitative patterns of cross type fitness were somewhat inconsistent among unique crosses for a given species. For limnetics, two of the three crosses recapitulated the main effect (i.e., $F_1 >$ pure $\geq F_2$), whereas this was only true for one of the three benthic crosses. The cross with the greatest among-cross differences was the benthic cross between Paxton and Priest lake, which were studied by Conte et al. (2015) and found to have approximately 40% of parallel QTL underlying adaptation. The variation among ponds is in agreement with the notion that stochastic processes play a major role in speciation by parallel natural selection (Mani and Clarke, 1990; Schluter, 2009). As a result, it might be difficult to predict the outcome of hybridization after parallel phenotypic evolution in any one population. Importantly, the main effects were largely consistent across benthics and limnetics, which might suggest that generalities can be found when looking across populations (Melo et al., 2019).

6.4.3 Caveats

The experiment makes a number of assumptions that should be acknowledged. First, an implicit assumption is that the experimental ponds are reasonable stand-ins for the lakes. This assumption is common to all studies taking place outside of species' source habitat. Unfortunately, introducing inter-lake hybrids into study lakes, even into enclosures, carries an unacceptable risk of escape and thus will never be attempted. Previous studies have found that the diet of benthic and limnetic stickleback in these same ponds closely match that of the natural habitat (Arnegard et al., 2014), which could alleviate this concern. My crosses also only capture natural selection, and important differences in sexual selection might have been overlooked (Keagy et al., 2016). I also assume that selection was parallel, when in reality there are probably some traits that are divergently selected between allopatric benthics and between allopatric limnetics. Finally, selection against larval fish can be extremely strong (China and Holzman, 2014) but could not be accounted for here because my experimental design required me to tag fish, which could not be done until fish were large enough to survive such handling.

6.4.4 Progress toward speciation via parallel natural selection?

In this chapter I have shown that parallel evolution can result in (some) ecological hybrid breakdown and also (some) heterosis. It remains unclear how this breakdown would affect progress toward speciation. In the sympatric benthic-limnetic species pairs, the species seem reproductively isolated because the genes that underlie adaptation also underlie mate choice (Bay et al., 2017). The species differ in body size, which acts as a reliable signal of species recognition (Conte and Schluter, 2013), and this selection has been honed through reinforcement (McKinnon and Rundle, 2002). Thus, although the allopatric benthic and limnetic species' genomes are weakly incompatible, it seems unlikely that the incompatibility is so strong that the allopatric 'parallel' species would remain distinct if brought together in sympatry. The observed heterosis might render a collapse even quicker. Thus, I conclude that there has been little progress toward speciation via parallel natural selection in this system. If this is general, it would imply that parallel natural selection is unlikely to generate substantial reproductive isolation via extrinsic post-zygotic isolation. Chance fixation of alleles with substantial 'intrinsic' effects (Melo et al., 2019) might be the primary pathway through which mutation-order speciation can be expected to act.

Chapter 7

Genetic evidence for environment-specific hybrid incompatibilities in threespine stickleback

7.1 Introduction

Hybrid incompatibilities—interactions among divergent genetic loci that reduce the fitness of hybrids—are a key component of reproductive isolation between diverging lineages (Coyne and Orr 2004). Incompatibilities have been studied most intensively in the context of sterility and embryo mortality, because these traits are conducive to reliable phenotyping in the laboratory and often seem to be underpinned by few loci (Maheshwari and Barbash 2011; Fishman and Sweigart 2018). These sorts of barriers have come to be called 'intrinsic' hybrid incompatibilities due to the fact that there are conflicts within the hybrid genome that are expected to severely impact hybrids in most environmental contexts (though note that the strength of selection against some intrinsic incompatibilities can vary across environments [Demuth and Wade 2007]). Studies have shown that the number of intrinsic incompatibilities increases deterministically with genetic divergence between parents (Matute et al. 2010; Moyle and Nakazato 2010; Wang et al. 2013) and that incompatibilities are common throughout the genomes of isolated populations of a given species (Corbett-Detig et al., 2013). Collectively, these studies have made substantial progress toward identifying generalities about the evolutionary genetics of intrinsic hybrid incompatibilities.

Incompatibilities can also be caused by exogenous ecological selection if particular allele combinations render hybrids unable to perform ecological functions such as predator avoidance or prey capture. Several recent studies have shown patterns to this effect, wherein hybrids have 'mismatched' trait combinations and reduced fitness as a result (Arnegard et al., 2014; Thompson et al., 2021). Such studies have successfully demonstrated incompatibilities via interactions among traits, but links to the underlying genetics have not been made. Perhaps the most significant barrier to progress in measuring 'ecological' hybrid incompatibilities is that they are likely difficult to detect because of their small individual effect sizes (Rockman 2012). For example, Arnegard et al. (2014) found that combining divergent jaw bone traits together in an F_2 stickleback (*Gasterosteus aculeatus* L.) hybrid reduced their fitness because these traits interacted in a manner that rendered hybrids unable to effectively generate suction for feeding. The underlying interacting jaw bones each map to several QTL that individually explain a small fraction of the phenotypic variance, thus rendering it difficult to study their individual fitness effects.

Recent theoretical advances, however, suggest ways to test for and measure the net effect of hybrid in-

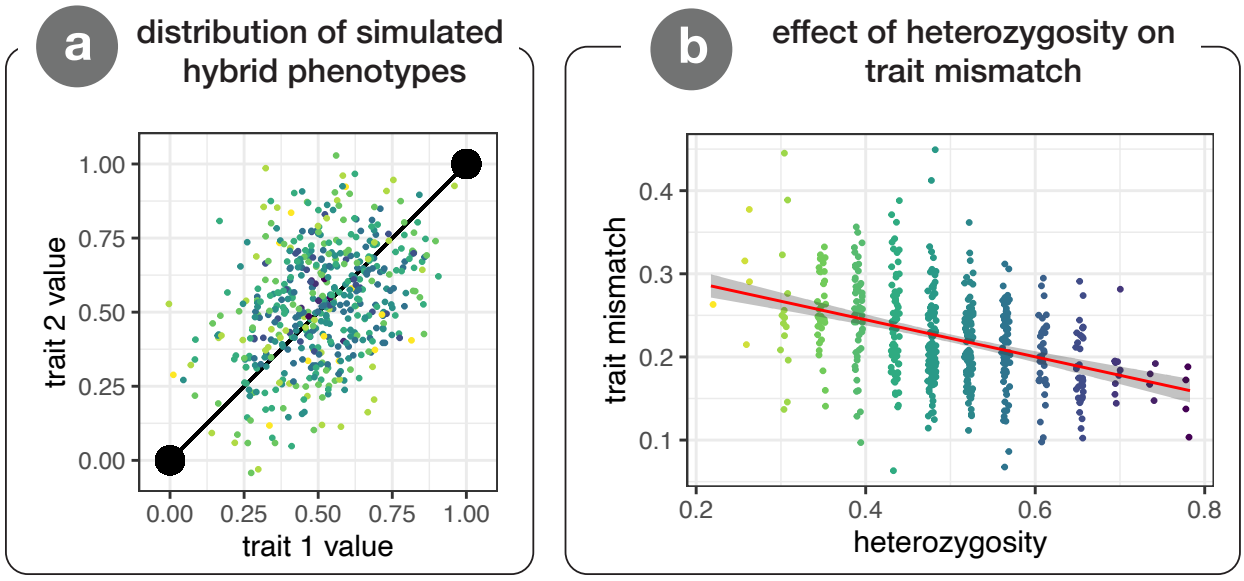


Figure 7.1: **Results from simulations illustrating an ecological mechanism underlying the heterozygosity-incompatibility relationship in F_2 hybrids.** Both panels depict results from a representative simulation run of adaptive divergence and hybridization between two populations. I consider an organism with two traits that have both diverged as a result of selection. Panel (A) depicts the distribution of 1,000 F_2 hybrid phenotypes in two-dimensional trait space. Large black points are the two parent phenotypes, which are connected by a black line indicating the 'axis of divergence'. Points are coloured by heterozygosity, as in (B). Panel (B) depicts the relationship between heterozygosity and trait 'mismatch' of individual hybrids. Mismatch is calculated as the shortest (i.e., perpendicular) distance between a hybrid's phenotype and the black line connecting parents. The plot shows that mismatch is lower in more heterozygous F_2s. Heterozygosity values are discrete because a small number of loci underlie adaptation in the plotted simulation run.

compatibilities using experimental crosses. Specifically, selection against hybrid incompatibilities in an F_2 hybrid cross manifests as directional selection for increased heterozygosity (Barton and Gale, 1993; Simon et al., 2018). This is expected because F_2 hybrids have a hybrid index of approximately 0.5—having half of their alleles from one parental species and half from the other. Given this, individuals with high heterozygosity have fewer pairs of homozygous loci with opposite ancestry than more homozygous individuals. Having many loci with opposite homozygous ancestry can result in hybrids with maladaptive 'mismatched' phenotypes, whereas highly heterozygous individuals are expected to be relatively intermediate (Fig. 7.1). Whether 'mismatch' affects fitness, however, ultimately depends on the ecology of the system and the underlying fitness landscape. Such 'coarse' approaches—coarse because they use summary statistics rather than fine mapping—are a promising means to identify the net strength of small-effect hybrid incompatibilities using field experiments.

In this study, I compare patterns of directional selection on heterozygosity between F_2 hybrid families raised in the lab to those from the same cross types raised in field enclosures. If ecological selection on 'trait mismatch' (i.e., selection favouring 'matched' trait values whether parental or somewhat intermediate) is operating in the field but not in the lab, selection for heterozygosity should be specific to the field (or at

least stronger than in the lab). I focus on hybrids between two types of ecologically divergent threespine stickleback populations. First, I consider hybridization between sympatric benthic and limnetic populations. These populations, which are in effect reproductively isolated species due to strong assortative mating (McKinnon et al. 2004) and reduced hybrid fitness due to extrinsic selection pressures (Hatfield and Schluter 1999), comprise sympatric species pairs that have evolved independently in at least five watersheds in British Columbia, Canada (McKinnon and Rundle 2002; McPhail 1992). Although reproductively isolated in practice, the species pairs have no known intrinsic barriers that reduce fitness in the lab (Hatfield and Schluter 1999). Second, I consider hybridization between allopatric populations of anadromous and solitary freshwater stickleback. Similarly to the benthic-limnetic species pairs, these populations are recently diverged and can readily hybridize. I pair an analysis of genetic data with a heuristic model to infer the strength of selection that is necessary to cause the observed patterns. My results provide compelling support for the hypothesis that environment-specific hybrid incompatibilities separate recently diverged stickleback populations.

7.2 Methods

7.2.1 Data sources

I used both previously published and unpublished data in my analyses. Summary information about each data source is listed in Table 7.1. I base my main inference on a comparison of heterozygosity in hybrids raised in aquaria to hybrids from the same crosses raised in experimental ponds. Besides their use in previous experiments, ponds are natural contained ecosystems. See Extended Data in Arnegard et al. (2014) for additional information about ponds. Pond-raised crosses capture both 'intrinsic' and 'extrinsic' incompatibilities, whereas aquarium-raised crosses are expected to capture 'intrinsic' incompatibilities that kill hybrids from the embryo to adult stage.

Studies raising fish from the same cross type (benthic $\times$ limnetic or marine $\times$ freshwater) in the same environment (aquaria or pond) were pooled for analysis. Two studies (Rennison et al., 2019; Schluter et al., 2021) genotyped both F_2 and F_3 hybrids, which I analyze together because mean heterozygosity did not differ between generations within a study (Fig. F.1). Similarly, data were analyzed together for four studies of benthic $\times$ limnetic hybrids from Paxton Lake raised in experimental ponds because heterozygosity was statistically indistinguishable among them (Fig. F.2). This grouping of studies was done only to simplify the presentation of results—patterns are highly repeatable across replicates and analyses showing results for each pond and/or study separately are shown in Fig. F.3. Relevant details of each data source are outlined below.

Studies genotyped fish using either microsatellites, SNPs, or Genotyping-by-Sequencing (GBS). All three lab studies used microsatellites, whereas the pond studies all used SNPs or GBS. I have no reason to believe that differences in methodology among studies influence the results presented herein. First, I only consider loci where parents have no alleles in common and thus can accurately assign ancestry. Second, I only use loci that were heterozygous for ancestry in F_1s, so loci with 'null' microsatellites or any difficulties in distinguishing alleles would be filtered out. In the largest microsatellite dataset (Rogers et al., 2012), 100 % of loci that were different in parents were accurately called as heterozygous across eight F_1s (288 of 288

loci across all eight F_1 fish). SNP genotypes of the same cross type similarly had 100 % heterozygosity in F_1s (Schluter et al., 2021). I also have no reason to suspect any systematic differences in the genetic constitution of wild fish (i.e., F_0s) used to found field vs. lab studies. In light of the above, it seems most likely that the results presented herein reflect biology rather than methodology.

Table 7.1: Summary of data sources.

species	design	study	population	method	gen	env.	*n* fish	*n* loci ± [1SD]
ben × lim	biparental	present	Paxton	microsatellites	F_2	lab	89	99.4 ± 6.2
ben × lim	biparental	Conte et al. (2015)	Paxton	SNP array	F_2	pond	636	64.0 ± 0.0
ben × lim	biparental	present	Priest	microsatellites	F_2	lab	92	22.8 ± 1.3
ben × lim	biparental	Conte et al. (2015)	Priest	SNP array	F_2	pond	412	90.0 ± 0.0
ben × lim	8 × F_0	Arnegard et al. (2014)	Paxton	SNP array	F_2	pond	615	62.5 ± 17.9
ben × lim	4 × biparental	Rennison et al. (2019)	Paxton	GBS (RAD)	F_2 & F_3	pond	649	85.1 ± 34.0
marine × fresh	biparental	Schluter et al. (2021)	LCR × Cranby	SNP array	F_2 & F_3	pond	723	120.3 ± 5.6
marine × fresh	biparental	Rogers et al. (2012)	LCR × Cranby	microsatellites	F_2	lab	374	58.2 ± 4.2

Benthic × limnetic crosses

I obtained data from three sources for the pond-raised benthic × limnetic hybrids.

Conte et al. (2015) generated a single F_1 family from each of the Priest and Paxton Lake species pairs. Both were founded by a single wild benthic female and a limnetic male that were collected and crossed in 2009. 35 adult F_1 Paxton Lake hybrids and 25 adult F_1 Priest Lake hybrids were released into separate ponds where they bred naturally to produce F_2 hybrids. F_2 adults were collected over one year later. 407 F_2s were genotyped from the Paxton Lake pond, and 324 F_2s were genotyped from the Priest Lake pond. Genotyping used a SNP microarray (Jones et al., 2012*a*), and 246 SNPs were found in the Paxton cross, and 318 were found in the Priest cross.

Arnegard et al. (2014) conducted a pond experiment with eight F_0 grandparents from Paxton Lake. Two crosses used a limnetic female and two crosses used a benthic female. Five F_1 males and five F_1 females from each family were added to the ponds in March 2008 where they bred naturally. 633 juvenile F_2s were collected in October of that same year and genotyped at 408 SNPs.

Bay et al. (2017) genotyped 383 F_2 hybrid females between Paxton Lake benthics and limnetics at 494 SNPs. Fish are from several crosses and study designs. One used a cross with four unique F_{0s} that were used to produce two F_1—one with a limnetic as dam and the other with a benthic as dam. A second had eight unique F_0s, where two F_1 crosses were in each direction. These two crosses used wild fish collected in 2007. A third set of crosses was done in 2009, one in each direction. The two F_1 families were released into separate ponds. Since the goal of the authors' study was to examine the genetics of mate choice, a large number of F_2 females were genotyped at a small number of microsatellite markers to determine whether a genotyped egg or fry was their offspring. 383 F_2 females were identified in this parentage analysis and were then genotyped at SNP markers. A total of 302 females were assigned to families with 10 or more full-sibs (which was necessary for linkage mapping), and these 302 females comprise my dataset.

Finally, Rennison et al. (2019) conducted a study with four unique Paxton Lake benthic × limnetic F_1 hybrid families that were each split between two ponds. One pond in each pair contained a cutthroat

trout predator (heterozygosity did not differ across pond types and data are pooled across all pairs and pond types). Wild fish were caught in 2011 and F_1s were released in 2012. Fish bred naturally and juvenile F_2s were sampled in September of that same year. F_3 hybrids were collected in September 2013. 50 fish from each pond and hybrid generation were genotyped at 2243 SNPs using restriction-site associated DNA sequencing (genotyping-by-sequencing).

The Paxton and Priest Lake laboratory cross data are original to this study. Crosses used a single wild-caught benthic female fish and a single wild-caught limnetic male fish as F_0 progenitors. Wild fish were crossed in 2003. Sibling mating of F_1 hybrids was used to produce a single F_2 hybrid family for analysis, and fish were raised in glass aquaria and fed *ad libitum*. 92 Priest Lake F_2s were genotyped at 85 microsatellite markers, and 86 Paxton Lake F_2s were genotyped at 216 microsatellite markers.

Marine $\times$ freshwater crosses

Schluter et al. (2021) conducted a pond experiment with anadromous (hereafter simply 'marine') $\times$ freshwater hybrids. This study crossed a marine female from the Little Campbell River, BC, with a freshwater male from Cranby Lake, BC. Over 600 juvenile F_2 hybrids were introduced into the ponds directly in August 2006. F_2s overwintered with an estimated over-winter survival rate of approximately 86 % (from mark-recapture). (Simulations designed to capture the experimental design, mortality, and sub-sampling procedures indicate that 86 % survivorship, if caused by viability selection on hybrid incompatibilities, leads to a mean heterozygosity value of 51.8 % $\pm$ 0.5 [SD]; not shown.) In spring 2007, surviving F_2s bred and were genotyped at 1,294 biallelic SNP markers. 500 of their F_3 hybrid offspring were collected in October 2007 and were genotyped with the same methodology.

The data for the laboratory comparison to the Schluter et al. (2021) data were originally published by Rogers et al. (2012). The population used a single Little Campbell River female as the F_0 dam and a single Cranby Lake male as the F_0 sire. Wild adult fish were captured in 2001 to generate a single F_1 hybrid family. Four F_2 crosses were made from individuals from this F_1 family (eight F_1 parents total) for a total of 374 genotyped F_2s. Fish were genotyped at 96 microsatellite markers.

7.2.2 Marker filtering and estimating heterozygosity

For each dataset, I restricted my analysis to loci where the F_0 progenitors of a given F_2 family had no alleles in common (e.g., all BB in benthic F_0s and all LL in limnetic F_0s) and where all F_1 hybrids were heterozygous for ancestry. GBS data were filtered to 20$\times$ coverage. Final sample sizes of fish and markers are given in Table 7.1. In all cases, the sex chromosome (chromosome 19) was not analyzed.

Because some studies have more individuals than loci, and others have more loci than individuals, I analyze heterozygosity both in individuals (averaged across loci) and at loci (averaged across individuals). I retained individuals for which at least 20 loci were genotyped, and retained loci for which at least 20 individuals were genotyped.

Ancestry proportion, directional selection, and segregation distortion reduce the expected heterozygosity below 50 %. I account for this and base my main inference on estimates of *excess* heterozygosity. Excess heterozygosity was calculated as observed heterozygosity (p_{AB}) minus expected heterozygosity ($2p_{\mathrm{A}}p_{\mathrm{B}}$, where

p_A and p_B are the frequencies of both ancestries at the locus or in the individual's genome). My conclusions are unchanged, however, if an uncorrected 'observed' heterozygosity is used as the response variable (Fig. F.4) or if the expected heterozygosity is adjusted for sample size (i.e., multiplying by $\frac{2N}{2N-1}$ following Irwin et al. 2018).

7.2.3 Data analysis

I evaluated whether excess heterozygosity differed between crosses using linear models. I first used simple *t*-tests to evaluate whether heterozygosity differed from the Hardy-Weinberg expectation (i.e., whether excess heterozygosity was significantly different from 0). This was done because studies are separate experiments and this is a useful way to evaluate patterns of heterozygosity. To avoid basing conclusions on such indirect comparisons, I also compared aquarium and pond studies directly in common linear models. Linear models had excess heterozygosity—of individuals or loci—as the response variable and environment (lab or pond) as a categorical predictor. For the benthic $\times$ limnetic data, 'lake' was included as a second categorical predictor. A lake $\times$ environment interaction term was non-significant and was omitted from the final model. Thus, two models were run: one for the benthic $\times$ limnetic cross type, and another for the marine $\times$ freshwater cross type.

7.2.4 Heuristic model

I used a simple heuristic model to infer the strength of selection on incompatibilities that is sufficient to cause the observed excess heterozygosity. I assume that incompatibilities only act between pairs of loci and that fitness is based only on the genotype of individual F_2 hybrids. There is no directional selection and fitness is not frequency- or density-dependent. I also assume that all possible pairs of incompatible loci reduce an individual's fitness. By assumption, I ignore higher-order interactions among loci (i.e., incompatibilities between three or more loci).

I consider two broad incompatibility architectures, recessive (Fig. 7.3A, left) and additive (7.3A, right). The recessive model assumes that only loci with opposite homozygous ancestry (i.e., homozygous at one locus for BB and at the other locus for LL) interact, whereas the additive model allows heterozygous loci to be involved in incompatibilities. The additive model is symmetric for simplicity and because it is plausible that populations could easily cross shallow fitness valleys such as those that typically underlie individual ecological incompatibilities.

The models simply sum the count of each incompatibility type, and penalize the individual's fitness according to a linear function where the penalty is determined by s—the selection coefficient acting against incompatibilities of opposite homozygous ancestries (i.e., BB-LL or FF-MM incompatibilities). For simplicity I use notation of benthic ('B') and limnetic ('L') ancestries, but this is arbitrary and one could substitute marine or freshwater. I assume the strength of incompatibilities involving heterozygous loci is half of that involving only homozygous loci (see below). The selection coefficient, s, can be interpreted as the reduction in fitness caused by individual incompatibilities. For example, an s value of 0.002 implies that each BB-LL or FF-MM incompatibility reduces an individual's fitness by 0.2 %.

For a case where all incompatibilities are assumed to be recessive, the fitness of an individual hybrid is calculated as:

$$W_i = 1 - n^2 \cdot p_{\mathrm{BB}} \cdot p_{\mathrm{LL}} \cdot s, \tag{7.1}$$

where W_i is the fitness of the i^{th} individual F_2 hybrid, n is the number of unlinked loci, p_{BB} & p_{LL} are the frequencies of homozygous ancestries among genotyped loci (here benthic or limnetic, but could also be marine and freshwater), and s is the selection coefficient acting against each BB-LL or FF-MM incompatibility. If an individual has 4 loci where ancestry is 1 × BB, 1 × LL, and 2 × BL, this individual has one possible pairwise incompatibility—the lone BB locus with the lone LL locus. If the individual has 5 loci with ancestries 2 × BB, 1 × LL, and 2 × BL, this individual has 2 pairwise incompatibilities—each BB with the lone LL. The fitness of a given individual is reduced by a given amount, s, for each incompatibility, so fitness in the first former case is $1 - s$ and fitness in the latter case is $1 - 2s$. I assume fitness is linear for simplicity, but note that this causes the model to return negative fitness values and thus fail under extended parameter combinations.

I next consider an additive model where heterozygous loci interact with homozygous loci to influence hybrid fitness. For systems where niche divergence is additive, as it likely is in threespine stickleback (Arnegard et al., 2014), an additive model might be more realistic than a recessive one. Under the additive model, an individual's fitness is calculated as:

$$W_i = 1 - n^2 \cdot p_{\mathrm{BB}} \cdot p_{\mathrm{LL}} \cdot s - (n^2 \cdot p_{\mathrm{BB}} \cdot p_{\mathrm{BL}} + n^2 \cdot p_{\mathrm{LL}} \cdot p_{\mathrm{BL}}) \cdot \frac{s}{2}, \tag{7.2}$$

where terms are as in Eqn. 7.1, and p_{BL} is the frequency of loci with heterozygous ancestry. The first term in the equation quantifies homozygote-homozygote incompatibilities (i.e., those between BB and LL), and the second term quantifies homozygote-heterozygote incompatibilities (i.e., those between BB and BL & LL and BL). This model assumes that the strength of selection on homozygous-heterozygous incompatibilities is half that acting on incompatibilities between loci of opposite homozygous ancestry. Consider again an individual with four loci, with 1:2:1 BB:BL:LL frequencies. The individual has 1 homozygote-homozygote incompatibility (BB-LL) and four heterozygote incompatibilities (BB-BL_1, LL-BL_2, LL-BL_1, and LL-BL_2).

I used this heuristic model as the basis of simple simulations to evaluate the connection between the above fitness functions and observed patterns of excess heterozygosity. These simulations interpret fitness from Eqns. 7.1 & 7.2 as survival probability. In simulations, I assumed that F_1 hybrid stickleback typically have between 1 and 1.5 recombination events per chromosome ($2n$ chromosomes = 42; Roesti et al. 2013), and therefore assume that F_2 hybrids inherit 50 unlinked haplotype blocks (hereafter referred to as the 50 genotyped 'loci'; approximately 1.2 recombination events per chromosome). I generated individuals with 50 loci where each locus had a 25 % probability of having ancestry 'BB', a 25 % probability of having 'LL', and a 50 % probability of being heterozygous, 'BL'. I calculated each individual's fitness under both Eqns. 7.1 and 7.2 then implemented probabilistic selective mortality according to an individual's predicted fitness (using random numbers). Finally, I recorded the mean excess heterozygosity of survivors. I repeat these simulations for various values of s to determine the strength of selection against incompatibilities that

returns estimates of excess heterozygosity that match those in the empirical data. As noted above, I assume incompatibilities only act between pairs of loci (i.e. they are digenic rather than trigenic, etc.) even though trigenic interactions are likely to be much more common than digenic ones (Kuzmin et al. 2018). Therefore, the resulting estimates of the minimum s required to observe a given amount of heterozygosity are very conservative and likely overestimated by one or more orders of magnitude.

7.3 Results and discussion

7.3.1 Patterns of excess heterozygosity

I begin this section by evaluating patterns of excess heterozygosity in aquaria and ponds separately for both benthic $\times$ limnetic and marine $\times$ freshwater crosses. These comparisons are indirect but are justified by the fact that the data are from separate studies measured in different years. I then make a direct comparison between these two environments in using linear models. Conclusions are consistent between approaches.

Aquarium-raised crosses

Mean individual excess heterozygosity was not significantly different from zero in any aquarium-raised stickleback cross (see left side [red] of panels in Fig. 7.2). Specifically, t-tests evaluating whether excess heterozygosity was significantly different from zero failed to reject the null hypothesis for both benthic $\times$ limnetic crosses (mean & 95 % CI Paxton Lake—β = 0.014 [-0.010, 0.037], t_{88} = 1.16, P = 0.25; Priest Lake—β = 0.0098 [-0.014, 0.034], t_{89} = 0.81, P = 0.42). The aquarium-raised marine $\times$ freshwater cross was similar (β = -0.0047 [-0.013, 0.00396], t_{373} = 0.28, P = 0.28).

The lack of excess heterozygosity in aquaria is not surprising given what is known about 'intrinsic' hybrid incompatibilities in stickleback. Previous studies of benthic $\times$ limnetic hybrids have found no evidence for intrinsic inviability in F_2 benthic $\times$ limnetic stickleback crosses, either in embryo development and hatching success McPhail (1984, 1992); Hatfield and Schluter (1999) or lifetime fitness (Hatfield and Schluter, 1999). In a recent review summarising the literature on reproductive isolation in threepsine stickleback, Lackey and Boughman (2017) report that 'intrinsic' barriers are typically weak to nonexistent. Both marine (i.e., anadromous) $\times$ freshwater and benthic $\times$ limnetic crosses had no evidence for intrinsic inviability in this review (Lackey and Boughman, 2017). By contrast, the authors found evidence for hybrid *ecological* inviability in both systems (Lackey and Boughman, 2017).

Pond-raised crosses

Mean individual excess heterozygosity exceeded zero in each pond-raised cross (see right side [blue] of panels in Fig. 7.2). This was the case for both Paxton and Priest Lake hybrids for the benthic $\times$ limnetic crosses (mean & 95 % CI Paxton Lake—β = 0.029 [-0.024, 0.033], t_{2222} = 13.16, $P < 0.0001$; Priest Lake—β = 0.039 [0.029, 0.048], t_{411}, $P < 0.0001$), as well as the marine $\times$ freshwater cross (β = 0.034 [0.027, 0.041], t_{722} = 9.33, $P < 0.0001$). Thus, there is support for the hypothesis that ponds select for heterozygosity.

Direct comparison between aquarium- and pond-raised crosses

I evaluated the difference between aquarium- and pond-raised crosses using simple linear models. For the benthic $\times$ limnetic crosses, I used a linear model where lake (Paxton or Priest), environment (aquaria or pond), and their interaction, were predictors. The interaction term was non-significant and was dropped leaving the two main effects. I found that individual mean excess heterozygosity was 2.2 % higher in pond-raised benthic $\times$ limnetic hybrids than in aquarium-raised hybrids ($\hat{\beta}$ = 0.022 $\pm$ 0.0083 [SE; units are heterozygote frequency], $F_{1,2509}$ = 6.89, P = 0.0087) (Fig. 7.2A; also see Fig F.5 for plots of individual hybrid index and heterozygosity). The 'lake' term was non-significant (P = 0.11). For the marine $\times$ freshwater crosses, I found that individual mean excess heterozygosity was 3.9 % higher in pond-raised fish than in aquarium-raised fish ($\hat{\beta}$ = 0.039 $\pm$ 0.0059 [SE], $F_{1,1095}$ = 41.88, P = 1.46 $\times 10^{-10}$) (Fig. 7.2B). I repeated these analyses using loci as the unit of replication, rather than individuals, and found virtually identical patterns to the analysis of individuals (Fig. 7.2C&D; Fig F.6). These results are consistent with the hypothesis that hybrid incompatibilities act more strongly in the wild than in the lab and implies that individuals with a greater number of incompatibilities are more likely to die than those with fewer.

7.3.2 Heuristic model

Analysis of the heuristic model revealed that, as expected, fitness (W) is lower when incompatibilities act additively rather than recessively (compare black and grey lines in Fig. 7.3B) and when F_2 individuals are more homozygous (compare dashed to solid lines in Fig. 7.3B). I used simulations of selective mortality to estimate the per-incompatibility s value that is sufficient to generate 3 % excess heterozygosity, which is the mean lab-pond difference of the two crosses. For the recessive model, I found that I observed a mean of 3 % excess heterozygosity when $s \geq 0.0036$. Under the additive model, I found that mean 3 % excess heterozygosity resulted when $s \geq 0.0017$. These results imply that if each two-way incompatibility reduces survival probability only by 0.0036 (i.e., 0.36 %) under a recessive model, or 0.0017 (i.e., 0.17%) under an additive incompatibility model, this would be sufficient to cause the observed patterns of excess heterozygosity. As noted above, it is possible that these values are substantial over-estimates because they only account for two-way incompatibilities.

7.3.3 Alternative explanations for excess heterozygosity

Although I framed my predictions on the hypothesis of hybrid incompatibilities, several other mechanisms could possibly underlie the observed pattern of excess heterozygosity. These alternative mechanisms—heterosis and dominance—involve processes operating at single loci rather than interactions among loci. Ultimately, the data cannot directly distinguish between single-locus processes like heterosis and multi-locus processes like incompatibilities, so indirect inference is required to evaluate the plausibility of causal processes.

Heterosis refers to the case where heterozygosity at a given locus is favoured *per se*, rather than as a means to escape incompatibilities caused by interactions involving homozygous alleles. Typically, heterosis is caused by the dominance of a fit allele over a less fit one. If individual loci were overdominant only

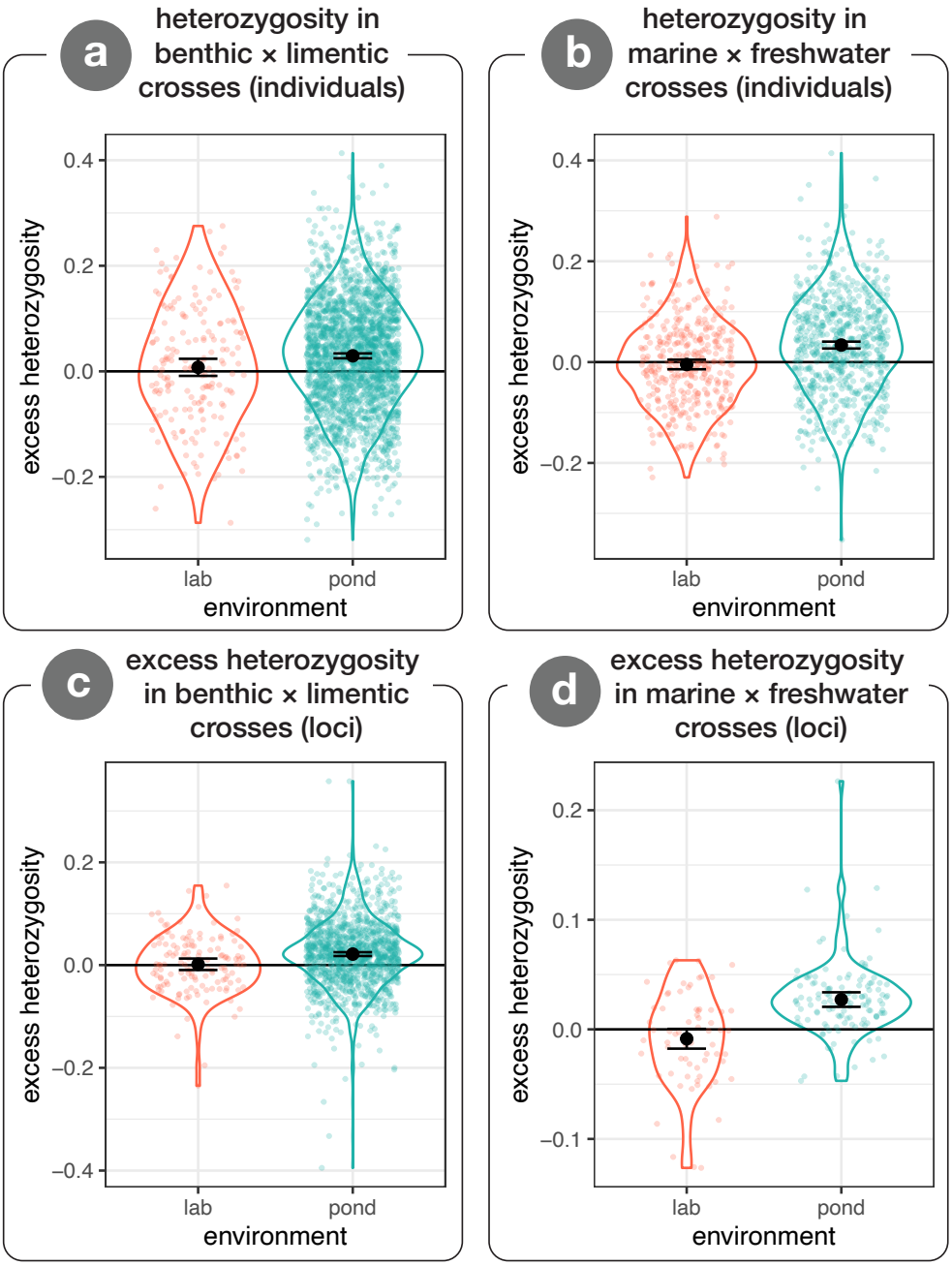


Figure 7.2: **Patterns of heterozygosity in stickleback hybrids across environments**. Panels (A) and (B) show excess heterozygosity for individual hybrids (averaged across loci), while panels (C) and (D) show excess heterozygosity for individual loci (averaged across individuals). Points represent the residual mean ± 95 % CI (from `visreg`). Violin overlays show the full distribution of the data. Paxton Lake pond data are from several different studies considered together (see Methods and Table 7.1).

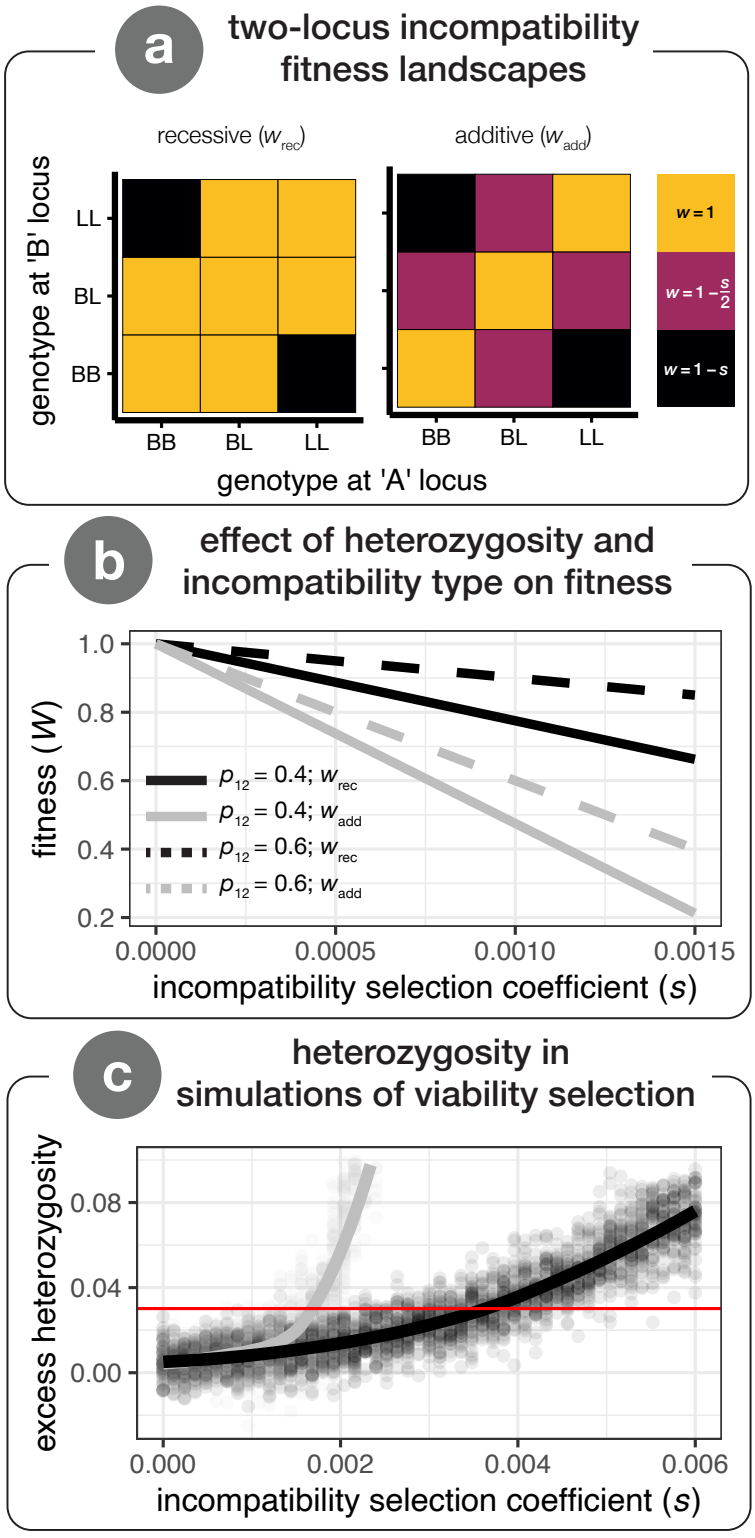


Figure 7.3: **Heuristic two-way incompatibility model to illustrate sufficient selection strengths to generate excess heterozygosity**. Panel (A) is an illustration of two two-locus incompatibility fitness landscape models where colour represents fitness. Panel (B) shows fitness values (W) calculated using Eqn. 7.1 (recessive incompatibility; black) and Eqn. 7.2 (additive incompatibility; grey) for different values of s. For each model, I show fitness for two values of heterozygote ancestry frequency (p_{12}). Panel (C) shows excess heterozygosity in simulations of F_2 hybrids experiencing different selection strengths (using Eqns. 7.1 and 7.2) (see Methods). The horizontal red line indicates average difference in excess heterozygosity between the lab and pond studies ($p_{12} = 0.03$).

under field conditions, then excess heterozygosity could result without any interactions among loci. In the benthic × limnetic crosses, this possibility can be disregarded based on prior knowledge about hybrid fitness in this system. If heterosis were common, then hybrids should have higher fitness than parents. However, F_1 and reciprocal backcross hybrids—which are highly heterozygous—have *lower* growth and/or survival than parent taxa in field experiments (Hatfield and Schluter 1999; Vamosi et al. 2000; Rundle 2002). In the lab, F_1 hybrid growth rate matches the additive expectation of parents (Hatfield, 1997; Hatfield and Schluter, 1999). These patterns simply would not occur if heterosis was a general feature of benthic-limnetic crosses. Less is known about heterosis in marine × freshwater crosses. Further evidence against heterosis comes from the fact that there is no relationship between body size and heterozygosity in the lab in any cross (Fig. F.7; family sizes are too small for a robust analysis in ponds). Thus, heterosis seems an unlikely explanation underlying selection for heterozygosity in stickleback hybrids.

Patterns of excess heterozygosity could also be caused by directional selection and dominance. If heterozygotes were just as likely to survive as the favoured homozygote, this would lead to a higher observed heterozygosity than expected (i.e., $>2p_{\mathrm{A}}p_{\mathrm{B}}$) for loci that are under directional selection. I quantified the strength of directional selection on each locus as $|p_{\mathrm{AA}} - p_{\mathrm{BB}}|$, the absolute difference between the frequencies of both homozygotes. Loci under strong directional selection should have larger values. If dominance was common, then excess heterozygosity should be greater at loci that are under stronger directional selection. I find that the excess heterozygosity of loci is not correlated with this metric of directional selection in either pond-raised benthic × limnetic hybrids nor marine × freshwater hybrids (both $P > 0.15$; Fig. F.8). Thus, because directional selection is not a strong predictor of heterozygosity, it is unlikely that dominance of beneficial alleles alone can explain the observed patterns of excess heterozygosity.

7.4 Conclusion

The main result of this study is that heterozygosity was elevated in F_2 stickleback hybrids raised in the field compared to those raised in aquaria. Because single-locus processes such as heterosis and dominance are unlikely to explain this pattern, I suggest that hybrid incompatibilities are the most likely causal mechanism. Selection on a per-incompatibility basis need not be terribly strong to cause levels of excess heterozygosity similar to what I observed. This finding has implications for our understanding of the genetic basis of extrinsic reproductive isolation. At least for the benthic × limnetic crosses, it is well-established that the species are reproductively isolated and hybrids perform worse than parents under field conditions. Thus, hybrid incompatibilities might underlie extrinsic post-zygotic isolation in this classic case of 'ecological speciation' (McKinnon and Rundle, 2002). Moreover, the analysis gives insight into the number of genes that might underlie speciation. The data presented here suggests that many genes of small effect (i.e., where their interactions reduce fitness by considerably less than 1 %), located throughout the genome, underlie reproductive isolation between ecologically divergent populations of threespine stickleback.

This study used crosses with different F_0 progenitors, genotyped with different methodologies, for its main comparison. Although this is less than ideal, there is no obvious mechanism that would cause the lab-raised fish to exhibit significant excess heterozygosity. In light of this, a valid though less conservative

approach would simply have tested if pond-raised crosses exhibit excess heterozygosity, which they almost invariably do. The fact that this pattern—excess heterozygosity in pond-raised hybrids—repeats across so many pond experiments from different years, might be seen as an even stronger conclusion than one based on a single comparison in a single year.

The evidence presented herein is consistent with the hypothesis that extrinsic hybrid incompatibilities, which operate between parental alleles at different loci, underlie extrinsic post-zygotic isolation in this system. To the extent that this study documents the existence of environment-specific hybrid incompatibilities, it cannot identify their underlying mechanisms. Experiments that directly manipulate individual phenotypes, or their interactions with the environment, are needed to establish such causality. Such studies represent an exciting new frontier for empirical research into the mechanisms of ecological speciation.

Chapter 8

Conclusion

8.1 Overview

My thesis used theory, data synthesis, and experiments to generate hypotheses, establish patterns, and test predictions about the evolutionary ecology of hybridization. The narrative thread throughout this thesis surrounded hybrid incompatibilities that affect the ecological performance of hybrids. I have tried to explore this topic using a variety of techniques and data types.

The first three data chapters of my thesis used theoretical approaches and previously published data. Chapter 2 used simulations to clarify how adaptation from standing variation affects progress toward speciation and used analytical models to arrive at generalities about when we should expect populations to use the same vs. different alleles for adaptation. Chapter 3 synthesized data from the literature to illustrate that divergent quantitative traits are often dominant and 'mismatched' in F_1 hybrids, and used data from a large-scale field experiment to illustrate that this mismatch has negative fitness consequences. Chapter 4 used data from the literature to provide support for a key assumption of models of hybrid fitness—that mutations used for adaptation are typically pleiotropic. These three chapters established a theoretical and empirical framework for studying the causes and consequences of ecological hybrid incompatibilities. I leveraged this framework in the final three data chapters which test theoretical predictions in the lab and field.

The final three data chapters of my thesis use threespine stickleback fish (*Gasterosteus aculeatus*) to test key predictions about ecological hybrid incompatibilities. Chapter 5 found support for the theoretical prediction that hybrid trait mismatch evolves in a manner that is positively associated with the magnitude of phenotypic divergence between parent populations. Chapter 6 used a pond experiment to illustrate that parallel phenotypic evolution can lead to the evolution of hybrid breakdown due to 'mutation-order' processes in both benthics and limnetics, but also found a surprising amount of heterosis. Finally, Chapter 7 found that the genetic signature of hybrid incompatibilities—excess heterozygosity—is reliably observed in pond-raised F_2 stickleback hybrids but is not observed in the lab. These three chapters collectively indicate that ecological hybrid incompatibilities might be an important form of post-zygotic reproductive isolation separating natural stickleback populations.

8.2 Areas requiring research

By illustrating that traits are often mismatched in hybrids and that hybrid incompatibilities can have an important ecological dimension, my thesis work has illuminated several research directions worth addressing. As I expressed at the outset, I hoped that this thesis would raise more questions than it answered. Though I

will leave it to the reader to decide whether this was accomplished, below I highlight several of the key areas that I see as high-priority for future research.

8.2.1 Mechanisms underlying ecological incompatibilities

Before my thesis, scientists had been making rapid progress in identifying (putative) mechanisms underlying ecological incompatibilities (Vinšálková and Gvoždík, 2007; Matsubayashi et al., 2010; Cooper et al., 2018; Martin and Wainwright, 2013; Keagy et al., 2016; Arnegard et al., 2014). These pioneering studies generated compelling and testable hypotheses about the mechanisms through which trait interactions affect fitness, but it was beyond their scope to quantify 'mismatch' more directly and directly link it to fitness. My thesis built on this strong foundation in several ways. First, in Chapter 3 I quantified individual-level mismatch in sunflowers and related it directly to fitness via seed count. Second, in Chapter 7 I showed that individuals with high 'genetic mismatch' (i.e., high homozygosity for a given hybrid index) were selected against under semi-natural conditions but not in the lab. While I think the above two conclusions are valuable and novel contributions to the literature, it is important to acknowledge their main limitation. Specifically, the link between mismatch and fitness is correlational because I cannot decouple environment-specific endogenous selection—intrinsic conflicts preventing development or gametogenesis—from environment-specific exogenous selection—selection imposed by the biotic or abiotic environment. Nor can I conclusively identify any agents of selection that link mismatch to fitness.

Future work on ecological incompatibilities should strive to identify the mechanisms of ecological selection underlying them. To accomplish this, I envision two main approaches. First, one can use population-level manipulations to change how groups of hybrids experience selection. For example, Rennison et al. (2019) split hybrid families into some ponds that contained trout and other ponds that didn't—such an approach could test for trout-mediated selection on incompatibilities if one hypothesized that it might exist. Rundle et al. (2003) conducted a similar experiment also removing predatory insects. Such approaches have the advantage that treatments are relatively easy to apply, but the disadvantage that they reduce sample size to the number of replicate enclosures (e.g., ponds, mesocosms, tanks) across which organisms are divided. For some groups of organisms, like fish, this is likely the only feasible approach because individuals cannot be reliably recaptured and manipulated during their lifetimes.

A second approach to studying mechanisms of ecological incompatibilities involves individual-level manipulations to randomize the selection regimes of individual organisms in a common garden. For example, one can provide supplemental resources or exclude mutualists (e.g., pollinators) to confirm a role for any particular selective agent that is hypothesized to underlie incompatibility. In my planned post-doctoral work, I plan to experimentally manipulate hand-pollination across flowers in individual F_2 hybrids between bird-pollinated *Penstemon centranthifolius* and wasp-pollinated *P. spectabilis*. Individual plants that are unable to attract and/or interact with pollinators (i.e., receive or deposit pollen) will have low seed production in the open-pollination treatment compared to those who are relatively well-suited to interacting with pollinators. Hand pollinations can confirm that low-fitness individuals have low fitness because of their interactions with pollinators, rather than some 'intrinsic' unfitness. Thus, a prediction is that the fitness effect of hand-pollination will be correlated with mismatch. If possible, one might also manipulate phenotypes directly to

decouple them from the genetic basis (Sinervo and Svensson, 2002). In these approaches using natural or experimental hybrids, hypotheses can be robustly tested by examining whether the manipulations affected selection on excess heterozygosity, the mismatch-fitness relationship, and/or patterns of hybrid breakdown.

8.2.2 Theoretical importance of ecological incompatibilities

The work in this thesis generally concerns early generation—F_1, F_2, or BC_1—hybrids, but future work must extend the findings herein to later generations. Models of hybrid zones or sympatric populations that consider how various genetic architectures influence the probability of collapse into a swarm are needed to resolve the efficacy of both intrinsic and extrinsic incompatibilities for maintaining reproductive isolation. One key difference between intrinsic incompatibilities and ecological incompatibilities is that extrinsic fitness landscapes are likely more dynamic and subject to a greater magnitude of negative frequency-dependent and density-dependent selection—more intrinsic incompatibilities are more likely subject to positive frequency-dependence. A truly intrinsic incompatibility, where selection is consistent across all plausible environments, could theoretically be removed (i.e., one set of compatible alleles reaches fixation) from a region of hybridzation (e.g., a hybrid zone) by selection rather easily if it involves only a small number of loci (Xiong and Mallet, 2021). If these incompatibilities are all that keep species apart, reproductive isolation could break down following the loss of one set of alleles. If the alleles underlying an ecological incompatibility reach low frequency, however, the lower frequency type might be increasingly favoured by selection. (Testing whether selection on ecological incompatibilities is frequency- or density-dependent is an exciting empirical opportunity). If selection is frequency-dependent, incompatibilities acted on by ecologically-mediated selection might be more robust in the face of potentially homogenizing gene flow. Future theory on this subject, perhaps integrating approaches from adaptive dynamics (wherein frequency-dependent selection is central [Dieckmann and Doebeli 1999]), will be valuable.

To better understand the role of ecological incompatibilities in speciation, we must also expand our phenotypic models of selection against hybrids. A given reduction in F_1 fitness could result from two general phenotypic mechanisms: relative intermediacy—where intermediate phenotypes are selected against due to their being no intermediate niche—or mismatch—where particular combinations of traits render hybrids unfit in any niche including a possible intermediate one. For a given reduction of F_1 fitness, is the maintenance of species more likely when one or the other mechanism is responsible? Selection against the F_1 is what it is, regardless of the underlying phenotypes or genetics. But patterns of selection against second generation hybrids and beyond (e.g., backcrosses and $F_{\geq 2}$s) will differ depending on whether traits are additive or dominant. Conducting simulations wherein population divergence is underpinned by additive vs. dominant loci—and where dominance differs in the same or different directions among traits—is likely the best way to determine how trait mismatch affects the persistence of reproductively isolated populations, and to ascertain the importance of trait mismatch for speciation.

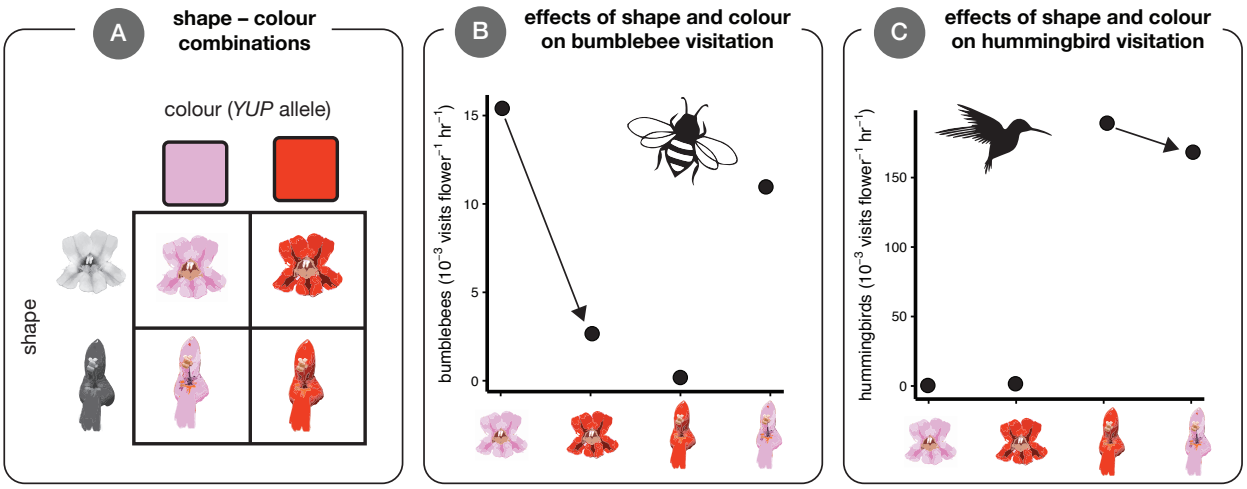


Figure 8.1: **A possible ecological incompatibility in monkeyflowers.** *Mimulus* (syn. *Erythranthe*) *lewisii* has pink petals and is pollinated by bumblebees, whereas *M. cardinalis* has red-coloured petals and is pollinated by hummingbirds. The species differ in a QTL at the YUP locus that confers flower colour differences. When the *cardinalis* YUP allele is introgressed into the *M. lewisii* background pollination by bumblebees decreases bee visitation by 83 % (arrow in B). By contrast introgressing the *lewisii* YUP allele into cardinalis only decreases hummingbird visitation by 10 % (arrow in panel C). This could represent an asymmetric incompatibility that is only identifiable in the field. Images digitized from Schemske and Bradshaw (1999) and data taken from Bradshaw and Schemske (2003).

8.3 A proposal to slightly refine the language of speciation

Increasing evidence, including that presented in this thesis, indicates that we must change the way we use the term 'hybrid incompatibility'. In this brief section, I argue two main points. First, the Bateson-Dobzhansky-Muller mechanism should be seen as a general mechanism of post-zygotic isolation, rather than only resulting from endogenous conflicts in hybrid genomes. Second, the general intrinsic vs. extrinsic classification has largely outlived its usefulness and its continued use is more harmful than productive.

As literature on speciation has accumulated in the past decades, the language we use to describe its underlying processes and mechanisms has grown. Noting this, Harrison (2012) cautioned against introducing new terms and re-configuring old definitions because often new language can reflect the idiosyncrasies of the systems we are most familiar with and many new terms are invented for describing long-understood processes. In their book, Coyne and Orr (2004) describe the mechanism of Bateson-Dobzhansky-Muller (Bateson, 1909; Dobzhansky, 1937; Muller, 1942) hybrid incompatibilities only when discussing embryo inviability and hybrid sterility. The BDMI mechanism is specifically included as a subsection of their discussion of 'intrinsic' postzygotic isolation. Even in books dedicated to the study of ecological speciation, incompatibilities are seldom if ever raised as the genetic basis of extrinsic postzygotic isolating barriers (Nosil, 2012). The sorts of mismatch-based incompatibilities described in this thesis act via an identical genetic mechanism to the more traditional 'intrinsic' BDM incompatibilities in that they involve epistasis between two or more loci. The fact that they might differ in their average fitness effects only implies that they differ in degree and not kind.

Although incompatibilities subject to ecologically-mediated selection are likely weak generally, this should not be seen as a bright line differentiating them from 'intrinsic' incompatibilities. Indeed, as scientists study ecological incompatibilities further, we are likely to discover many 'intrinsic' incompatibilities that have weak fitness-effects and many 'ecological' incompatibilities where the fitness consequences are substantial. Some in the latter category likely have already been discovered; for example, introgressing the allele for 'red' flowers from *Mimulus cardinalis* into bee-pollinated *M. lewisii* almost completely reduces visitation from bee pollinators from whom the flowers are adapted to receive pollen and on whose backs they are adapted to deposit it (Fig. 8.1). This substitution does not substantially affect visitation from hummingbirds. Although the fitness consequences of this substitution in terms of seed set were unmeasured, it is reasonable to conclude that a complete lack of pollination likely had a substantial effect on seed set that rivals many ovule- or pollen-sterilizing incompatibilities and thus would have a fitness effect rivalling many sterilizing incompatibilities.

Importantly, many incompatibilities that are called 'intrinsic' are sensitive both to the environmental context and the genetic background (Fig. 8.2). For example, using an experimental approach that mimics hybridization, Ono et al. 2017 used yeast to examine the fitness consequences of combining two alleles that evolved in an experimental evolution experiment and individually confer resistance to a fungicide. In the environment under which the alleles evolved, combining both in a common genetic background caused a reduction in fitness compared to each type examined alone (compare purple to red and blue in Fig. 8.2A. However, in a different environment with a higher concentration of the fungicide, combining these alleles improved fitness rather than reduced it. Thus, even incompatibilities that might appear 'intrinsic' because they reduce growth rate in the lab can have an important environmental context.

The word 'intrinsic' can literally mean 'essential', and incompatibilities described as such are often thought to be permanent or irreversible. However, recent evidence indicates that incompatibilities can also interact in a manner that refutes the hypothesis that they are in fact 'intrinsic'. For example, Guerrero et al. (2017) investigated interactions among different incompatibility-causing alleles in tomato (*Solanum* spp.). The authors compared plants where one region of the genome had hybrid ancestry (i.e., single-introgression lines) to those with two regions of hybrid ancestry (double-introgression lines). The goal was to determine if incompatibilities typically act additively (e.g., if same phenotypic effect, then double-introgression lines are twice as bad as single-introgression lines), synergistically (double-introgression more than twice as bad), or antagonistically (double-introgression less than twice as bad). The authors found that antagonistic interactions were pervasive, and occasionally even found that double mutants had higher fitness than both single mutants. If two incompatibility-causing alleles *improve* hybrid fitness in the presence of one another, then substituting an 'incompatibility' allele into some genetic backgrounds is beneficial. Calling such incompatibilities 'intrinsic' might mislead us into thinking they are invariably deleterious. In fact, whether or not an incompatibility is 'intrinsic' is fundamentally unfalsifiable because there could always be a genetic background or environment that impacts its fitness effect. We can avoid this pitfall by simply referring to the phenotypic effect of incompatibilities. By calling them 'sterility-causing' or 'lethality-causing' incompatibilities rather than intrinsic, we get our point across while remaining agnostic to any environment-dependence.

Coyne and Orr (2004) acknowledge the somewhat fuzzy distinction between intrinsic and extrinsic iso-

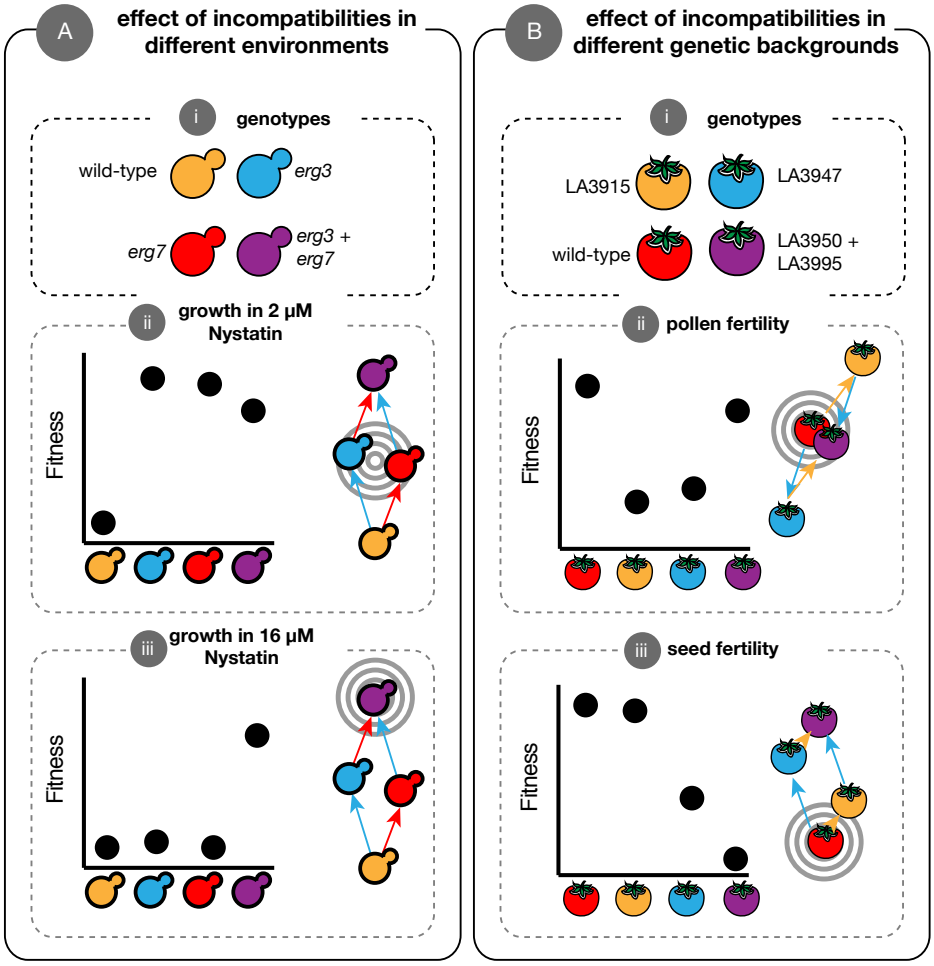


Figure 8.2: **Fitness effects of incompatibility loci vary across environments and genetic backgrounds.** (A) Adapted from Ono et al. (2017). In their experiment, an ancestral strain is divided into (i) different populations of brewer's yeast (*Saccharomyces cerevisiae*) that fixed independent mutations in the ergosterol biosynthesis pathway (ERG3 and ERG5) in response to treatment with the fungicide Nystatin. Nystatin binds ergosterol embedded in the cell membrane, forming channels that cause cells to leach their contents; mutations reduce the quantity of ergosterol in the membrane. (ii) In the presence of 2 μM nystatin, these mutations are individually beneficial because they reduce the target size of nystatin, at the cost of making a less permeable membrane. However, individuals with both mutations have reduced fitness because their membranes are too impermeable, acting as a BDMI. (iii) In higher concentrations of Nystatin, the single mutants are unfit whereas double-mutants have high fitness, illustrating that the incompatibility is environment-specific. (B) Adapted from Guerrero et al. (2017). (i) In this experiment, short chromosomal regions from *Solanum habrochaites* (here LA3947 and LA3915) reduced fitness when introgressed onto the genetic background of *S. lycopersicum*. When combined in a single genetic background (purple tomato), these alleles often restored fitness for at least one fitness component (restored in [ii], further reduced in [iii]).

lating barriers. In light of the emerging evidence, some of which I have briefly touched on above, I advocate for a slight modification to the way we use the term 'hybrid incompatibility'. Namely, we need to recognize that 'incompatibility' need not mean sterility or embryo inviability. Rather, all interactions among alleles of divergent ancestry that interact epistatically to reduce hybrid fitness fully deserve the title of incompatibility—regardless of the reason why they reduce fitness. We should stop adding the preface of 'intrinsic' or 'extrinsic' when describing incompatibilities and be more agnostic unless we are confident in having identified a particular mechanism. Incompatibilities have myriad selective mechanisms and, in my view, dissecting these mechanisms represents some of the most exciting work yet to be done on the evolutionary genetics and evolutionary ecology of hybridization.

Bibliography

Abramoff, M. D., P. J. Magalhaes, and S. J. Ram. 2004. Image processing with ImageJ. Biophotonics International 11:36–42.

Abramowitz, M., and I. A. Stegun. 1972. Handbook of mathematical functions: with formulas, graphs, and mathematical tables, vol. 55. United States Department of Commerce, Washington DC, USA.

Adam, D., N. Dimitrijevic, and M. Schartl. 1993. Tumor suppression in *Xiphophorus* by an accidentally acquired promoter. Science 259:816–819.

Alves, J. M., M. Carneiro, J. Y. Cheng, A. L. de Matos, M. M. Rahman, L. Loog, P. F. Campos, N. Wales, A. Eriksson, A. Manica, T. Strive, S. C. Graham, S. Afonso, D. J. Bell, L. Belmont, J. P. Day, S. J. Fuller, S. Marchandeau, W. J. Palmer, G. Queney, A. K. Surridge, F. G. Vieira, G. McFadden, R. Nielsen, M. Thomas, P. Gilbert, P. J. Esteves, N. Ferrand, and F. M. Jiggins. 2019. Parallel adaptation of rabbit populations to myxoma virus. Science 363:1319–1326.

Anderson, C. J. R., and L. Harmon. 2014. Ecological and mutation-order speciation in digital organisms. The American Naturalist 183:257–268.

Archambeault, S. L., L. R. Bärtschi, A. D. Merminod, and C. L. Peichel. 2020. Adaptation via pleiotropy and linkage: Association mapping reveals a complex genetic architecture within the stickleback *Eda* locus. Evolution Letters 4:282–301.

Arnegard, M. E., M. D. McGee, B. Matthews, K. B. Marchinko, G. L. Conte, S. Kabir, N. Bedford, S. Bergek, Y. F. Chan, F. C. Jones, D. M. Kingsley, C. L. Peichel, and D. Schluter. 2014. Genetics of ecological divergence during speciation. Nature 511:307–311.

Arnold, S. J. 1983. Morphology, performance and fitness. Integrative and Comparative Biology 23:347–361.

Auguie, B. 2019. egg: Extensions for 'ggplot2': Custom Geom, Custom Themes, Plot Alignment, Labelled Panels, Symmetric Scales, and Fixed Panel Size. R package version 0.4.5.

Bachtrog, D., K. Thornton, A. Clark, and P. Andolfatto. 2006. Extensive introgression of mitochondrial DNA relative to nuclear genes in the *Drosophila yakuba* species group. Evolution 60:292.

Barghi, N., R. Tobler, V. Nolte, A. M. Jaksic, F. Mallard, K. Otte, M. Dolezal, T. Taus, R. Kofler, and C. Schloetterer. 2019. Genetic redundancy fuels polygenic adaptation in *Drosophila*. PLoS Biology 17:e3000128.

Barrett, R. D. H., S. M. Rogers, and D. Schluter. 2008. Natural selection on a major armor gene in threespine stickleback. Science 322:255–7.

Barrett, R. D. H., and D. Schluter. 2008. Adaptation from standing genetic variation. Trends in Ecology and Evolution 23:38–44.

Barton, N. 1989. The divergence of a polygenic system subject to stabilizing selection, mutation and drift. Genetical Research 54:59–78.

Barton, N., and G. Hewitt. 1985. Analysis of hybrid zones. Annual review of Ecology and Systematics 16:113–148.

Barton, N. H. 2001. The role of hybridization in evolution. Molecular Ecology 10:551–568.

Barton, N. H., A. M. Etheridge, and A. Véber. 2017. The infinitesimal model: Definition, derivation, and implications. Theoretical Population Biology 118:50–73.

Barton, N. H., and K. S. Gale. 1993. Genetic analysis of hybrid zones. Pages 13–45 *in* R. G. Harrison, ed. Hybrid zones and the evolutionary process.

Bates, D., M. Maechler, B. Bolker, and W. S. 2014. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7.

Bateson, W. 1909. Heredity and variation in modern lights. Pages 85–101 *in* A. C. Seward, ed. Darwin and Modern Science. Cambridge University Press, Cambridge.

Bay, R. A., M. E. Arnegard, G. L. Conte, J. Best, N. L. Bedford, S. R. McCann, M. E. Dubin, Y. F. Chan, F. C. Jones, D. M. Kingsley, D. Schluter, and C. L. Peichel. 2017. Genetic Coupling of Female Mate Choice with Polygenic Ecological Divergence Facilitates Stickleback Speciation. Current Biology 27:3344–3349.e4.

Bell, M. A., and S. A. Foster. 1994. The evolutionary biology of the threespine stickleback. Oxford University Press.

Bolnick, D. I., R. D. H. Barrett, K. B. Oke, D. J. Rennison, and Y. E. Stuart. 2018. (Non)Parallel Evolution. Annual Review of Ecology, Evolution, and Systematics 49:303–330.

Bolnick, D. I., and T. J. Near. 2005. Tempo of hybrid inviability in centrarchid fishes (Teleostei: Centrarchidae). Evolution 59:1754–1767.

Bradshaw, H. D., K. G. Otto, B. E. Frewen, J. K. McKay, and D. W. Schemske. 1998. Quantitative trait loci affecting differences in floral morphology between two species of monkeyflower (*Mimulus*). Genetics 149:367–382.

Bradshaw, H. D., and D. W. Schemske. 2003. Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. Nature 426:176–178.

Bratteler, M., C. Lexer, and A. Widmer. 2006. Genetic architecture of traits associated with serpentine adaptation of *Silene vulgaris*. Journal of Evolutionary Biology 19:1149–1156.

Brawand, D., C. E. Wagner, Y. I. Li, M. Malinsky, I. Keller, S. Fan, O. Simakov, A. Y. Ng, Z. W. Lim, E. Bezault, J. Turner-Maier, J. Johnson, R. Alcazar, H. J. Noh, P. Russell, B. Aken, J. Alföldi, C. Amemiya, N. Azzouzi, J. F. Baroiller, F. Barloy-Hubler, A. Berlin, R. Bloomquist, K. L. Carleton, M. A. Conte, H. D'Cotta, O. Eshel, L. Gaffney, F. Galibert, H. F. Gante, S. Gnerre, L. Greuter, R. Guyon, N. S. Haddad, W. Haerty, R. M. Harris, H. A. Hofmann, T. Hourlier, G. Hulata, D. B. Jaffe, M. Lara, A. P. Lee, I. MacCallum, S. Mwaiko, M. Nikaido, H. Nishihara, C. Ozouf-Costaz, D. J. Penman, D. Przybylski, M. Rakotomanga, S. C. Renn, F. J. Ribeiro, M. Ron, W. Salzburger, L. Sanchez-Pulido, M. E. Santos, S. Searle, T. Sharpe, R. Swofford, F. J. Tan, L. Williams, S. Young, S. Yin, N. Okada, T. D. Kocher, E. A. Miska, E. S. Lander, B. Venkatesh, R. D. Fernald, A. Meyer, C. P. Ponting, J. T. Streelman, K. Lindblad-Toh, O. Seehausen, and F. Di Palma. 2015. The genomic substrate for adaptive radiation in African cichlid fish. Nature 513:375–381.

Breheny, P., and W. Burchett. 2017. Visualization of regression models using visreg. R Journal 9:56–71.

Brodie, E. D. 1992. Correlational selection for color pattern and antipredator behavior in the garter snake *Thamnophis ordinoides*. Evolution 46:1284–1298.

Burton, R. S., C. K. Ellison, and J. S. Harrison. 2006. The sorry state of F₂ hybrids: Consequences of rapid mitochondrial DNA evolution in allopatric populations. The American Naturalist 168:S14–S24.

Carlson, S. M., A. Kottas, and M. Mangel. 2010. Bayesian analysis of size-dependent overwinter mortality from size-frequency distributions. Ecology 91:1016–1024.

Castle, W. E. 1921. An improved method of estimating the number of genetic factors concerned in cases of blending inheritance. Science 54:223.

Chamberlain, S. A., and E. Szöcs. 2013. Taxize: Taxonomic search and retrieval in R.

Charlesworth, B., and K. Hughes. 1998. The maintenance of genetic variation in life history traits. *In* R. Singh and C. Krimbas, eds., Evolutionary Genetics from Molecules to Morphology. Cambridge University Press, Cambridge, UK.

Chevin, L.-M., G. Decorzent, and T. Lenormand. 2014. Niche dimensionality and the genetics of ecological speciation. Evolution 68:1244–1256.

Chevin, L. M., and F. Hospital. 2008. Selective sweep at a quantitative trait locus in the presence of background genetic variation. Genetics 180:1645–1660.

China, V., and R. Holzman. 2014. Hydrodynamic starvation in first-feeding larval fishes. Proceedings of the National Academy of Sciences of the United States of America 111:8083–8088.

Churchill, G. A., and R. W. Doerge. 1994. Empirical threshold values for quantitative trait mapping. Genetics 138:963–971.

Colosimo, P. F., K. E. Hosemann, S. Balabhadra, G. Villarreal, H. Dickson, J. Grimwood, J. Schmutz, R. M. Myers, D. Schluter, and D. M. Kingsley. 2005. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. Science 307:1928–1933.

Colosimo, P. F., C. L. Peichel, K. Nereng, B. K. Blackman, M. D. Shapiro, D. Schluter, and D. M. Kingsley. 2004. The genetic architecture of parallel armor plate reduction in threespine sticklebacks. PLoS Biology 2:e109.

Conte, G. L., M. E. Arnegard, J. Best, Y. F. Chan, F. C. Jones, D. M. Kingsley, D. Schluter, and C. L. Peichel. 2015. Extent of QTL reuse during repeated phenotypic divergence of sympatric threespine stickleback. Genetics 201:1189–1200.

Conte, G. L., and D. Schluter. 2013. Experimental confirmation that body size determines mate preference via phenotype matching in a stickleback species pair. Evolution 67:1477–1484.

Cooper, B. S., A. Sedghifar, W. T. Nash, A. A. Comeault, and D. R. Matute. 2018. A maladaptive combination of traits contributes to the maintenance of a Drosophila hybrid zone. Current Biology 28:2940–2947.

Corbett-Detig, R. B., J. Zhou, A. G. Clark, D. L. Hartl, and J. F. Ayroles. 2013. Genetic incompatibilities are widespread within species. Nature 504:135–137.

Coughlan, J. M., and D. R. Matute. 2020. The importance of intrinsic postzygotic barriers throughout the speciation process: Intrinsic barriers throughout speciation. Philosophical Transactions of the Royal Society B: Biological Sciences 375:20190533.

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

———. 1997. "Patterns of Speciation in Drosophila" Revisited. Evolution 51:295.

———. 2004. Speciation. Sinauer.

Crnokrak, P., and D. a. Roff. 1995. Dominance variance: associations with selection and fitness. Heredity 75:530–540.

Dagilis, A. J., M. Kirkpatrick, and D. I. Bolnick. 2019. The evolution of hybrid fitness during speciation. PLoS Genetics 15:e1008125.

David, J. R., P. Gibert, G. Pétavy, and B. Moreteau. 2002. Variable modes of inheritance of morphometrical traits in hybrids between *Drosophila melanogaster* and *Drosophila simulans*. Proceedings of the Royal Society B: Biological Sciences 269:127–135.

Demuth, J. P., and M. J. Wade. 2007. Population differentiation in the beetle *Tribolium castaneum* . I. Genetic architecture. Evolution 61:494–509.

Dieckmann, U., and M. Doebeli. 1999. On the origin of species by sympatric speciation. Nature 400:354–357.

Dobzhansky, T. 1937. Genetics and the Origin of Species. Columbia University Press, New York.

Dominigues, A., and J. Albornoz. 1987. Environment-dependent heterosis in *Drosophila melanogaster*. Génétique, Sélection, Évolution 19:37–48.

Drès, M., and J. Mallet. 2002. Host races in plant-feeding insects and their importance in sympatric speciation. Philosophical Transactions of the Royal Society B: Biological Sciences 357:471–492.

East, E. M. 1916. Inheritance in Crosses between Nicotiana langsdorffii and Nicotiana alata. Genetics 1:311.

Edmands, S. 1999. Heterosis and outbreeding depression in interpopulation crosses spanning a wide range of divergence. Evolution 53:1757–1768.

Eyre-Walker, A., and P. D. Keightley. 2007. The distribution of fitness effects of new mutations. Nature Reviews Genetics 8:610–618.

Fiévet, J. B., T. Nidelet, C. Dillmann, and D. de Vienne. 2018. Heterosis is a systemic property emerging from non-linear genotype-phenotype relationships: Evidence from in vitro genetics and computer simulations. Frontiers in Genetics 9:159.

Fisher, R. A. 1930. The Genetical Theory of Natural Selection. Oxford University Press, Oxford, UK.

———. 1931. The Evolution of Dominance. Biological Reviews 6:345–368.

Fishman, L., P. M. Beardsley, A. Stathos, C. F. Williams, and J. P. Hill. 2015. The genetic architecture of traits associated with the evolution of self-pollination in *Mimulus*. New Phytologist 205:907–917.

Fishman, L., and A. L. Sweigart. 2018. When Two Rights Make a Wrong: The Evolutionary Genetics of Plant Hybrid Incompatibilities. Annual Review of Plant Biology 69:707–731.

Fraïsse, C., P. A. Gunnarsson, D. Roze, N. Bierne, and J. J. Welch. 2016. The genetics of speciation: Insights from Fisher's geometric model. Evolution 70:1450–1464.

Funk, D. J., P. Nosil, and W. J. Etges. 2006. Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. Proceedings of the National Academy of Sciences of the United States of America 103:3209–3213.

Gavrilets, S. 2004. Fitness Landscapes and the Origin of Species. Princeton University Press, Princeton, NJ.

Gompert, Z., E. G. Mandeville, and C. A. Buerkle. 2017. Analysis of population genomic data from hybrid zones. Annual Review of Ecology, Evolution, and Systematics 48:207–229.

Grant, V. 1981. Plant speciation. Columbia University Press, New York, NY.

Guerrero, R. F., C. D. Muir, S. Josway, and L. C. Moyle. 2017. Pervasive antagonistic interactions among hybrid incompatibility loci. PLoS Genetics 13:e1006817.

Haenel, Q., M. Roesti, D. Moser, A. D. C. MacColl, and D. Berner. 2019. Predictable genome-wide sorting of standing genetic variation during parallel adaptation to basic versus acidic environments in stickleback fish. Evolution Letters 3:28–42.

Hagen, D. W. 1967. Isolating Mechanisrns in Threespine (Gasterosteus). Journal Fisheries Research Board of Canada 24:1637–1692.

Haldane, J. B. S. 1924. A mathematical theory of natural and artificial selection, Part I. Transactions of the Cambridge Philosophical Society 23:19–41.

———. 1927. A mathematical theory of natural and artificial selection, Part V: selection and mutation. Mathematical Proceedings of the Cambridge Philosophical Society 23:838–844.

Harrison, R. G. 1993. Hybrid zones and the evolutionary process. Oxford University Press, New York.

———. 2012. The language of speciation. Evolution 66:3643–3657.

Hartl, D. 2002. Compensatory Nearly Neutral Mutations: Selection without Adaptation. Journal of Theoretical Biology 182:303–309.

Hatfield, T. 1997. Genetic divergence in adaptive characters between sympatric species of stickleback. The American Naturalist 149:1009–1029.

Hatfield, T., and D. Schluter. 1999. Ecological speciation in sticklebacks: environment-dependent hybrid fitness. Evolution 53:866–873.

Henry, L., and H. Wickham. 2019. purrr: Functional programming tools.

Hermann, K., U. Klahre, J. Venail, A. Brandenburg, and C. Kuhlemeier. 2015. The genetics of reproductive organ morphology in two *Petunia* species with contrasting pollination syndromes. Planta 241:1241–1254.

Hermisson, J., and P. S. Pennings. 2005. Soft sweeps: Molecular population genetics of adaptation from standing genetic variation. Genetics 169:2335–2352.

Hersh, A. H. 1934. On Mendelian dominance and the serial order of phenotypic effects in the bar series of *Drosophila melanogaster*. The American Naturalist 68:186–190.

Hill, W. G., and X. S. Zhang. 2012. On the pleiotropic structure of the genotype-phenotype map and the evolvability of complex organisms. Genetics 190:1131–1137.

Hopkins, R. 2013. Reinforcement in plants. New Phytologist 197:1095–1103.

Hothorn, T., F. Bretz, P. Westfall, and R. M. Heiberger. 2008. multcomp: simultaneous inference for general linear hypotheses, 2008.

Howe, D. K., and D. R. Denver. 2008. Muller's Ratchet and compensatory mutation in *Caenorhabditis briggsae* mitochondrial genome evolution. BMC Evolutionary Biology 8:62.

Hubbs, C. L. 1940. Speciation of Fishes. The American Naturalist 74:198–211.

———. 1955. Hybridization between fish species in nature. Systematic Zool 4:1–20.

Hudson, R. R., M. Slatkin, and W. P. Maddison. 1992. Estimation of levels of gene flow from DNA sequence data. Genetics 132:583–589.

Husemann, M., M. Tobler, C. McCauley, B. Ding, and P. D. Danley. 2017. Body shape differences in a pair of closely related Malawi cichlids and their hybrids: Effects of genetic variation, phenotypic plasticity, and transgressive segregation. Ecology and Evolution 7:4336–4346.

Ingram, T., R. Svanbäck, N. J. Kraft, P. Kratina, L. Southcott, and D. Schluter. 2012. Intraguild predation drives evolutionary niche shift in threespine stickleback. Evolution 66:1819–1832.

Irwin, D. E. 2020. Assortative mating in hybrid zones is remarkably ineffective in promoting speciation. The American Naturalist 195:150–167.

Irwin, D. E., B. Milá, D. P. Toews, A. Brelsford, H. L. Kenyon, A. N. Porter, C. Grossen, K. E. Delmore, M. Alcaide, and J. H. Irwin. 2018. A comparison of genomic islands of differentiation across three young avian species pairs. Molecular Ecology 27:4839–4855.

Jacquemyn, H., R. Brys, O. Honnay, and I. Roldán-Ruiz. 2012. Asymmetric gene introgression in two closely related *Orchis* species: Evidence from morphometric and genetic analyses. BMC Evolutionary Biology 12:178.

Jaeger, B. 2017. r2glmm: Computes R Squared for Mixed (Multilevel) Models.

Jiang, Y., D. I. Bolnick, and M. Kirkpatrick. 2013. Assortative mating in animals. American Naturalist 181:E125–E138.

Johansen-Morris, A. D., and R. G. Latta. 2006. Fitness Consequences of Hybridization Between Ecotypes of Avena Barbata: Hybrid Breakdown, Hybrid Vigor, and Transgressive Segregation. Evolution 60:1585.

Jones, F. C., C. Brown, J. M. Pemberton, and V. A. Braithwaite. 2006. Reproductive isolation in a threespine stickleback hybrid zone. Journal of Evolutionary Biology 19:1531–1544.

Jones, F. C., Y. F. Chan, J. Schmutz, J. Grimwood, S. D. Brady, A. M. Southwick, D. M. Absher, R. M. Myers, T. E. Reimchen, B. E. Deagle, D. Schluter, and D. M. Kingsley. 2012*a*. A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. Current Biology 22:83–90.

Jones, F. C., M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli, J. Johnson, R. Swofford, M. Pirun, M. C. Zody, S. White, E. Birney, S. Searle, J. Schmutz, J. Grimwood, M. C. Dickson, R. M. Myers, C. T. Miller,

B. R. Summers, A. K. Knecht, S. D. Brady, H. Zhang, A. A. Pollen, T. Howes, C. Amemiya, J. Baldwin, T. Bloom, D. B. Jaffe, R. Nicol, J. Wilkinson, E. S. Lander, F. Di Palma, K. Lindblad-Toh, and D. M. Kingsley. 2012*b*. The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484:55–61.

Kassambara, A. 2020. ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.4.0. https://CRAN.R-project.org/package=ggpubr.

Kassen, R. 2014. Experimental evolution and the nature of biodiversity. Roberts.

Keagy, J., L. Lettieri, and J. W. Boughman. 2016. Male competition fitness landscapes predict both forward and reverse speciation. Ecology Letters 19:71–80.

Kenward, M. G., and J. H. Roger. 1997. Small sample inference for fixed effects from restricted maximum likelihood. Biometrics 53:983–997.

Kimura, M. 1965. A stochastic model concerning the maintenance of genetic variability in quantitative characters. Proceedings of the National Academy of Sciences of the United States of America 54:731–736.

———. 1969. The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. Genetics 61:893–903.

Koelling, V. A., and R. Mauricio. 2010. Genetic factors associated with mating system cause a partial reproductive barrier between two parapatric species of *Leavenworthia* (Brassicaceae). American Journal of Botany 97:412–422.

Kumar, S., G. Stecher, M. Suleski, and S. B. Hedges. 2017. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. Molecular biology and evolution 34:1812–1819.

Kuzmin, E., B. VanderSluis, W. Wang, G. Tan, R. Deshpande, Y. Chen, M. Usaj, A. Balint, M. M. Usaj, J. Van Leeuwen, E. N. Koch, C. Pons, A. J. Dagilis, M. Pryszlak, J. Z. Y. Wang, J. Hanchard, M. Riggi, K. Xu, H. Heydari, B. J. S. Luis, E. Shuteriqi, H. Zhu, N. Van Dyk, S. Sharifpoor, M. Costanzo, R. Loewith, A. Caudy, D. Bolnick, G. W. Brown, B. J. Andrews, C. Boone, and C. L. Myers. 2018. Systematic analysis of complex genetic interactions. Science 360:6386.

Kuznetsova, A., P. B. Brockhoff, and H. B. Christensen. 2014. lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). v 2.0-11. http://CRAN.R-project.org/package=lmerTest.

Kvitek, D. J., and G. Sherlock. 2011. Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. PLoS Genetics 7:e1002056.

Lackey, A. C., and J. W. Boughman. 2017. Evolution of reproductive isolation in stickleback fish. Evolution 71:357–372.

Lamb, T., and J. C. Avise. 1987. Morphological variability in genetically defined categories of anuran hybrids. Evolution 41:157–165.

Lande, R. 1979. Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. Evolution 33:402–416.

———. 1981. The minimum number of genes contributing to quantitative variation between and within populations. Genetics 99:541–553.

Langerhans, R. B., and R. Riesch. 2013. Speciation by selection: A framework for understanding ecology's role in speciation. Current Zoology 59:31–52.

Lee, K. M., and G. Coop. 2017. Distinguishing among modes of convergent adaptation using population genomic data. Genetics 207:1591–1619.

Lenth, R., H. Singmann, J. Love, P. Buerkner, and M. Herve. 2020. Package ' emmeans ' R topics documented :. R package version 1.15-15 34:216–221.

Li, S. 2011. Concise formulas for the area and volume of a hyperspherical cap. Asian Journal of Mathematics Statistics 4:66–70.

Lynch, M., and B. Walsh. 1998. Genetics and analysis of quantitative traits. 1st ed. Sinauer, Sunderland, MA.

MacNair, M. R., V. E. MacNair, and B. E. Martin. 1989. Adaptive speciation in *Mimulus*: an ecological comparison of *M. cupriphilus* with its presumed progenitor, *M. guttatus*. New Phytologist 112:269–279.

MacPherson, A., and S. L. Nuismer. 2017. The probability of parallel genetic evolution from standing genetic variation. Journal of Evolutionary Biology 30:326–337.

Maheshwari, S., and D. A. Barbash. 2011. The Genetics of Hybrid Incompatibilities. Annual Review of Genetics 45:331–355.

Makiyama, K. 2016. magicfor: Magic Functions to Obtain Results from for Loops.

Makowski, D., M. S. Ben-Shachar, I. Patil, and D. Lüdecke. 2019. Methods and Algorithms for Correlation Analysis in R. Journal of Open Source Software 5:2306.

Mallet, J. 1986. Hybrid zones of *Heliconius* butterflies in Panama and the stability and movement of warning colour clines. Heredity 56:191–202.

———. 2005. Hybridization as an invasion of the genome. Trends in Ecology and Evolution 20:229–237.

Mani, G. S., and B. C. Clarke. 1990. Mutational order: A major stochastic process in evolution. Proceedings of the Royal Society B: Biological Sciences 240:29–37.

Martin, C. H., and P. C. Wainwright. 2013. Multiple fitness peaks on the adaptive landscape drive adaptive radiation in the wild. Science 339:208–11.

Martin, G., and T. Lenormand. 2015. The fitness effect of mutations across environments: Fisher's geometrical model with multiple optima. Evolution 69:1433–1447.

Matsubayashi, K. W., I. Ohshima, and P. Nosil. 2010. Ecological speciation in phytophagous insects. Entomologia Experimentalis et Applicata 134:1–27.

Matuszewski, S., J. Hermisson, and M. Kopp. 2014. Fisher's geometric model with a moving optimum. Evolution 68:2571–2588.

———. 2015. Catch me if you can: Adaptation from standing genetic variation to a moving phenotypic optimum. Genetics 200:1255–1274.

Matute, D. R., I. A. Butler, D. A. Turissini, and J. A. Coyne. 2010. A test of the snowball theory for the rate of evolution of hybrid incompatibilities. Science 329:1518–1521.

Matute, D. R., and B. S. Cooper. 2021. Comparative studies on speciation: 30 years since Coyne and Orr. Evolution In press.

Mayo, O., and R. Bürger. 1997. The evolution of dominance: A theory whose time has passed? Biological Reviews of the Cambridge Philosophical Society 72:97–110.

Mayr, E. 1963. Animal species and evolution. The Belknap Press, Cambridge, MA.

McGee, M. D., J. W. Reustle, C. E. Oufiero, and P. C. Wainwright. 2015. Intermediate Kinematics Produce Inferior Feeding Performance in a Classic Case of Natural Hybridization. The American Naturalist 186:807–814.

McGee, M. D., D. Schluter, and P. C. Wainwright. 2013. Functional basis of ecological divergence in sympatric stickleback. BMC Evolutionary Biology .

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.

McKinnon, J. S., and H. D. Rundle. 2002. Speciation in nature: the threespine stickleback model systems. Trends in Ecology Evolution 17:480–488.

McPhail, J. D. 1984. Ecology and evolution of sympatric sticklebacks (Gasterosteus): morphological and genetic evidence for a species pair in Enos Lake, British Columbia. Canadian Journal of Zoology 62:1402–1408.

———. 1992. Ecology and evolution of sympatric sticklebacks (Gasterosteus): evidence for a species-pair in Paxton Lake, Texada Island, British Columbia. Canadian Journal of Zoology 70:361–369.

Melo, M. C., M. E. James, F. Roda, D. Bernal-Franco, M. J. Wilkinson, H. Liu, G. M. Walter, and D. Ortiz-Barrientos. 2019. Evidence for mutation-order speciation in an Australian wildflower. bioRxiv page 692673.

Merilä, J., and B. C. Sheldon. 1999. Genetic architecture of fitness and nonfitness traits: Empirical patterns and development of ideas. Heredity 83:103–109.

Merker, M., M. Barbier, H. Cox, J. P. Rasigade, S. Feuerriegel, T. A. Kohl, R. Diel, S. Borrell, S. Gagneux, V. Nikolayevskyy, S. Andres, U. Nübel, P. Supply, T. Wirth, and S. Niemann. 2018. Compensatory evolution drives multidrug-resistant tuberculosis in central Asia. eLife 7:e38200.

Miller, C. T., A. M. Glazer, B. R. Summers, B. K. Blackman, A. R. Norman, M. D. Shapiro, B. L. Cole, C. L. Peichel, D. Schluter, and D. M. Kingsley. 2014. Modular skeletal evolution in sticklebacks is controlled by additive and clustered quantitative trait loci. Genetics 197:405–420.

Miller, S. E., M. Roesti, and D. Schluter. 2019. A Single Interacting Species Leads to Widespread Parallel Evolution of the Stickleback Genome. Current Biology 29:530–537.e6.

Mione, T., and G. J. Anderson. 2017. Genetics of floral traits of Jaltomata procumbens (Solanaceae). Brittonia 69:1–10.

Moher, D., A. Liberati, J. Tetzlaff, D. G. Altman, and PRISMA. 2009. Preferred reporting items for systematic reviews and meta-analyses. Annals of Internal Medicine 151:264–269.

Morris, M. R., E. Bowles, B. E. Allen, H. A. Jamniczky, and S. M. Rogers. 2018. Contemporary ancestor? Adaptive divergence from standing genetic variation in Pacific marine threespine stickleback. BMC Evolutionary Biology 18:113.

Moyle, L. C., and T. Nakazato. 2010. Hybrid incompatibility "snowballs" between Solanum species. Science 329:1521–1523.

Muller, H. J. 1942. Isolating mechanisms, evolution and temperature. Biol. Symp 6:71–125.

Münkemüller, T., S. Lavergne, B. Bzeznik, S. Dray, T. Jombart, K. Schiffers, and W. Thuiller. 2012. How to measure and test phylogenetic signal. Methods in Ecology and Evolution 3:743–756.

Nakagawa, S., and H. Schielzeth. 2013. A general and simple method for obtaining R2 from generalized linear mixed-effects models. Methods in Ecology and Evolution 4:133–142.

Nei, M., and W.-H. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proceedings of the National Academy of Sciences of the United States of America 76:5269–5273.

Nelson, T. C., J. G. Crandall, C. M. Ituarte, J. M. Catchen, and W. A. Cresko. 2019. Selection, linkage, and population structure interact to shape genetic variation among threespine stickleback genomes. Genetics 212:1367–1382.

Nelson, T. C., and W. A. Cresko. 2018. Ancient genomic variation underlies repeated ecological adaptation in young stickleback populations. Evolution Letters 2:9–21.

Nosil, P. 2012. Ecological speciation. Oxford University Press, Oxford, UK.

Nosil, P., and S. M. Flaxman. 2011. Conditions for mutation-order speciation. Proceedings of the Royal Society B: Biological Sciences 278:399–407.

Ohta, T. 1971. Associative overdominance caused by linked detrimental mutations. Genetical Research 18:277–286.

Oke, K. B., G. Rolshausen, C. LeBlond, and A. P. Hendry. 2017. How parallel is parallel evolution? A comparative analysis in fishes. The American Naturalist 190:1–16.

Ono, J., A. C. Gerstein, and S. P. Otto. 2017. Widespread genetic incompatibilities between first-step mutations during parallel adaptation of *Saccharomyces cerevisiae* to a common environment. PLoS Biology 15:e1002591.

Orr, H. A. 1995. The population genetics of speciation: The evolution of hybrid incompatibilities. Genetics 139:1805–1813.

———. 1998. The population genetics of adaptation: The distribution of factors fixed during adaptive evolution. Evolution 52:935.

———. 2000. Adaptation and the cost of complexity. Evolution 54:13–20.

———. 2001. The genetics of species differences. Trends in Ecology and Evolution 16:343–350.

———. 2005. The genetic theory of adaptation: a brief history. Nature Reviews Genetics 6:119–127.

Orr, H. A., and A. J. Betancourt. 2001. Haldane's sieve and adaptation from the standing genetic variation. Genetics 157:875–884.

Osmond, M. M., and S. P. Otto. 2015. Fitness-valley crossing with generalized parent-offspring transmission. Theoretical Population Biology 105:1–16.

Ostevik, K. L., B. T. Moyers, G. L. Owens, and L. H. Rieseberg. 2012. Parallel ecological speciation in plants? International Journal of Ecology 2012:939862.

Otto, S. P., and D. Bourguet. 1999. Balanced polymorphisms and the evolution of dominance. The American Naturalist 153:561–574.

Otto, S. P., and C. D. Jones. 2000. Detecting the undetected: Estimating the total number of loci underlying a quantitative trait. Genetics 156:2093–2107.

Owens, G. L., G. J. Baute, S. Hubner, and L. H. Rieseberg. 2019. Genomic sequence and copy number evolution during hybrid crop development in sunflowers. Evolutionary Applications 12:54–65.

Paradis, E., and K. Schliep. 2018. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics .

Pettengill, J. B., and D. A. Moeller. 2012. Phylogeography of speciation: Allopatric divergence and secondary contact between outcrossing and selfing *Clarkia*. Molecular Ecology 21:4578–4592.

Phillips, P. C. 1996. Waiting for a compensatory mutation: phase zero of the shifting-balance process. Genetical Research 67:271–283.

Poon, A., and L. Chao. 2005. The rate of compensatory mutation in the DNA bacteriophage ϕX174. Genetics 170:989–999.

Presgraves, D. C. 2003. A fine-scale genetic analysis of hybrid incompatibilities in Drosophila. Genetics 163:955–972.

———. 2010. Speciation genetics: Search for the missing snowball. Current Biology 20.

Pritchard, V. L., V. L. Knutson, M. Lee, J. Zieba, and S. Edmands. 2013. Fitness and morphological outcomes of many generations of hybridization in the copepod *Tigriopus californicus*. Journal of Evolutionary Biology 26:416–433.

R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Raeymaekers, J. A., L. Delaire, and A. P. Hendry. 2009. Genetically based differences in nest characteristics between lake, inlet, and hybrid threespine stickleback from the Misty system, British Columbia, Canada. Evolutionary Ecology Research 11:905–919.

Ralph, P. L., and G. Coop. 2015. The role of standing variation in geographic convergent adaptation. The American Naturalist 186:S5–S23.

Reid, N. M., D. A. Proestou, B. W. Clark, W. C. Warren, J. K. Colbourne, J. R. Shaw, S. I. Karchner, M. E. Hahn, D. Nacci, M. F. Oleksiak, D. L. Crawford, and A. Whitehead. 2016. The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. Science 354:1305–1308.

Rennison, D. J., K. Heilbron, R. D. Barrett, and D. Schluter. 2015. Discriminating selection on lateral plate phenotype and its underlying gene, *Ectodysplasin*, in threespine stickleback. American Naturalist 185:150–156.

Rennison, D. J., S. M. Rudman, and D. Schluter. 2019. Genetics of adaptation: Experimental test of a biotic mechanism driving divergence in traits and genes. Evolution Letters 3:513–520.

Revell, L. J. 2012. phytools: An R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution 3:217–223.

Rieseberg, L. H., M. a. Archer, and R. K. Wayne. 1999. Transgressive segregation, adaptation and speciation. Heredity 83:363–372.

Rieseberg, L. H., and N. C. Ellstrand. 1993. What can molecular and morphological markers tell us about plant hybridization? Critical Reviews in Plant Sciences 12:213–241.

Rieseberg, L. H., S. C. Kim, R. A. Randell, K. D. Whitney, B. L. Gross, C. Lexer, and K. Clay. 2007. Hybridization and the colonization of novel habitats by annual sunflowers. Genetica 129:149–165.

Robinson, D., A. Hayes, and S. Couch. 2020. broom: Convert Statistical Objects into Tidy Tibbles.

Robinson, J. A., C. Brown, B. Y. Kim, K. E. Lohmueller, and R. K. Wayne. 2018. Purging of Strongly Deleterious Mutations Explains Long-Term Persistence and Absence of Inbreeding Depression in Island Foxes. Current Biology 28:3487–3494.e4.

Rockman, M. V. 2012. The QTN program and the alleles that matter for evolution: All that's gold does not glitter. Evolution 66:1–17.

Roesti, M., S. Gavrilets, A. P. Hendry, W. Salzburger, and D. Berner. 2014. The genomic signature of parallel adaptation from shared genetic variation. Molecular Ecology 23:3944–3956.

Roesti, M., D. Moser, and D. Berner. 2013. Recombination in the threespine stickleback genome - Patterns and consequences. Molecular Ecology 22:3014–3027.

Rogers, S. M., P. Tamkee, B. Summers, S. Balabahadra, M. Marks, D. M. Kingsley, and D. Schluter. 2012. Genetic signature of adaptive peak shift in threespine stickleback. Evolution 66:2439–2450.

Rohatgi, A. 2019. WebPlotDigitizer (https://automeris.io/WebPlotDigitizer).

Rosas, U., N. H. Barton, L. Copsey, P. B. de Reuille, and E. Coen. 2010. Cryptic variation between species and the basis of hybrid performance. PLoS Biology 8:e1000429.

Rundle, H. D. 2002. A test of ecologically dependent postmating isolation between sympatric sticklebacks. Evolution 56:322–9.

Rundle, H. D., S. M. Vamosi, and D. Schluter. 2003. Experimental test of predation's effect on divergent selection during character displacement in sticklebacks. Proceedings of the National Academy of Sciences of the United States of America 100:14943–14948.

Rundle, H. D., and M. C. Whitlock. 2001. A genetic interpretation of ecologically dependent isolation. Evolution 55:198–201.

Schemske, D. W., and H. D. Bradshaw. 1999. Pollinator preference and the evolution of floral traits in monkeyflowers (*Mimulus*). Proceedings of the National Academy of Sciences of the United States of America 96:11910–11915.

Schluter, D. 1993. Adaptive radiation in sticklebacks: size, shape, and habitat use efficiency. Ecology 74:699–709.

———. 1996. Ecological speciation in postglacial fishes. Philosophical Transactions of the Royal Society B: Biological Sciences 351:807–814.

———. 2000. The Ecology of Adaptive Radiation. Oxford University Press, New York.

———. 2001. Ecology and the origin of species. Trends in Ecology and Evolution 16:372–380.

———. 2009. Evidence for ecological speciation and its alternative. Science 323:737–741.

Schluter, D., and G. L. Conte. 2009. Genetics and ecological speciation. Proceedings of the National Academy of Sciences of the USA 106.Sup1:9955–62.

Schluter, D., K. B. Marchinko, M. E. Arnegard, H. Zhang, S. D. Brady, F. C. Jones, M. A. Bell, and D. M. Kingsley. 2021. Fitness maps to a large-effect locus in introduced stickleback populations. Proceedings of the National Academy of Sciences 118:e1914889118.

Schluter, D., and J. D. McPhail. 1992. Ecological Character Displacement and Speciation in Sticklebacks. The American Naturalist 140:85–108.

Schluter, D., T. D. Price, and L. Rowe. 1991. Conflicting selection pressures and life history trade-offs. Proceedings of the Royal Society B: Biological Sciences 246:11–17.

Schluter, D., M. Roesti, and T. Veen. 2017. Mark-recapture estimates of stickleback population sizesin Paxton and Priest Lakes in 2016. Tech. rep., University of British Columbia; BC Ministry of Environment; Fisheries and Oceans Canada.

Schumer, M., G. G. Rosenthal, and P. Andolfatto. 2014. How common is homoploid hybrid speciation? Evolution 68:1553–1560.

Scopece, G., A. Croce, C. Lexer, and S. Cozzolino. 2013. Components Of Reproductive Isolation Between Orchis Mascula And Orchis Pauciflora. Evolution 67:2083–2093.

Selz, O. M., R. Thommen, M. E. Maan, and O. Seehausen. 2014. Behavioural isolation may facilitate homoploid hybrid speciation in cichlid fish. Journal of Evolutionary Biology 27:275–289.

Shafer, A. B. A., and J. B. W. Wolf. 2013. Widespread evidence for incipient ecological speciation: A meta-analysis of isolation-by-ecology. Ecology Letters 16:940–950.

Shore, J. S., and S. C. Barrett. 1990. Quantitative genetics of floral characters in homostylous *Turnera ulmifolia* var. *angustifolia* Willd. (Turneraceae). Heredity 64:105–112.

Simmons, M. J., and J. F. Crow. 1977. Mutations affecting fitness in Drosophila populations. Annual review of genetics 11:49–78.

Simon, A., N. Bierne, and J. J. Welch. 2018. Coadapted genomes and selection on hybrids: Fisher's geometric model explains a variety of empirical patterns. Evolution Letters 2:472–498.

Sinervo, B., and E. Svensson. 2002. Correlational selection and the evolution of genomic architecture. Heredity 89:329–338.

Slatkin, M., and R. Lande. 1994. Segregation variance after hybridization of isolated populations. Genetical Research 64:51–56.

Stearns, F. W. 2010. One hundred years of pleiotropy: A retrospective. Genetics 186:767–773.

Stelkens, R., and O. Seehausen. 2009. Genetic distance between species predicts novel trait expression in their hybrids. Evolution 63:884–897.

Stelkens, R. B., C. Schmid, O. Selz, and O. Seehausen. 2009. Phenotypic novelty in experimental hybrids is predicted by the genetic distance between species of cichlid fish. BMC Evolutionary Biology 9:283.

Stuart, Y. E., T. Veen, J. N. Weber, D. Hanson, M. Ravinet, B. K. Lohman, C. J. Thompson, T. Tasneem, A. Doggett, R. Izen, N. Ahmed, R. D. H. Barrett, A. P. Hendry, C. L. Peichel, and D. I. Bolnick. 2017. Contrasting effects of environment and genetics generate a continuum of parallel evolution. Nature Ecology Evolution 1:0158.

Tave, D. 1986. Genetics for fish hatchery managers. Tave, D. Genetics for Fish Hatchery Managers. Xv+299p. .

Taylor, E. B. 1999. Species pairs of north temperate freshwater fishes: Evolution, taxonomy, and conservation. Reviews in Fish Biology and Fisheries 9:299–324.

Taylor, E. B., and J. D. McPhail. 2000. Historical contingency and ecological determinism interact to prime speciation in sticklebacks, Gasterosteus. Proceedings of the Royal Society B: Biological Sciences 267:2375–2384.

Tenaillon, O. 2014. The utility of Fisher's geometric model in evolutionary genetics. Annual Review of Ecology, Evolution, and Systematics 45:179–201.

Thompson, K. A. 2020. Experimental hybridization studies suggest that pleiotropic alleles commonly underlie adaptive divergence between natural populations. American Naturalist 196:E16–E22.

Thompson, K. A., M. M. Osmond, and D. Schluter. 2019. Parallel genetic evolution and speciation from standing variation. Evolution Letters 3:129–141.

Thompson, K. A., M. Urquhart-Cronish, K. D. Whitney, L. H. Rieseberg, and D. Schluter. 2021. Patterns, predictors, and consequences of dominance in hybrids. The American Naturalist In press.

Turelli, M. 1984. Heritable genetic variation via mutation-selection balance: Lerch's zeta meets the abdominal bristle. Theoretical Population Biology 25:138–193.

———. 1985. Effects of pleiotropy on predictions concerning mutation-selection balance for polygenic traits. Genetics 111:165–195.

Turissini, D. A., A. A. Comeault, G. Liu, Y. C. G. Lee, and D. R. Matute. 2017. The ability of Drosophila hybrids to locate food declines with parental divergence. Evolution 71:960–973.

Vallejo-Marín, M., M. Quenu, S. Ritchie, and S. Meeus. 2017. Partial interfertility between independently originated populations of the neo-allopolyploid *Mimulus peregrinus*. Plant Systematics and Evolution 303:1081–1092.

Vamosi, S. M., T. Hatfield, and D. Schluter. 2000. A test of ecological selection against young-of-the-year hybrids of sympatric sticklebacks. Journal of Fish Biology 57:109–121.

Vasseur, F., L. Fouqueau, D. De Vienne, T. Nidelet, C. Violle, and D. Weigel. 2019. Nonlinear phenotypic variation uncovers the emergence of heterosis in *Arabidopsis thaliana*. PLoS Biology 17:e3000214.

Vines, T., A. Dalziel, A. Albert, T. Veen, P. Schulte, and D. Schluter. 2016. Cline coupling and uncoupling in a stickleback hybrid zone. Evolution 70:1023–1038.

Vinšálková, T., and L. Gvoždík. 2007. Mismatch between temperature preferences and morphology in F1 hybrid newts (*Triturus carnifex* × *T. dobrogicus*). Journal of Thermal Biology 32:433–439.

Wagner, G. P., J. P. Kenney-Hunt, M. Pavlicev, J. R. Peck, D. Waxman, and J. M. Cheverud. 2008. Pleiotropic scaling of gene effects and the 'cost of complexity'. Nature 452:470–472.

Wagner, G. P., and J. Zhang. 2011. The pleiotropic structure of the genotype-phenotype map: The evolvability of complex organisms. Nature Reviews Genetics 12:204–213.

Wang, M. 2018. The Genetic Basis of Adaptation and Speciation in Benthic and Limnetic Threespine Stickleback (Gasterosterus aculeatus). Ph.D. thesis.

Wang, R. J., M. A. White, and B. A. Payseur. 2015. The pace of hybrid incompatibility evolution in house mice. Genetics 201:229–242.

Wang, S., M. Ore, E. K. Mikkelsen, J. Lee-Yaw, S. Rohwer, and D. E. Irwin. 2020. Signatures of mitonuclear climate adaptation in a warbler species complex. bioRxiv 2020.04.06.

Wang, W., C. Krner, Z. Zhang, R. Wu, Y. Geng, W. Shi, and X. Ou. 2013. No slope exposure effect on alpine treeline position in the Three Parallel Rivers Region, SW China. Alpine Botany 123:87–95.

Wang, Z., B. Y. Liao, and J. Zhang. 2010. Genomic patterns of pleiotropy and the evolution of complexity. Proceedings of the National Academy of Sciences of the United States of America 107:18034–18039.

Wei, X., and J. Zhang. 2018. The optimal mating distance resulting from heterosis and genetic incompatibility. Science Advances 4:eaau5518.

Weir, J. T., and D. Schluter. 2004. Ice sheets promote speciation in boreal birds. Proceedings of the Royal Society B: Biological Sciences 271:1881–1887.

Westram, A. M., M. Rafajlović, P. Chaube, R. Faria, T. Larsson, M. Panova, M. Ravinet, A. Blomberg, B. Mehlig, K. Johannesson, and R. Butlin. 2018. Clines on the seashore: The genomic architecture underlying rapid divergence in the face of gene flow. Evolution Letters 2:297–309.

Whitney, K. D., R. A. Randell, and L. H. Rieseberg. 2006. Adaptive introgression of herbivore resistance traits in the weedy sunflower *Helianthus annuus*. The American Naturalist 167:794.

———. 2010. Adaptive introgression of abiotic tolerance traits in the sunflower *Helianthus annuus*. New Phytologist 187:230–239.

Wickham, H. 2017. Tidyverse: Easily Install and Load the 'Tidyverse'. R package version 1.2.1.

Wilkinson, S. P. 2018. The 'aphid' package for analysis with profile hidden Markov models. R package version 1.1.0. https://cran.r-project.org/package=aphid.

Williams, G. C. 1957. Pleiotropy, Natural Selection, and the Evolution of Senescence. Evolution 11:398–411.

Willis, J. H. 1999. Inbreeding load, average dominance and the mutation rate for mildly deleterious alleles in *Mimulus guttatus*. Genetics 153:1885–1898.

Winter, D. J. 2017. rentrez: An R package for the NCBI eUtils API. The R journal .

Wright, S. 1931. Evolution in mendelian populations. Genetics 16:97–159.

———. 1934. Physiological and Evolutionary Theories of Dominance. The American Naturalist 68:24–53.

———. 1968. Evolution and the Genetics of Populations Vol. 1. Genetic and Biometric Foundations. University of Chicago Press, Chicago, IL.

Xiong, T., and J. Mallet. 2021. The collapse of genetic incompatibilities in a hybridizing population. bioRxiv page 2021.01.08.425971.

Yamaguchi, R., and S. P. Otto. 2020. Insights from Fisher's geometric model on the likelihood of speciation under different histories of environmental change. Evolution 74:1603–1619.

Zhang, L., G. R. Hood, A. M. Roush, S. A. Shzu, M. S. Comerford, J. R. Ott, and S. P. Egan. 2021. Asymmetric, but opposing reductions in immigrant viability and fecundity promote reproductive isolation among host-associated populations of an insect herbivore. Evolution 75:476–489.

Appendices

Appendix A

Appendix for Chapter 2

A.1 Geometric basis for rapid loss of parallelism

Here, I outline an explanation for why genetic parallelism decreases rapidly with the angle of divergence, θ (Fig. 2.2A) and distance between optima (Fig. A.15B). My explanation focuses on the extent of phenotypic space wherein mutations improve the fitness of both adapting populations in their respective environments. At the time of founding both adapting populations have the same mean phenotype, which is the mean ancestral phenotype. Mutations that move this ancestral mean phenotype into the region that leads to higher fitness in both parental environments are thus beneficial in both populations. The region of phenotypic space that has higher fitness than the mean phenotype in one environment is a hypersphere (of dimensionality, m), centred on the optimum with a radius equal to the distance between the mean phenotype and the optimum, d. A similar hypersphere characterizes the phenotypic space that has higher fitness than the mean phenotype in the other parental environment. The region that is mutually beneficial is then the intersection of two hyperspheres, which is the union of two hyperspherical caps.

Fortunately, the volume of a hyperspherical cap is known for any dimension, m (Li, 2011). It depends on the dimensionality (m), the radii of the two hyperspheres (D), and the distance between their centers (δ). In my case the distance between the two centres is $\delta = 2d \times \sin(\theta/2)$. The amount of phenotypic space that is beneficial in a given environment is simply the volume of one of the hyperspheres. Thus, dividing the volume of the mutually-beneficial space (the union of the hyperspherical caps) by the volume of the space beneficial in a given environment (one of the hyperspheres) gives the fraction of beneficial mutations which are mutually beneficial. Using the formula given by Li (2011; their eqn 3) for the volume of a hyperspherical cap created by the intersection of two m-dimensional hyperspheres with radii d whose centres are distance δ apart, the fraction of beneficial mutations that are expected to be beneficial in both, O, is:

$$O = I_x[1 + m/2, 1/2] \tag{A.1}$$

where $I_x[a, b]$ is the regularized incomplete beta function (Equation 6.6.2 in Abramowitz and Stegun 1972) and here $x = \cos(\theta/2)^2$. Eqn. A.1 depends on only m and θ, that is the solution is independent of the distance from the ancestor to the new optima, d. I refer to Eqn. A.1 as the fraction of overlap in the main text, but note that this is only true when $d_1 = d_2$ (the formula is more complex when $d_1 \neq d_2$, but can easily be used, e.g., Fig. A.16B). The incomplete regularized beta function arises from integrating $\sin^m(\theta)$ over θ (Li, 2011).

The solution of Eqn. A.1 exhibits a rapid decrease with θ for all values of $m > 0$, and the decrease is faster for greater values of m (Fig. 2.2B). Thus, if standing genetic variation was uniformly distributed throughout the beneficial hyperspheres, the percent of segregating beneficial mutations that were beneficial

in both parental populations, and thus expected to potentially fix in both, would decrease rapidly with the angle of divergence.

The above analysis considers only the very onset of adaptation, when the two parental populations have the same mean phenotypes, such that the fraction of phenotypic space that is beneficial in one population that is also beneficial in the other population (call this X) is equivalent to the fraction of possible beneficial mutations (if uniformly distributed across the hyperspheres) that are beneficial in both populations (call this Y). As adaptation proceeds the mean phenotypes of the parental populations depart from one another and X therefore no longer equals Y. This is because mutations are vectors that move a phenotype in a particular direction, and thus a mutually beneficial point in phenotypic space is only guaranteed to be a mutually beneficial mutation if both populations have the same mean phenotype.

To account for the inequality between phenotypic space (X) and mutational vectors (Y) during adaptation we must shift the mean phenotypes so that they are at the same point in phenotypic space and move their optima by an identical translation (see Fig. A.20). We then have $X = Y$. One way to imagine this is to keep the mean phenotypes in place at the mean ancestral phenotype (the origin) and consider adaptation as the movement of the optima closer to the mean phenotypes. From this perspective, adaptation's effect is a shrinking of the radii of the hyperspheres (at roughly equivalent rates in the two populations if adaptation proceeds relatively deterministically). Thus, because the fraction of overlap (Eq. A.1) does not depend on the radii of the hyperspheres, the fraction of overlap is expected to remain relatively constant throughout adaptation.

In reality and in my simulations, standing genetic variation is not uniformly distributed, the probability of fixation varies across the region of overlap, and adaptation uses up some of the standing variation so that the distribution of standing variation changes with time. Taking the first two complications into account would require weighted averages across the space contained in the hyperspherical caps, which is beyond the scope of my study. The third complication is yet more involved and would require an analysis of how standing genetic variation is used as adaptation proceeds (i.e., how the distribution of segregating effects and allele frequencies shift as alleles fix). Such a calculation is also beyond the scope of this chapter. Despite these complications, it seems as though the simple analysis above qualitatively captures the essence of why genetic parallelism decreases rapidly with the angle of divergence.

A.2 Supplementary figures

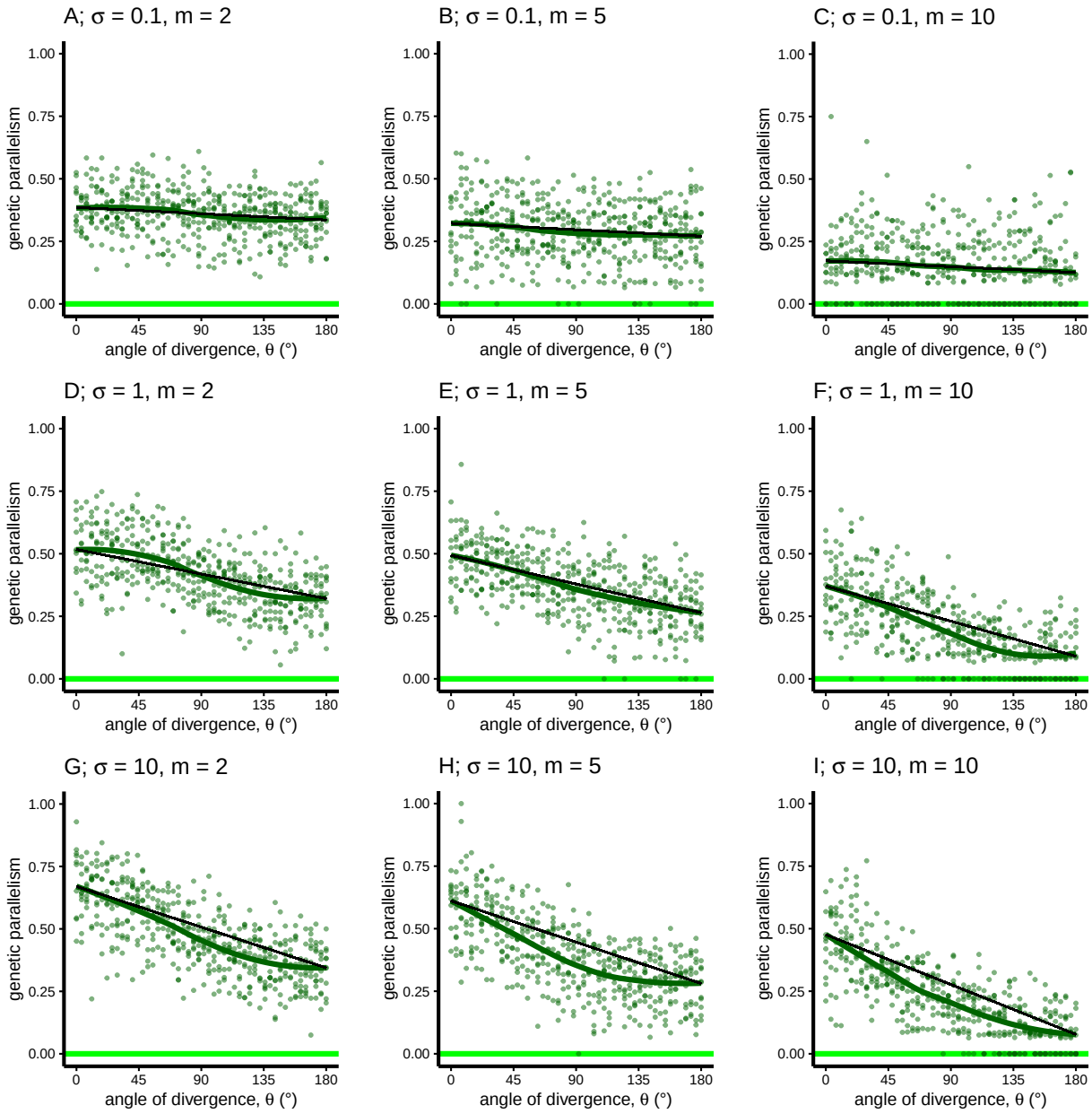


Figure A.1: **Genetic parallelism across the continuum of parallel to divergent natural selection (N = 100)**. This figure presents simulations similar to Fig. 2.2A in the main text but with varying parameter values (selection [σ] and dimensionality [m]). I ran these particular simulations for T = 5000 generations. All other parameters as in main text.

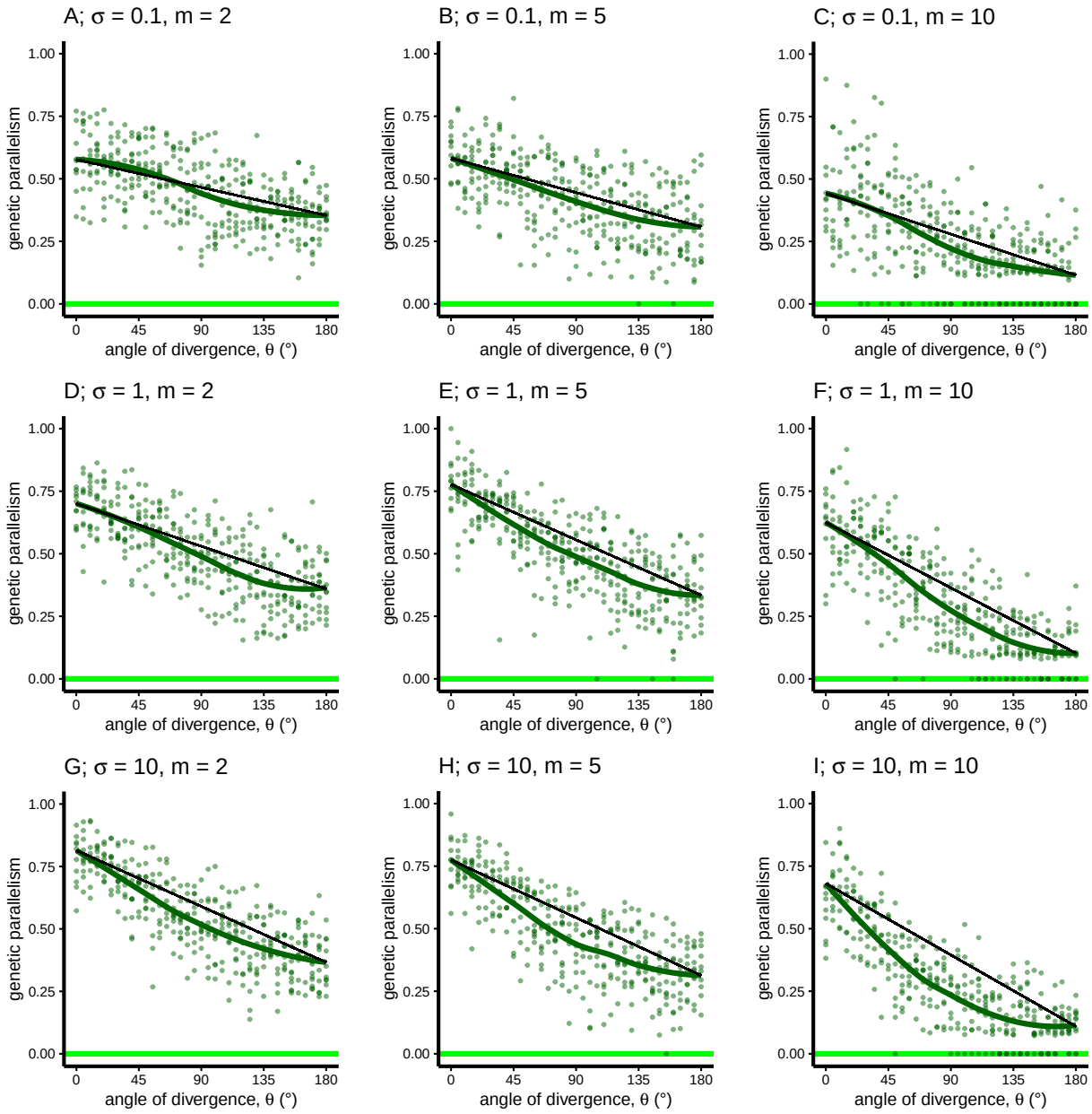


Figure A.2: **Genetic parallelism across the continuum of parallel to divergent natural selection (N = 1000).** This figure presents simulations similar to Fig. 2.2A in the main text but with varying parameter values (selection [σ] and dimensionality [m]). I ran these particular simulations for T = 2000 generations. All other parameters as in main text.

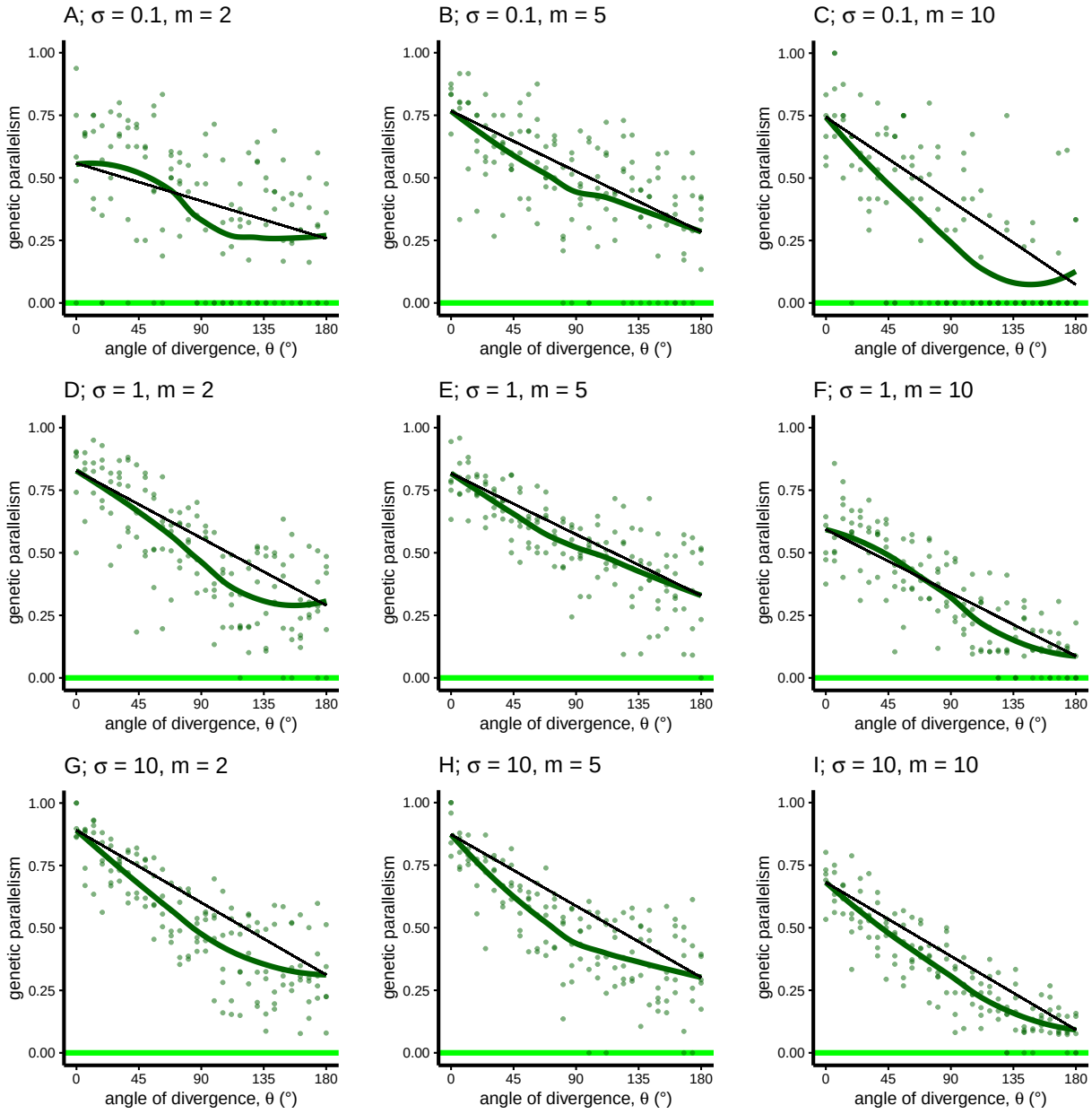


Figure A.3: **Genetic parallelism across the continuum of parallel to divergent natural selection (N = 5000)**. This figure presents simulations similar to Fig. 2.2A in the main text but with varying parameter values (selection [σ] and dimensionality [m]). I ran these particular simulations for T = 1000 generations. All other parameters as in main text. These simulations are computationally intensive and were therefore not run for as many replicates as those plotted in Fig. A.1 or A.2.

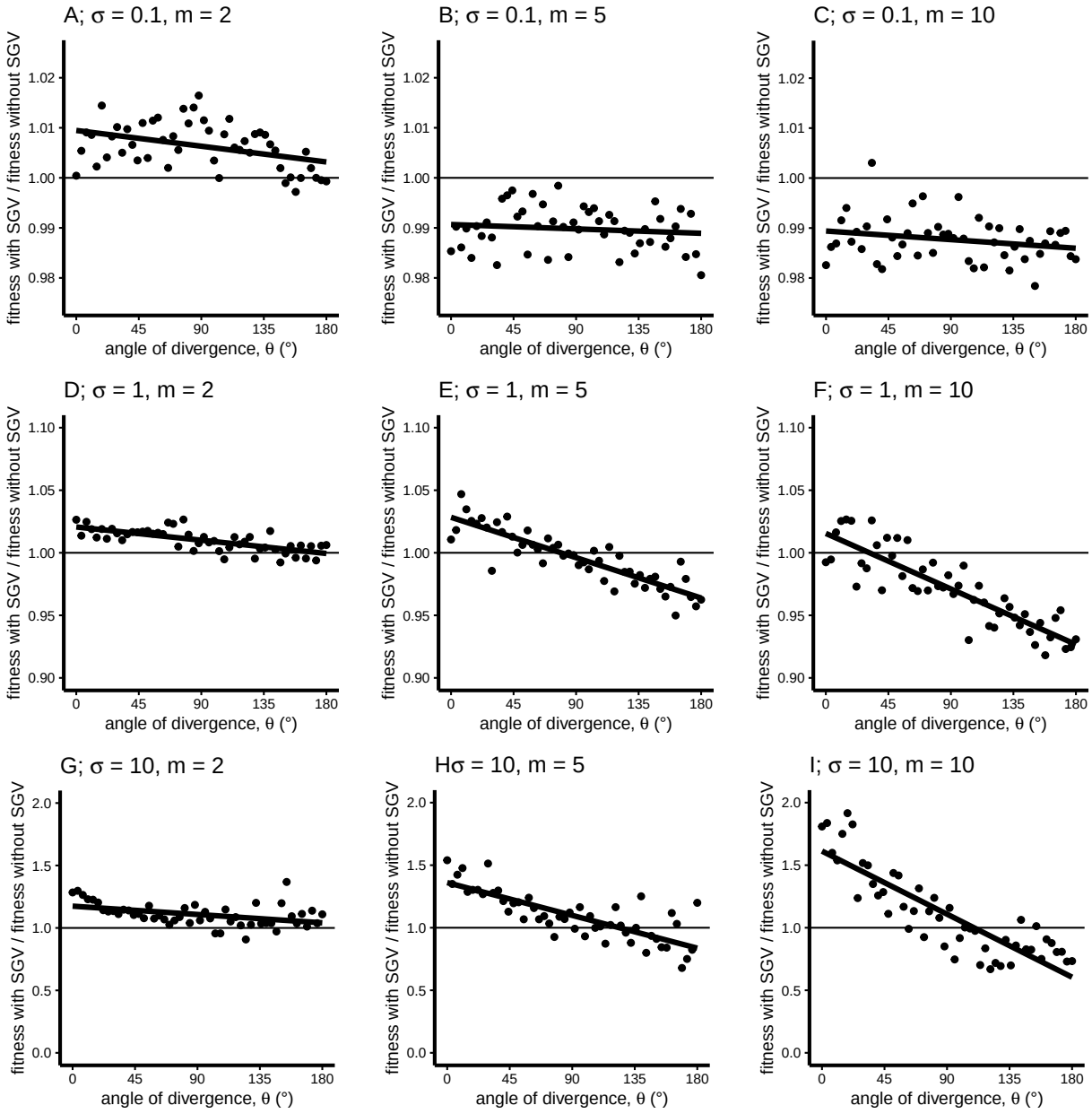


Figure A.4: **Effect of standing genetic variation on hybrid fitness across the continuum of parallel to divergent natural selection ($N = 100$)**. This figure presents simulations similar to Fig. 2.4B in the main text but with varying parameter values (selection [σ] and dimensionality [m]). I ran these particular simulations for $T = 5000$ generations. All other parameters are as in the main text. Note different y-axis scales across rows.

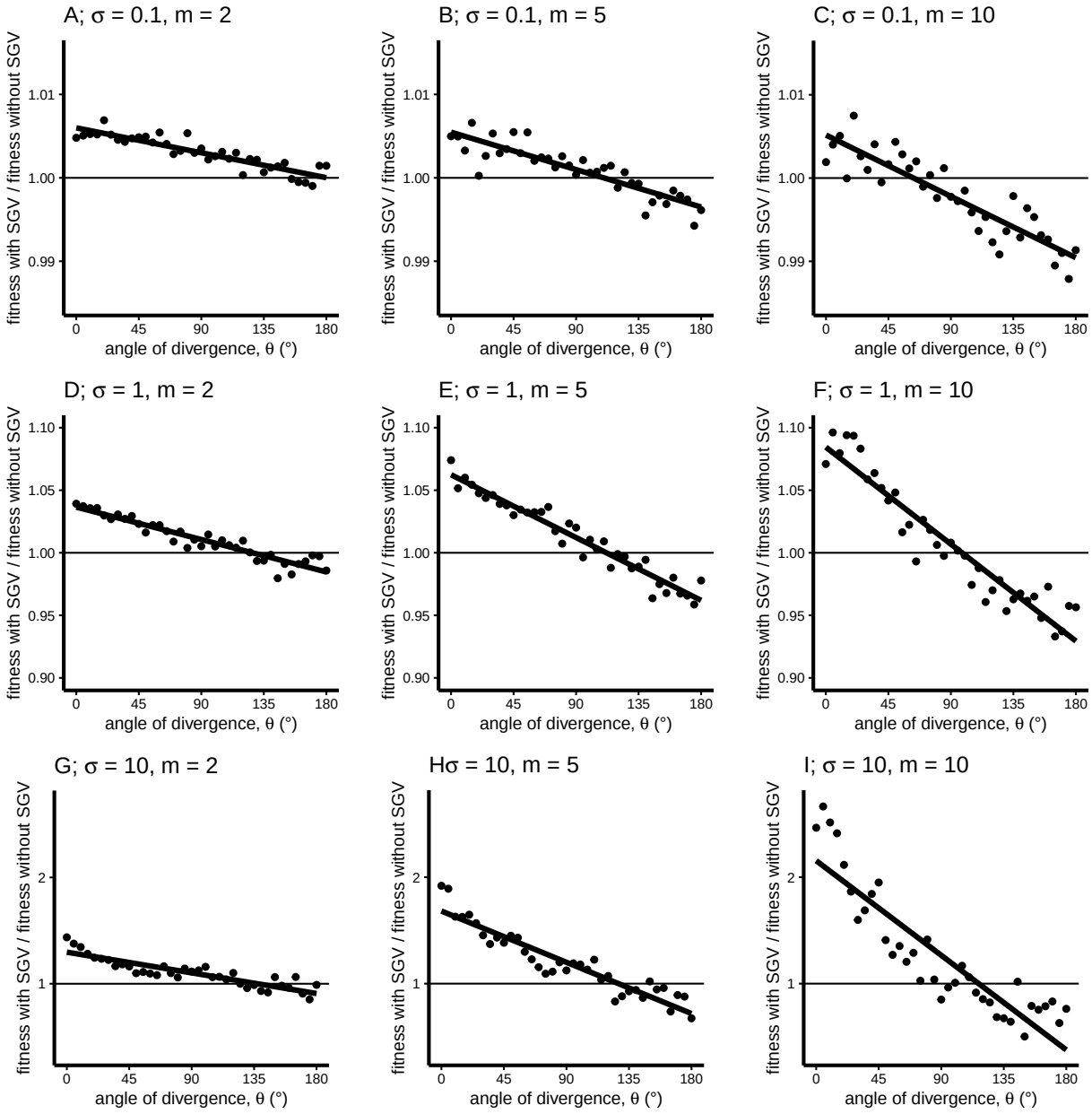


Figure A.5: **Effect of standing genetic variation on hybrid fitness across the continuum of parallel to divergent natural selection (N = 1000)**. This figure presents simulations similar to Fig. 2.4B in the main text but with varying parameter values (selection [σ] and dimensionality [m]). I ran these particular simulations for T = 2000 generations. All other parameters as in main text. Note different y-axis scales across rows.

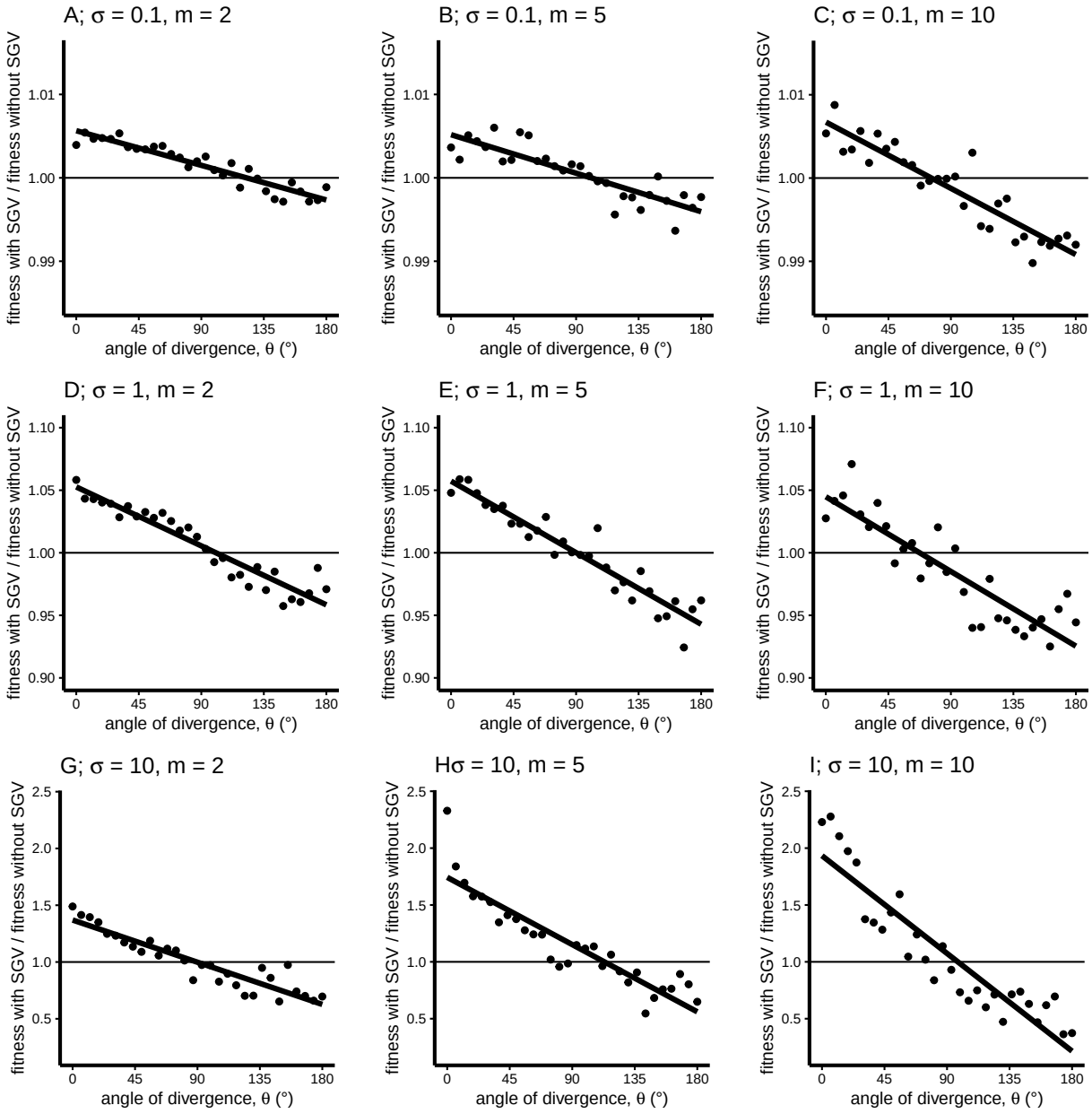


Figure A.6: **Effect of standing genetic variation on hybrid fitness across the continuum of parallel to divergent natural selection (N = 5000)**. This figure presents simulations similar to Fig. 2.4B in the main text but with varying parameter values (selection [σ] and dimensionality [m]). I ran these particular simulations for T = 1000 generations. All other parameters as in main text. Note different y-axis scales across rows.

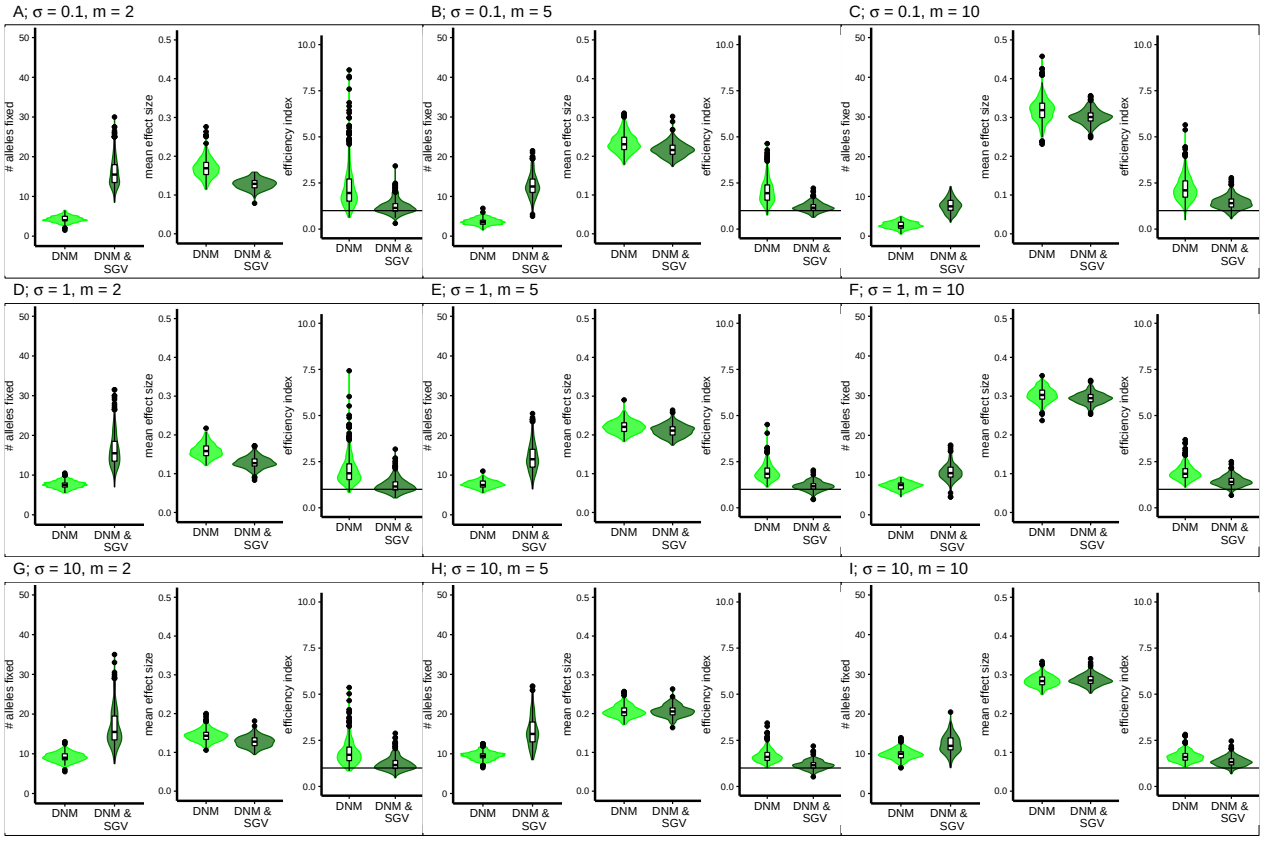


Figure A.7: **Properties of fixed mutations under a variety of parameter combinations (N = 1000)**. This figure presents simulations similar to Fig. 2.6 in the main text but with varying parameter values (selection [σ] and dimensionality [m]). See main text and panel description of Fig. 2.6 for more detail. Patterns were similar for other population sizes.

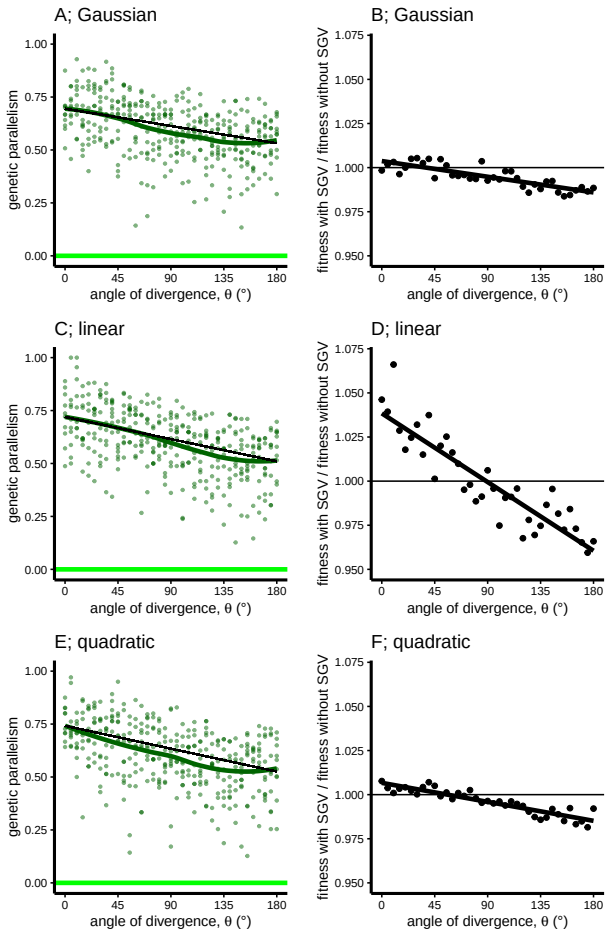


Figure A.8: **Simulations under various fitness functions.** Here I plot simulations across environments for (A & B) Gaussian ($W = \exp(\sigma \|z - o\|2/2)$), (C & D) linear ($W = 1 - \sigma \|z - o\|$), and (E & F) quadratic ($W = 1 - \sigma \|z - o\|^2/2$) fitness functions. I show results for both genetic parallelism and the effect of standing variation on hybrid fitness. I ran these simulations with a nearer optimum and weaker selection ($d = 0.5$, $\sigma = 0.5$, $N = 1000$, $m = 5$) because populations otherwise became extinct with linear/quadratic fitness functions. Under these conditions, the non-linear decrease in parallelism is less substantial for all parameter values. Nevertheless, the patterns are qualitatively similar among the three sets of simulations (note differences in y-axis scales).

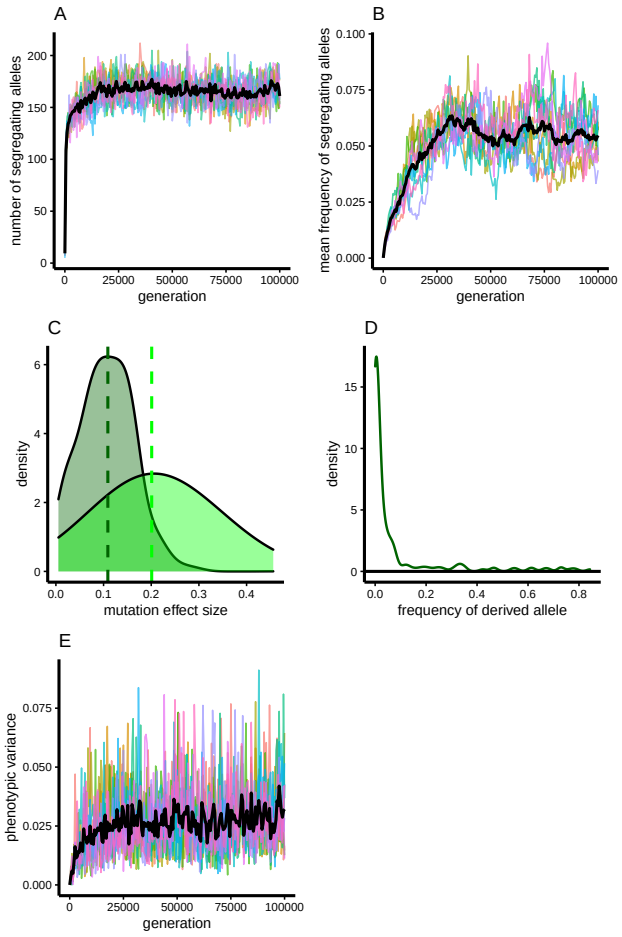


Figure A.9: **Mutation-selection balance and mutation effect sizes in ancestral populations**. In panel (A) I am showing the number of segregating sites in each of 10 ancestral populations and (B) the mean frequency of the derived alleles at each of these sites in the ancestral populations. The black line is plotted through the mean of all populations at each generation, and all ten burn-ins used to generate my main text results are shown. Panel (C) illustrates the distribution of mutation effect sizes—the Euclidean distance of a mutational vector in phenotypic space—at the end of a single representative burn-in simulation (dark green), as compared to the distribution of mutations that arise *de novo* (light green). The vertical lines represent the median mutation effect size for each group. Panel (D) represents the site-frequency spectrum for segregating sites (excluding sites that have fixed). And panel (E) shows the phenotypic variance in the ancestral population over time. ($\sigma = 0.01$; $m = 5$ for all simulations shown; for rest of parameters see Table 2.1).

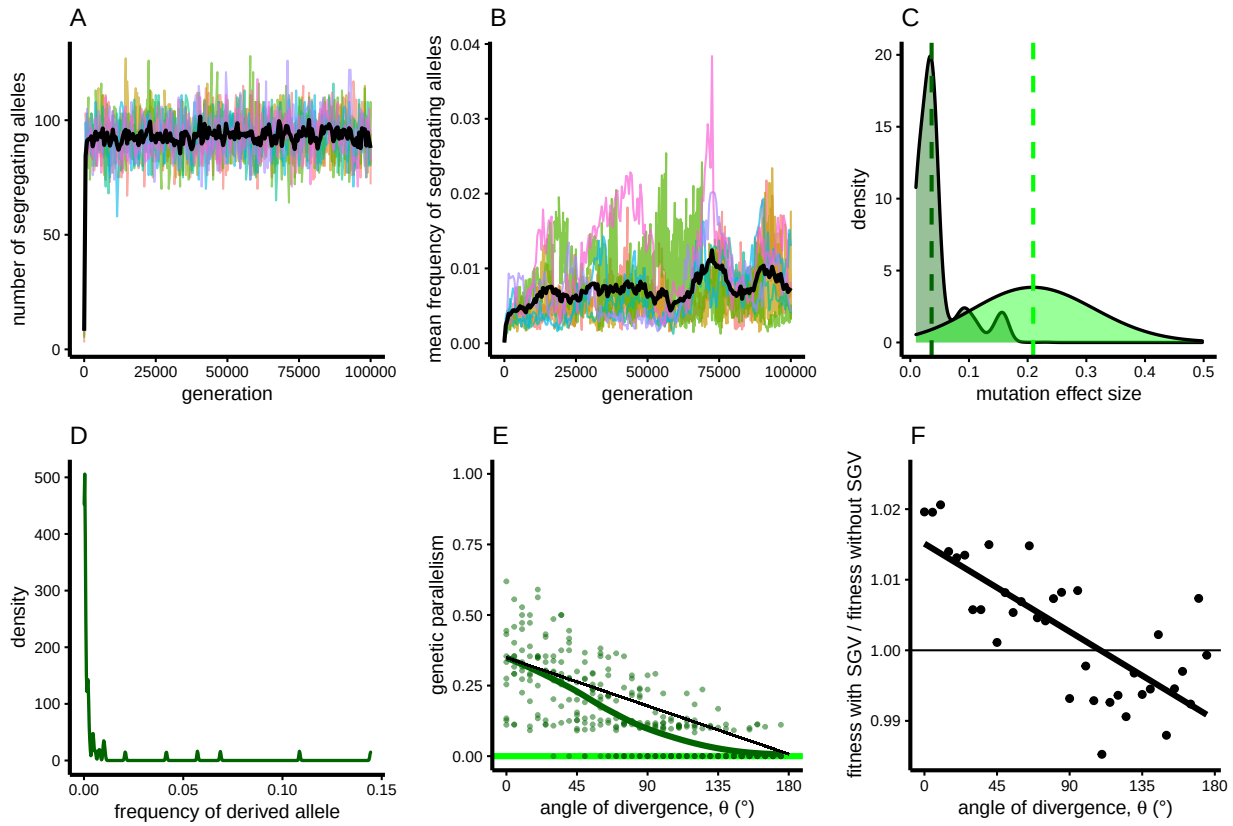


Figure A.10: **Mutation-selection balance and mutation effect sizes in ancestral populations under stronger selection (σanc = 1)**. These parameter values imply $\mu \ll \alpha^2 \sigma$, as in the House-of-Cards regime (Turelli, 1984, 1985) from a Gaussian regime under an alternative set of parameters. (A) The number of segregating sites in each of 10 ancestral populations and (B) the mean frequency of derived alleles at each of these sites in the ancestral populations. The black line is plotted through the mean of all populations at each generation, and all ten burn-ins used to generate the results ([e] and [f]) are shown. Panel (C) illustrates the distribution of mutation effect sizes—the absolute value of a mutation's effect on the phenotype—at the end of a single burn-in simulation (dark green), as compared to the distribution of mutations that arise *de novo* (light green). The vertical lines represent the median mutation effect size for each group. Panel (D) represents the site-frequency spectrum histogram for segregating sites. (Compare these to Fig. A.9). Panels (E) and (F) are as in Fig. 2.2A and 2.4B in the main text. For unspecified parameters see Table 2.1 in the main text.

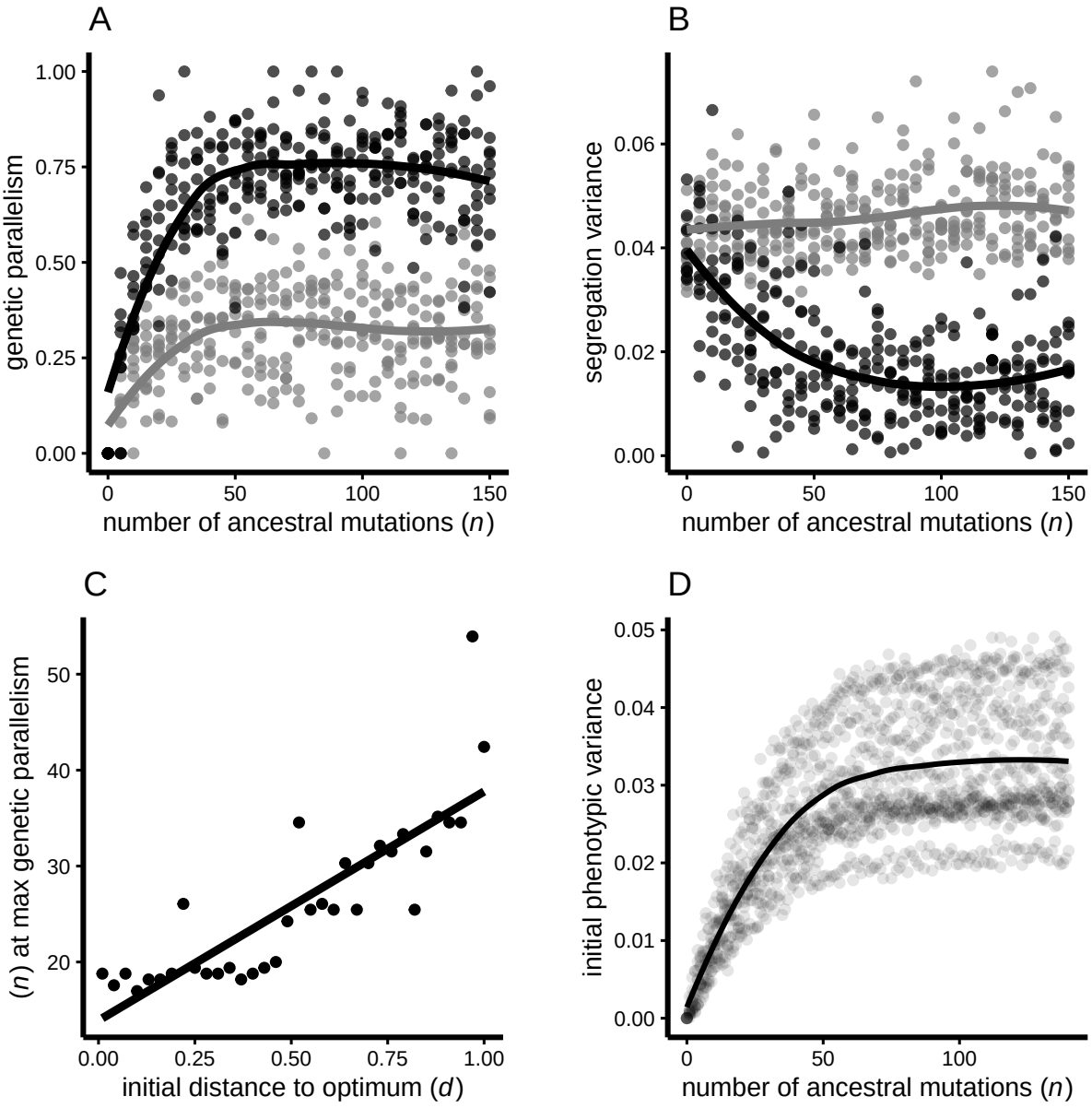


Figure A.11: **The effects of standing genetic variation on genetic parallelism and phenotypic segregation variance in hybrids under parallel and divergent natural selection**. I show (A) genetic parallelism (main text equation 2.1) and (B) net segregation variance for populations founded with varying quantities of ancestral standing variation (n: number of ancestral mutations). Populations were subject to either parallel ($\theta = 0°$; black) and divergent ($\theta = 180°$; grey) selection, with $d = 1$, and there were 10 replicate simulations per parameter combination. Genetic parallelism values of 0 indicate no parallelism and values of 1 indicate complete parallelism (main text Eq. 2.2). The curves are loess fits. Panel (C) shows that the quantity of ancestral standing variation that maximizes genetic parallelism under parallel selection ($\theta = 0°$) increases when populations adapt to more distant optima. A value of $d = 1$ is 10 mutational SDs. The line is a linear regression. Panel (D) shows the relationship between the genetic (phenotypic) variation in a parental population as a function of n.

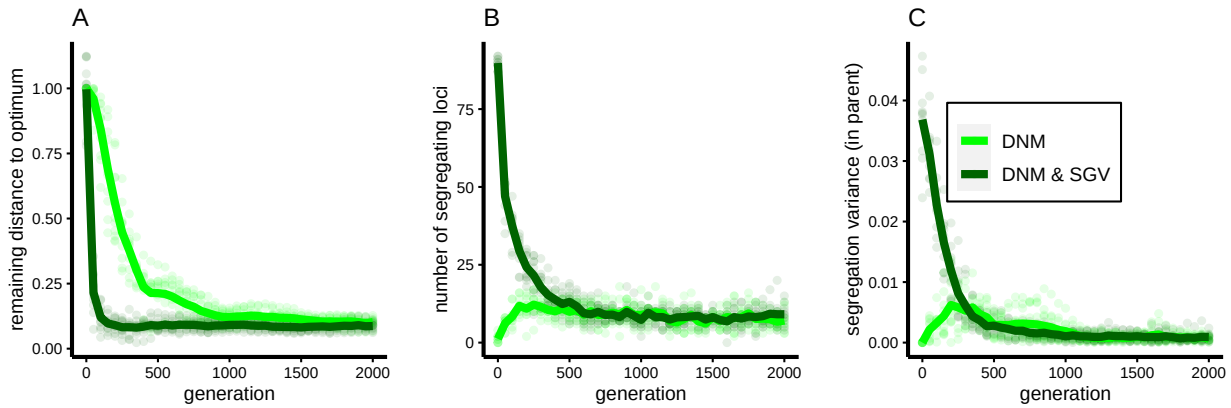


Figure A.12: **Effect of standing variation on the pace of adaptation and attainment of mutation-selection-drift balance**. (A) Populations that adapt with standing variation in addition to new mutation (DNM & SGV; $n = 100$ segregating alleles; dark green) reach the phenotypic optimum more quickly than populations that adapt from new mutation only (DNM; $n = 0$ segregating alleles; light green). (B) Although populations equipped with standing variation adapt more quickly than populations adapting from new mutation only, they both reach mutation-selection-drift balance by generation 2000. (C) The phenotypic (and genotypic) variance in parental populations, calculated as it is in hybrids (see main text), is stable and near zero by the end of each simulation. The initial distance to the optima, d, is $d = 1$ for all simulations. I plot 10 replicate simulations, and lines connect the mean values at each sampled generation. For unspecified parameters see Table 2.1 in the main text.

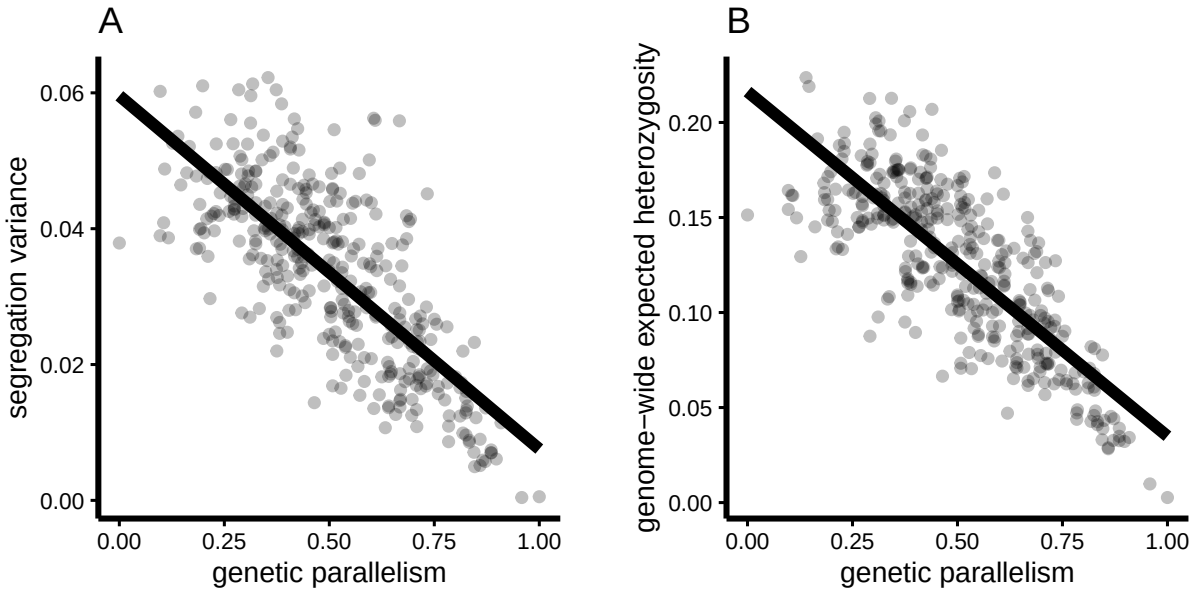


Figure A.13: **Relationship between genetic parallelism and (A) segregation variance and (B) expected heterozygosity**. Our metric of genetic parallelism (main text equation 2) is on the x-axis. This is the data plotted in Fig. 2.2A 2.2C of the main text. I show the correlation between genetic parallelism and (A) segregation variance ($r^2 = 0.56$) and (B) genome-wide expected heterozygosity ($2p[1-p]$), averaged across all loci ($r^2 = 0.63$). Patterns were similar for F_{ST} (Hudson et al., 1992) and net π (Nei and Li, 1979) (not shown).

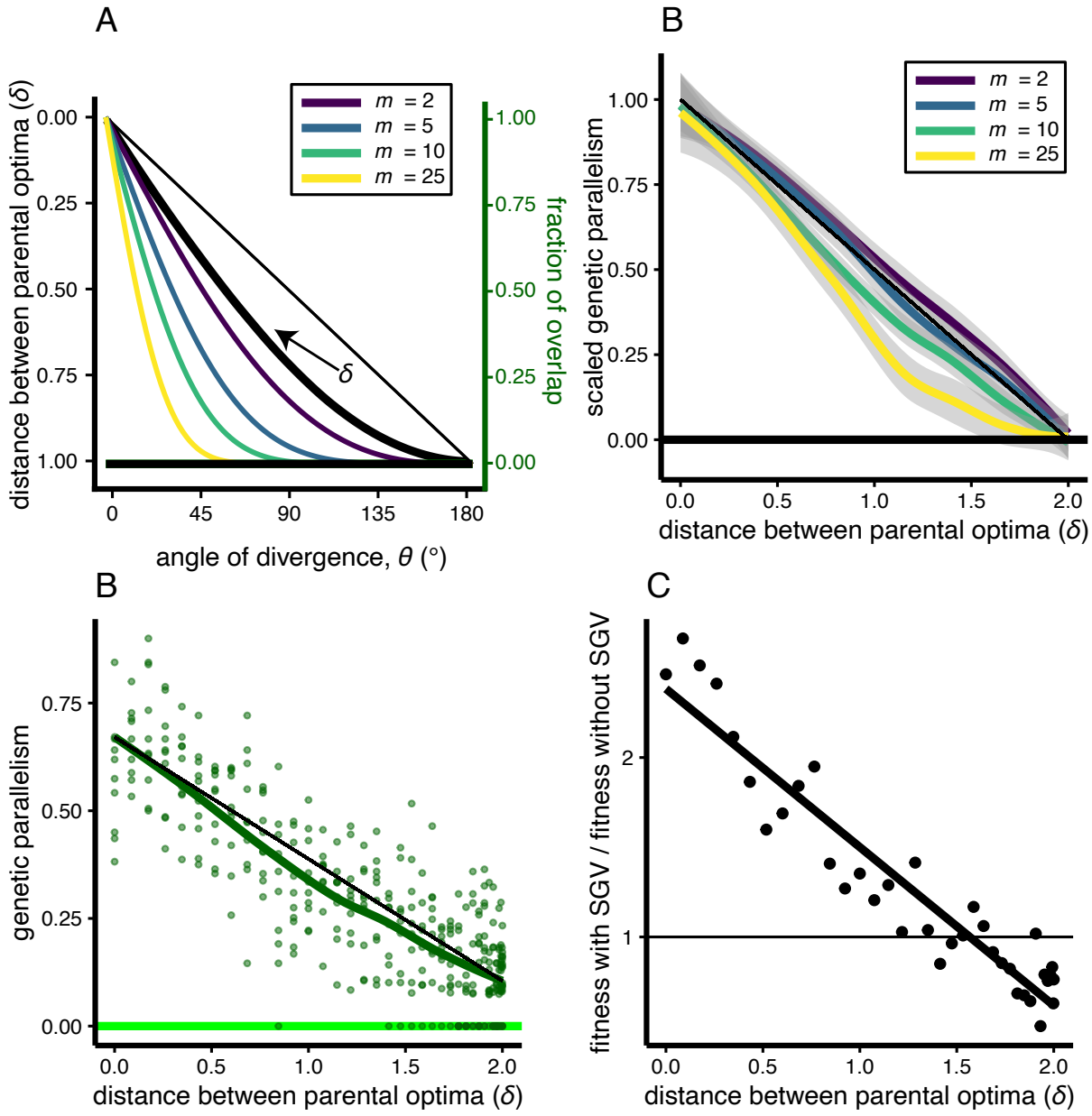


Figure A.14: **Alternative presentation of simulation results across environments: distance between optima** (δ). Panel (A) plots the relationship between the angle of divergence, θ, and the Euclidean distance between parental optima, δ (thick black line; note reversal of left y-axis; scaled between 0 and 1 by dividing by 2d). I also plot the fraction of non-overlap (right [green] y-axis) as in the main text Fig. 2.3A for four different dimensionalities (m; coloured lines). Panel (B) shows observed (scaled) genetic parallelism vs. δ for the same dimensionalities as plotted in (A). For a given value of θ, δ is invariant with dimensionality (i.e., the distance between optima does not change as dimensionality increases). Accordingly, the non-linearity emerges even when considering δ, but only appreciably when considering higher dimensions ($m > 5$). In both panels, the thin and straight black line connects the fit at 0° with 180° for visual reference. In panels (C) and (D) I show the raw data for genetic parallelism and relative hybrid fitness in simulations conducted for simulations conducted 10 dimensions ($m = 10$, $\sigma = 1$, $N = 1000$).

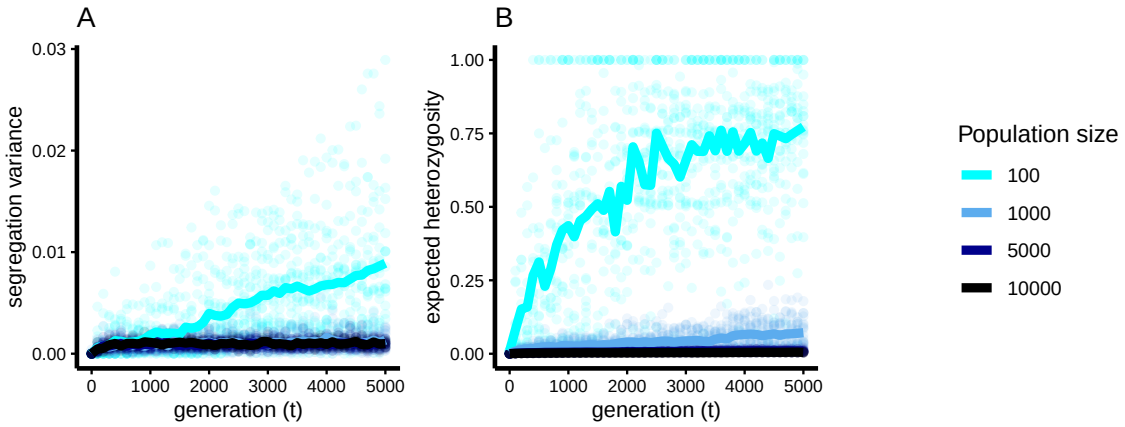


Figure A.15: **The effect of population size on the rate of divergence between populations due to drift**. I show populations held at a common optimum with no standing variation (i.e. $d = 0$, $N = 0$) and plot (A) segregation variance and (B) expected heterozygosity in hybrids over time for 5,000 generations. The evolution of segregation variance is proportional to the rate of evolution of reproductive isolation under parallel natural selection. Greater drift in smaller populations leads to greater segregation variance and heterozygosity. The lines are drawn as the average of 10 replicate simulations ($m = 5$, $\sigma = 1$).

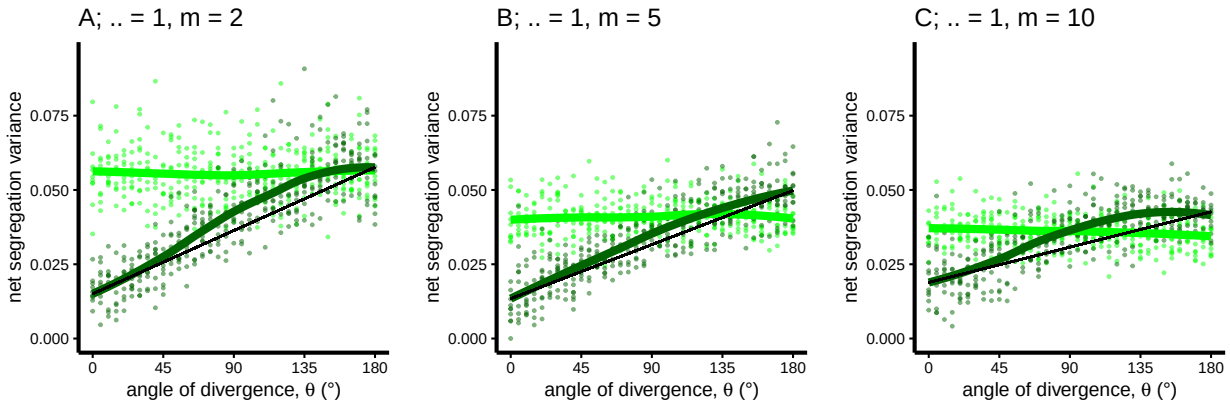


Figure A.16: **Effect of dimensionality on net segregation variance**. These plots are similar to Fig. 2.2C in the main text except I show results for three different dimensionalities. Under divergent natural selection, simulations where populations adapted from standing variation (dark green) had higher segregation variance—relative to simulations where populations adapted only from *de novo* mutation (light green)—in higher dimensions. Note the overall trend of a decrease in net segregation variance as dimensionality increases.

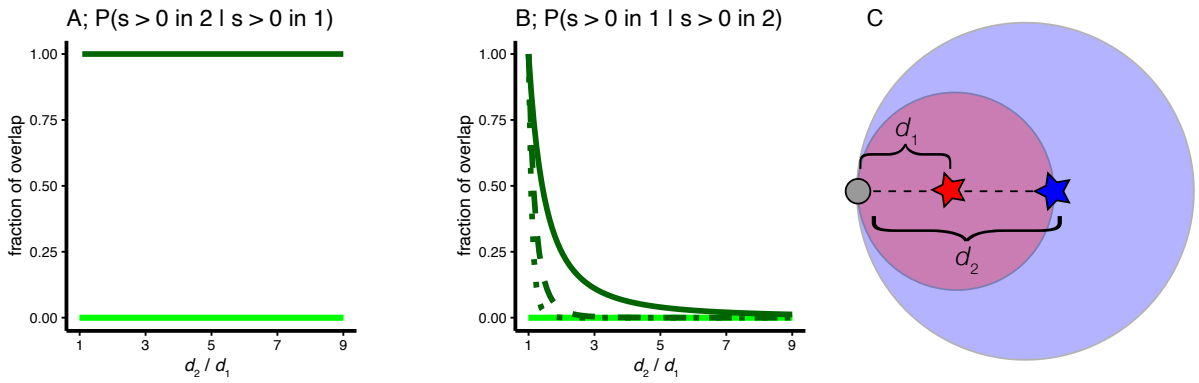


Figure A.17: **Fraction of overlap of beneficial mutations with parallel selection (θ = 0°) but unequal distance ($d_1 \neq d_2$).** The main text explores how the fraction of overlap changes with theta while holding $d_1 = d_2 = d$ constant. Here I explore how the fraction of overlap changes with the ratio d_2/d_1 when holding θ = 0° constant. Unlike the metric presented in the main text this metric is asymmetrical because one population is completely contained within the other. Panel (A) plots the fraction of overlap for population 1 (the fraction of alleles that are beneficial in population 1 are also beneficial in population 2) as a function of d_2/d_1. With d1 $<$ d2 the value is 1 for any ratio d_2/d_1 because population 1's hypersphere is contained within population 2's. Panel (B) plots the fraction of alleles that are beneficial in population 2 that are also beneficial in population 1. This latter result mirrors what is seen in the main text Fig. 2.3A: as the locations of the optima depart from one another the fraction of overlap rapidly approaches zero and does so most rapidly at the onset of departure. Panel (C) shows a cartoon example of a case in 2-dimensions where $d_2 = 2d_1$.

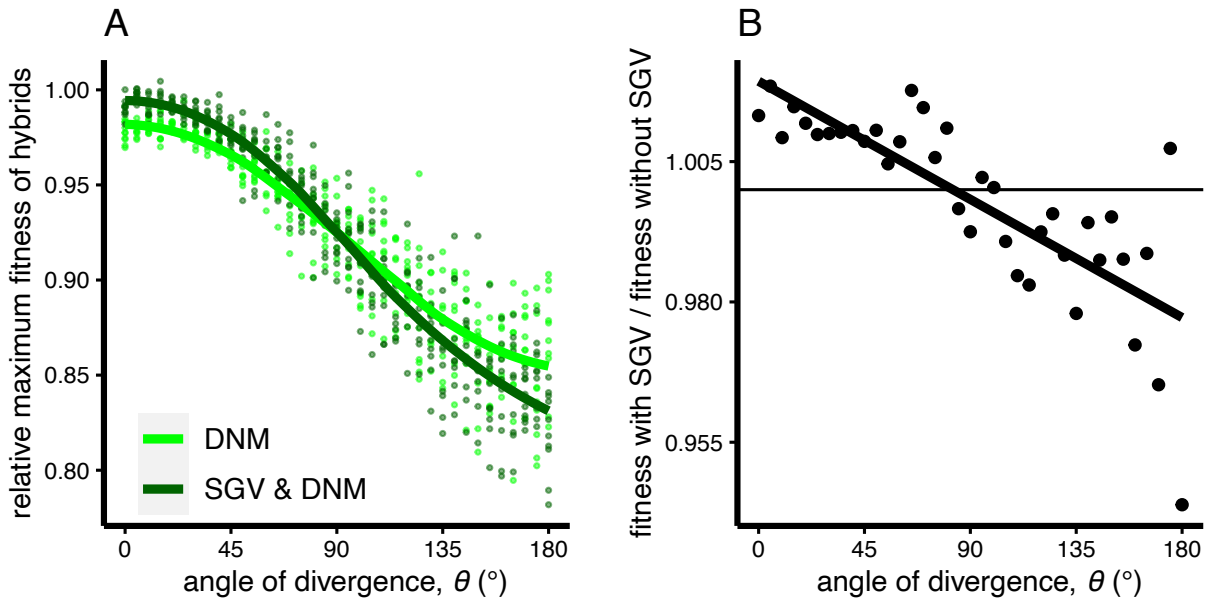


Figure A.18: **The effect of standing genetic variation (SGV) on relative maximum hybrid fitness across environments**. Data are from simulations plotted in the main text, but instead of mean fitness of all hybrids I depict the mean fitness of the top 5 % of hybrids relative to the mean fitness of parents. I plot both the (A) raw values of relative maximum fitness and (B) the effect of standing variation on maximum hybrid fitness (dark green divided by light green).

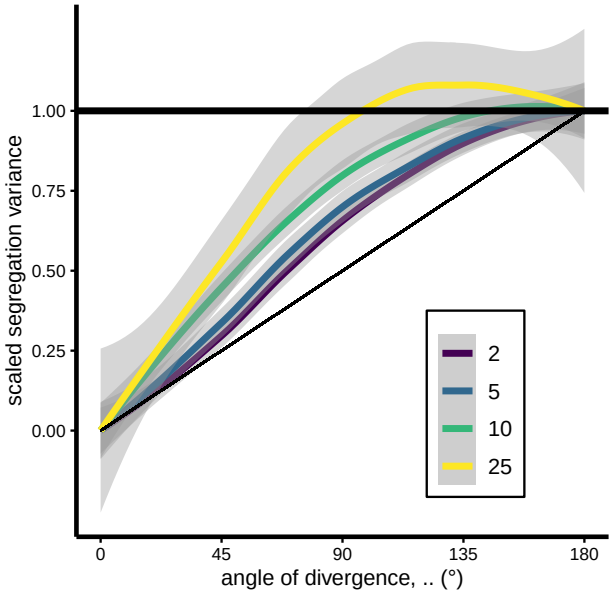


Figure A.19: **The relationship between segregation variance and θ for different dimensionalities**. I plot the loess fits of proof-of-concept simulation results with 95 % confidence intervals conducted in four different dimensionalities (colours), each scaled between 0 (at 0°) and 1 (at 180°). Simulations were conducted with strong natural selection (σ = 10) to minimize the effect of drift.

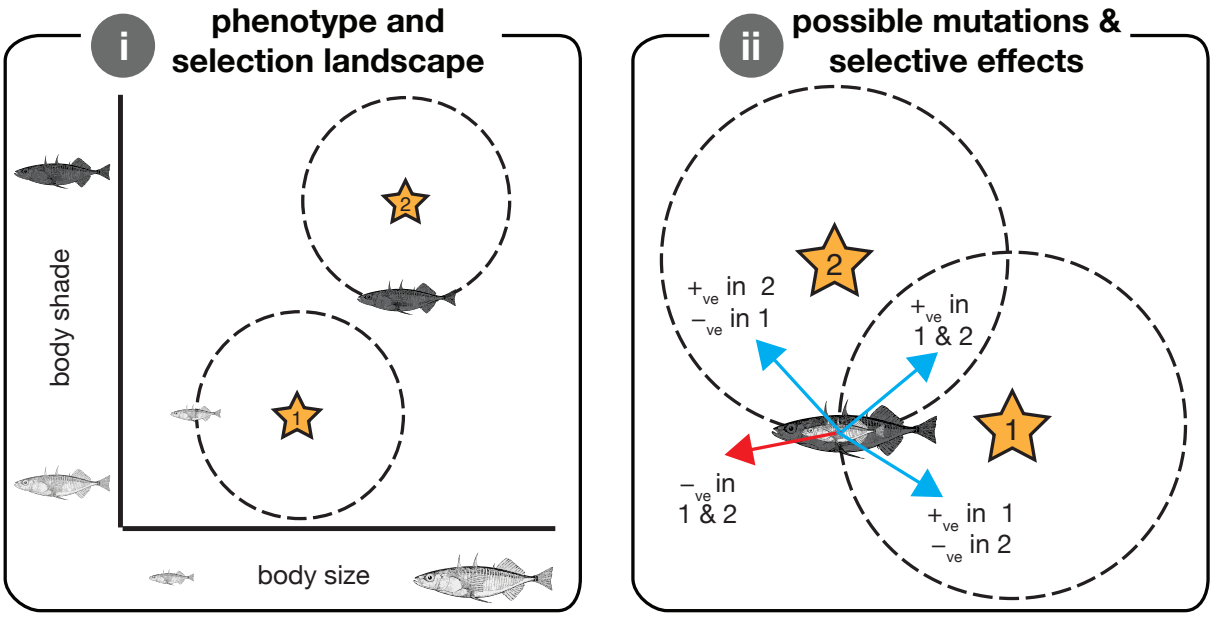


Figure A.20: **Cartoon illustration of why divergence among populations does not affect whether an allele is beneficial in both of them**. Panel (i) depicts the phenotype landscape and selection landscape. Variation in the horizontal dimension reflects phenotypic variation in body size, and the vertical dimension reflects variation in body shade. I depict two 'populations' with differences in body size and shade (small light; big dark). The stars reflect local optima after a hypothetical environmental shift—selection favours adaptation toward a larger body size in population 1 and selection for darker body shade in population 2. If I illustrate the circle of beneficial mutational space (dashed circles) with respect to the current phenotypic position they do not overlap. Panel (ii) illustrates the selection landscape as it is 'experienced' by each population. An allele that slightly increases body size and darkens the body shade from the current phenotype (the position of the fish cartoons) is beneficial (blue) in both of populations. Some alleles are beneficial in only one population, and others are deleterious in both (red). Thus, even though the spheres do not overlap in (i) it is not the case that they populations will undergo non-parallel genetic evolution.

Appendix B

Appendix for Chapter 3

B.1 Supplementary methods

B.1.1 Search strategy

I searched the literature for studies that made measurements of traits in F_1 hybrids and their parents. To identify studies for possible inclusion, I conducted a systematic literature search using Web of Science. I included all papers that resulted from a general topic search of "Castle-Wright", and from a topic search of "F_1 OR hybrid OR inherit*" in articles published (from any year) in *Evolution*, *Proceedings of the Royal Society B*, *Journal of Evolutionary Biology*, *Heredity*, or *Journal of Heredity*. These journals were selected because a preliminary search indicated that they contained nearly half of all suitable studies. These searches returned 82 of the 198 studies deemed suitable after screening. The literature search included all studies published until 31 December 2017.

To be more comprehensive, I then examined journals outside of the original group of 5 focal journals. I performed a forward search for articles that both cited key references (Dobzhansky 1937; Hubbs 1955; Mayr 1963; Grant 1981; Lande 1981; Tave 1986; Churchill and Doerge 1994; Bradshaw et al. 1998; Lynch and Walsh 1998; Hatfield and Schluter 1999; Schluter 2000; Coyne and Orr 2004) and contained the keywords described above. After screening, this search identified the remaining 116 studies of the 198 analyzed in my study. These searches collectively returned 14,048 studies and after removing duplicates this left 11,287 studies to be screened for possible inclusion. The full literature search results are available in the archived data.

Comments on systematic nature of review

I attempted to follow PRISMA (Moher et al. 2009) guidelines to the best of my ability. Most of the criteria have been addressed in the main text but a few other comments are warranted. In particular, I have no reason to suspect that any bias was introduced about dominance. This is because no studies seemed to have *a priori* hypotheses about such patterns. Accordingly, I do not believe that my estimates suffer from a file drawer problem. I emphasise that a formal meta-analytic framework—wherein data from multiple studies are aggregated with various weights—is not appropriate because I am not comparing studies that had any experimental treatment. Because the studies in my dataset do not have anything resembling an 'effect size', a simple summary across all of them is most appropriate.

B.1.2 Evaluation of studies

I required studies to meet several criteria to merit inclusion in the database. Most of these criteria are summarised in the main text, but additional details are given here. I excluded all domesticated species and most laboratory populations (or 'strains') because dominance patterns differ considerably following domestication (Crnokrak and Roff 1995). However, laboratory populations were included if founders were 10 or fewer generations removed from the wild. If populations were maintained in a lab for more than 10 generations but were found by comparison to still strongly resemble the source population, I included the study (n = 2 studies). I also excluded studies where the origin of the study populations was ambiguous. Hybrids had to be formed via the union of gametes from parental taxa, so I excluded studies that used techniques like somatic fusion. If authors had genotyped wild hybrids, I typically included hybrids where the probability of correct category assignment was $> 95\%$; in some cases the authors themselves used a different cut-off in which case I went with their cut-offs.

I only included traits that I thought were not directly linked to fitness. The majority of cases were not difficult to assess, but I have included reasons for excluding particular studies or traits in the database screening notes (see **Data accessibility** in main text).

Traits had to be measured in a quantitative manner to be included in the dataset. For example, if a trait was reported as being in categories related to parents or intermediacy ('parent-like' or 'intermediate'), I did not include it. Some traits such as mate choice must often be scored discretely (in the absence of multiple trials per individual), even though the trait can vary on independent trials. Accordingly, I included discretely scored traits—like mate choice—when it was possible in principle to obtain a different outcome on independent trials. Such traits are recorded as 0s and 1s, but hybrids can be intermediate if both outcomes occurred with equal frequencies. I included traits where authors devised their own discrete scale for quantification. When suitable data were collected by the authors but not obtainable from the article, I wrote to the authors and requested the data. If the author cited a dissertation as containing the data, I attempted to locate the data therein because dissertations are not indexed by Web Of Science. I included multivariate trait summaries (e.g., PC axis scores) when reported. If traits reported both the raw trait values and the PC axis scores for a summary of those same traits, I collected both sets of data but omitted the PCs in my main analyses.

Using the above criteria, I screened each article for suitability. As a first pass, I quickly assessed each article for suitability by reading the title and abstract and, if necessary, consulting the main text. After this initial search, I retained 407 studies. Since the previous steps were done by a team of five, the primary author conducted an in-depth evaluation of each study flagged for possible inclusion. If deemed suitable, I next evaluated whether the necessary data could be obtained. After this second assessment, 198 studies remained. The reasons for exclusion of each study are documented the archived data (see **Data accessibility** in main text).

B.1.3 Data collection

For each study, I recorded several types of data. First, I recorded the mean, sample size, and an estimate of uncertainty (if available; e.g., SD or variance) for each measured trait for each parental taxon and hybrid category (cross generation and/or direction). In most cases, these data were included in tables or could be

extracted from figures. For figure data extraction, I used 'WebPlotDigitizer' (Rohatgi 2019). In some cases, I contacted authors for the raw data or summary data.

Each study contributed a minimum of three records to the larger database: one trait measured in each parent and the F_1 generation. If the same traits were measured over ontogeny, I used only the final data point. When data were reported from multiple 'trials' or 'sites' I pooled them within and then across sites. If data were reported for different cross directions and/or sexes I recorded data for each cross direction/sex combination separately. I recorded whether each variable was a linear measurement (1D), area measurement (2D), volume measurement (3D; e.g., mass), categorical, or discrete. I did not find meaningful differences in dominance among trait types (ANOVA, $P = 0.998$) and so these data do not factor in to the present analyses. Data processing was greatly aided by the functions implemented in the `tidyverse` (Wickham 2017).

Occasionally, different studies analysed different traits of individuals from the same crosses. In these cases, I simply grouped them as being the same cross before analysis.

B.1.4 Estimating genetic divergence and divergence time

I estimated genetic distance for crosses where sequence data were available for both parents. I downloaded sequences in R using the `rentrez` package (Winter 2017), and retained up to 40 sequences per species. Sequences were then aligned with the profile hidden Markov models implemented in the `align` function in the package, `aphid` (Wilkinson 2018). After aligning sequences I calculated genetic distance by simply counting the number of sites that differed between two aligned sequences, implemented using the `raw` model option in the `dist.dna` function within `ape` (Paradis and Schliep 2018). I made this computation for each pair of sequences and then calculated the average over all pairs for one summary estimate of parental genetic divergence per cross.

I also used timetree (Kumar et al. 2017) to obtain estimates of divergence time for each species pair in their database in years. After obtaining estimates of divergence time I regressed divergence time against the response and predictor variables used in the main analysis. The conclusions from the timetree data do not differ from those using genetic distance and I do not discuss them further (see archived analysis code for these analyses).

B.1.5 Phylogenetic signal

In the data from the systematic review, I was interested in evaluating whether there was phylogenetic signal in dominance. I retrieved NCBI taxonomy IDs for my species using the `taxize` `R` package (Chamberlain and Szöcs 2013), and used these IDs (one arbitrarily chosen per cross) to generate a phylogeny using phyloT (https://phylot.biobyte.de/). Because branch lengths negligibly affect estimates of phylogenetic signal (Münkemüller et al. 2012), I assigned all branches equal lengths and used the `phylosig` function implemented in `phytools` (Revell 2012) to test for phylogenetic signal via Pagel's λ. Assigning random branch lengths never affected my conclusions. Because I am not testing a causal model, and also because there was no evidence of phylogenetic signal, I do not use the phylogeny in my main text analyses.

B.1.6 Field experiment with sunflowers

Whitney et al. (2010, 2006) generated artificial hybrids to resemble the presumed early ancestors of an existing natural hybrid sunflower, *Helianthus annuus* ssp. *texanus*, which grows in Texas, USA. For further details on the cross, experimental setup and trait measurements, see Whitney et al. (2010, 2006). The BC_1 generation was obtained by first mating *H. debilis* ssp. *cucumerifolius* from Texas to wild *H. annuus* ssp. *annuus* from Oklahoma to produce F_1 progeny in the greenhouse. In order to produce enough BC_1 seed for replicate field populations, a single progeny from the F_1 generation was propagated vegetatively to produce 14 F_1 clones. A single *H. a. annuus* pollen donor was mated to the F_1 clones to produce 3,758 BC_1 seeds.

To obtain seedlings for the field experiment, seeds were germinated on damp filter paper in late February 2003. Approximately six-day old seedlings were transplanted into peat pots containing field soil and grown in a greenhouse for four weeks before transplanting to the field at the Lady Bird Johnson Wildflower Center, Austin, Texas (hereafter LBJ; 30°10.886' N, 97°52.58' W). Prior to planting, plots were tilled to remove standing vegetation. All plants were planted at 90 cm spacing. Plots were fenced with plastic deer fencing to reduce disturbance by deer and rabbits. After planting, local vegetation was allowed to colonise the plots unhindered. BC_1 individuals were planted in the 'selection plot' of Whitney et al. (2006)), and parent individuals were planted into a second common garden plot approximately 500 m away from the BC_1 plants but within the same site (see Table S2 for sample sizes of parents and of BC_1s grown at that same second common garden.). A photograph of the experiment is included as Fig. B.16. A parallel experiment at a second site, the Brackenridge Field Laboratory, Texas, was analyzed and showed similar patterns as LBJ, but results are not reported here for brevity.

Plant traits and fitness were measured from March–September 2003. Viable seed production was chosen as the measure of fitness in these annual plants. Bags made from plastic mesh (DelStar Technologies, Delaware, USA) were secured onto flowerheads with twist-ties to prevent seed loss in the field. Flowerheads were bagged throughout the season (June to September) to obtain a representative sample. Seed production was estimated by multiplying the total number of heads (bagged & unbagged) by the average number of viable seeds per head in a pooled sample of the bagged heads. In addition to fitness, 30 traits comprising architectural, floral, ecophysiological, phenological, and herbivore resistance traits were measured on each plant, and those traits included herein after filtering are described in Table S1. Further details on trait measurement protocols and the relevance of individual traits to plant performance are given by Whitney et al. (2006, 2010).

B.2 Supplementary Tables & Figures

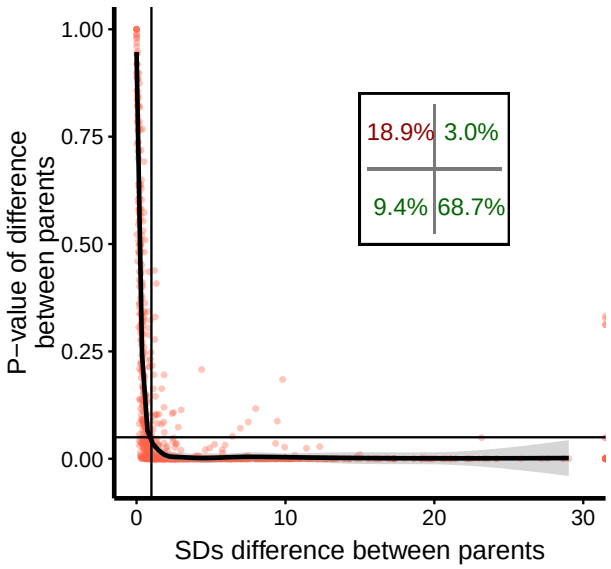


Figure B.1: **Relationship between SD divergence between parents (SDs in units of smaller parent value) and *P*-value of *t*-test evaluating whether parent phenotypes are statistically distinguishable.** The horizontal line is at $P = 0.05$ and the vertical line is at SD = 1. All traits except for those in the upper-left quadrant—with < 1 SD difference and $P > 0.05$— were used to measure dominance. The percentage of traits in each quadrant is shown in the inset. Few traits with > 1 SD divergence have non-significant *P*-values (upper-right quadrant).

Table B.1: Description of sunflower traits used to quantify parent-bias and mismatch

Trait Name	**Trait Description and units**	**Trait Category**
Height of uppermost branch	cm	Architectural
Height of lowest branch	cm	Architectural
Relative branch diameter	Mean primary branch basal diameter / stem diameter	Architectural
Plant volume	Volume of the main stem (cm3)	Architectural
Seed weight	Avg mass of individual seed	Architectural
Specific leaf area	ratio of leaf area to mass, $cm^2 \cdot g^{-1}$	Ecophysiological
Water-use efficiency	$\delta^{13}C$	Ecophysiological
Leaf Carbon:Nitrogen ratio		Ecophys. & palat.
Disk diameter	Diameter of the inflorescence disk (mm)	Floral
Ligule number		Floral
Ligule length	cm	Floral
Ligule width	cm	Floral
Phyllary length	mm	Floral
Phyllary width	mm	Floral
Bud initiation time	Time between planting and initiation of first inflorescence bud (days)	Phenological
Longevity	Time between planting and death	Phenological
Seed maturation time	Period between pollination and seed maturity (days)	Phenological
Glandular trichome density	On bottom surface of leaf; these contain sesquiterpene lactones (mm^2)	Palatability
Nonglandular trichome density	On bottom surface of leaf; physical rather than chemical defenses (mm^2)	Palatability

Table B.2: Mean ± SE (*n*) for parent species and BC_1 hybrids at the common garden site.

Trait*	*H. annuus***	BC_1	*H. debilis*
Bud initiation time	71.429 ± 2.886 (14)	70.378 ± 1.315 (45)	43.297 ± 1.803 (37)
Disk diameter	39.887 ± 1.498 (13)	32.649 ± 0.694 (45)	17.629 ± 0.35 (36)
Glandular trichome density	22.974 ± 2.143 (14)	15.752 ± 1.1 (45)	2.423 ± 0.585 (37)
Height of uppermost branch	132.786 ± 7.683 (14)	94.844 ± 3.496 (45)	17.183 ± 0.849 (36)
Height of lowest branch	44.429 ± 6.946 (14)	25.111 ± 2.584 (45)	8.417 ± 0.733 (36)
Leaf Carbon:Nitrogen ratio	10.213 ± 0.52 (7)	9.907 ± 0.202 (18)	9.487 ± 0.476 (18)
Ligule length	39.435 ± 1.072 (14)	37.222 ± 0.641 (45)	23.523 ± 0.629 (37)
Ligule number	22.571 ± 0.824 (14)	20.156 ± 0.218 (45)	14.568 ± 0.314 (37)
Ligule width	11.265 ± 0.412 (14)	12.511 ± 0.288 (45)	9.537 ± 0.204 (37)
Longevity	169.357 ± 4.435 (14)	183.111 ± 1.729 (45)	188.515 ± 6.051 (33)
Nonglandular trichome density	15.965 ± 1.117 (14)	17.369 ± 0.647 (45)	12.722 ± 0.774 (37)
Phyllary length	21.234 ± 0.619 (14)	20.31 ± 0.524 (45)	11.857 ± 0.264 (37)
Phyllary width	9.254 ± 0.47 (14)	7.673 ± 0.191 (45)	2.045 ± 0.047 (37)
Plant volume	636.321 ± 89.82 (14)	378.758 ± 24.027 (45)	39.886 ± 6.162 (37)
Relative branch diameter	0.28 ± 0.011 (14)	0.305 ± 0.007 (45)	0.538 ± 0.017 (37)
Seed maturation time	29.923 ± 0.836 (13)	23.8 ± 0.349 (45)	19.25 ± 0.42 (36)
Seed weight	9.842 ± 0.268 (14)	7.401 ± 0.177 (45)	2.146 ± 0.076 (37)
Specific leaf area	140.693 ± 5.923 (14)	137.769 ± 2.939 (45)	164.722 ± 3.866 (37)
Water-use efficiency	−28.704 ± 0.299 (7)	−28.638 ± 0.058 (18)	−29.317 ± 0.188 (18)

*For trait descriptions, see Table S1. I include only a subset of BC_1 individuals here, which represent those grown in the exact same location as parents. The 475 BC_1s analyzed for fitness were grown nearby.

**Values differ from those reported for three *H. a. annuus* populations in Whitney et al. (2006, 2010) because the values here refer to the single population, RAR59, which served as a cross parent to the BC_1 hybrids.

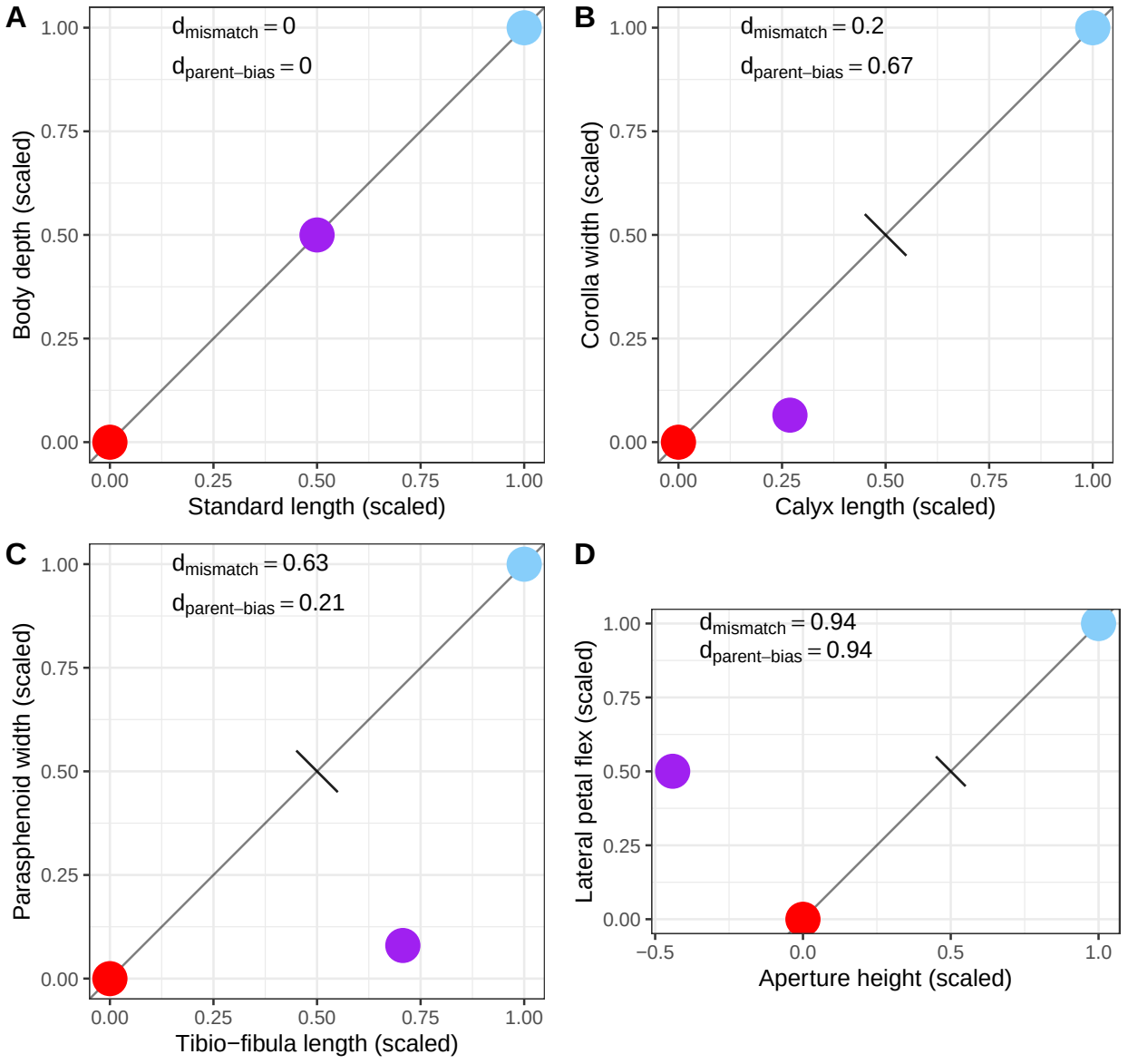


Figure B.2: **Example visualisations of four cases of bivariate trait expression**. Here I show scaled trait values of parents and hybrids. Panel (A) shows a case where hybrids are intermediate for both traits (Hatfield and Schluter 1999). Panel (B) shows a case where hybrids are similar to one parent for both traits (Fishman et al. 2015). Panel (C) shows a case where hybrids are similar to one parent for one trait and to the other for the second trait (Lamb and Avise 1987). Panel (D) shows a case where the hybrids are transgressive for one trait and intermediate for the other (Bradshaw et al. 1998; 'LC' cross). I show values for mismatch and parent-bias in each case.

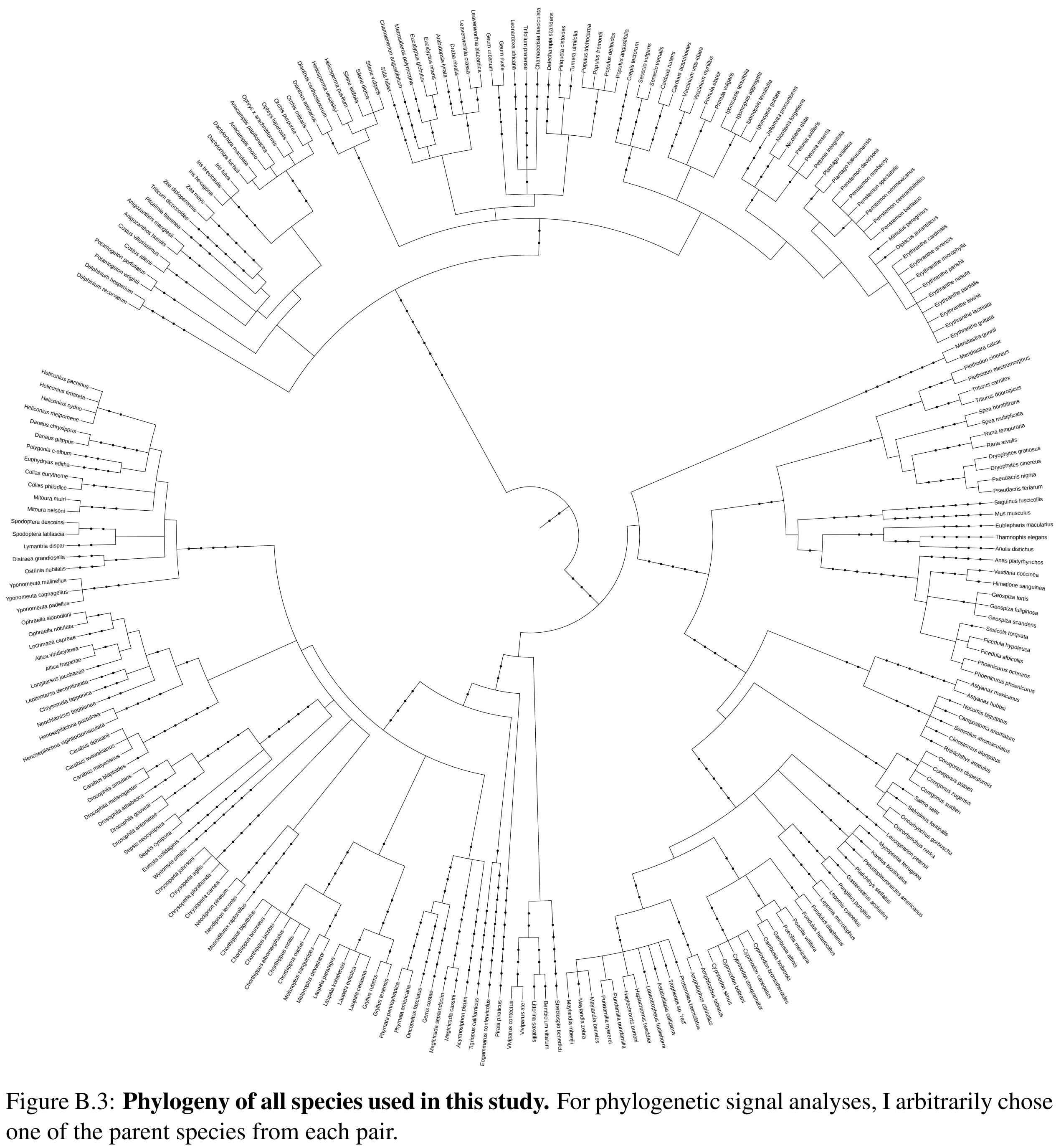


Figure B.3: **Phylogeny of all species used in this study.** For phylogenetic signal analyses, I arbitrarily chose one of the parent species from each pair.

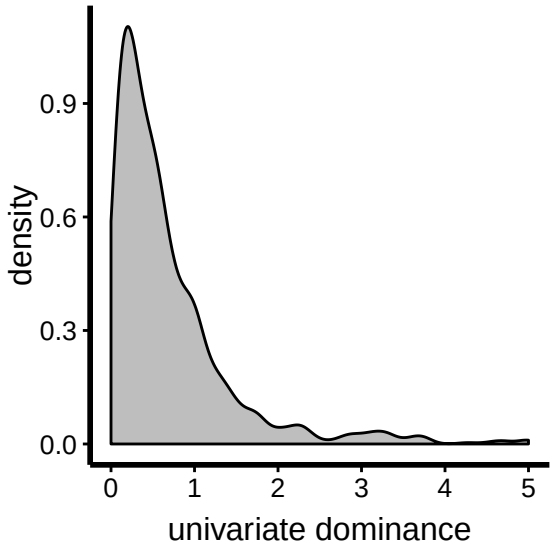


Figure B.4: **$d_{univariate}$ across all traits in the datasets (*n* = 1046 traits).** The *x*-axis is truncated at $d_{univariate}$ = 5.

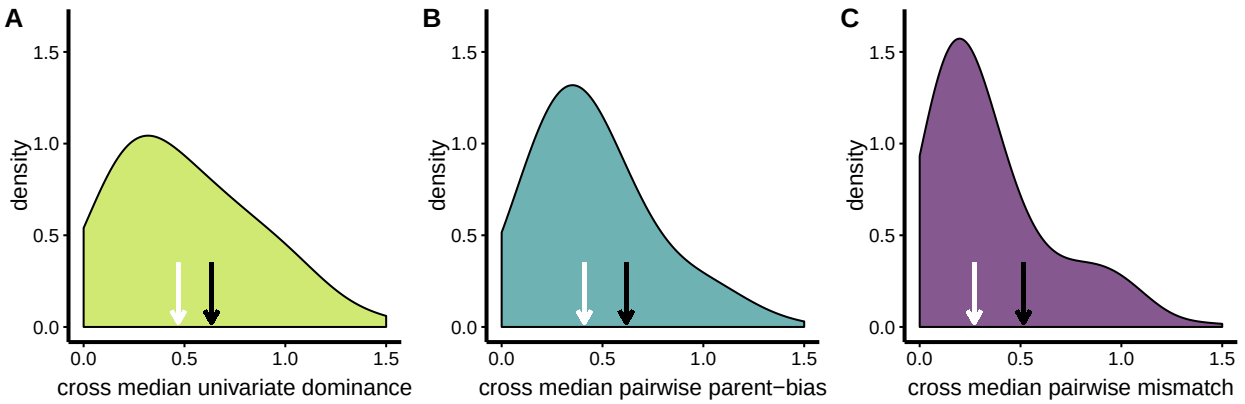


Figure B.5: **Summary of cross median dominance metrics with each cross contributing a single value.** Everything is the same as Fig. 2.2 of the main text, except here each cross contributed the median value instead of the mean for each dominance metric. The density plots (y-axis standardized across panels) show the three main dominance metrics contained herein, with each cross contributing a single value per panel. Values of 0 indicate no dominance, values of 1 indicate the maximum without transgression, and values $>$ 1 indicate transgression. The x-axis is truncated at 1.5, but the mean of cross medians (black arrows) and median of cross medians (white arrows) are calculated from the whole dataset. Panel **a** shows the univariate dominance ($d_{univariate}$; eqn. 1), panel **b** shows parent-bias (pairwise $d_{parent\text{-}bias}$; eqn. 3), and panel **c** shows mismatch (pairwise $d_{mismatch}$; eqn. 4). Panel **a** contains one value from all crosses (n = 233) while panels **b** and **c** only contain information from crosses wherein two or more traits were measured (n = 165).

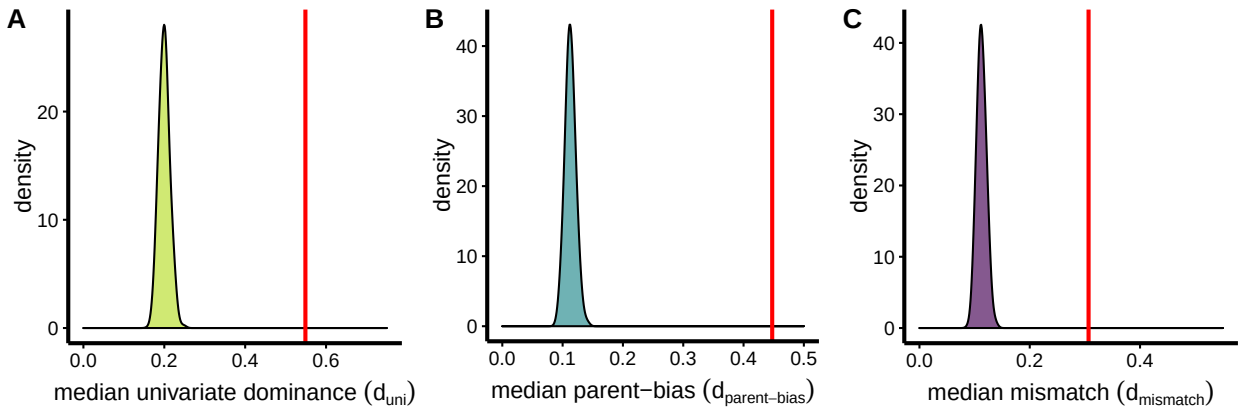


Figure B.6: **Expected patterns based on sampling error alone (from simulations).** The vertical lines are the median values observed in the main text. Density plots are the distribution of median values (n = 1000 simulations) of each metric in simulations where the true (scaled) mean was 0.5 for each trait and individual phenotypes were simulated with an identical SD and identical sample sizes to the real data. These simulations reveal the extent of 'dominance' caused by sampling error alone.

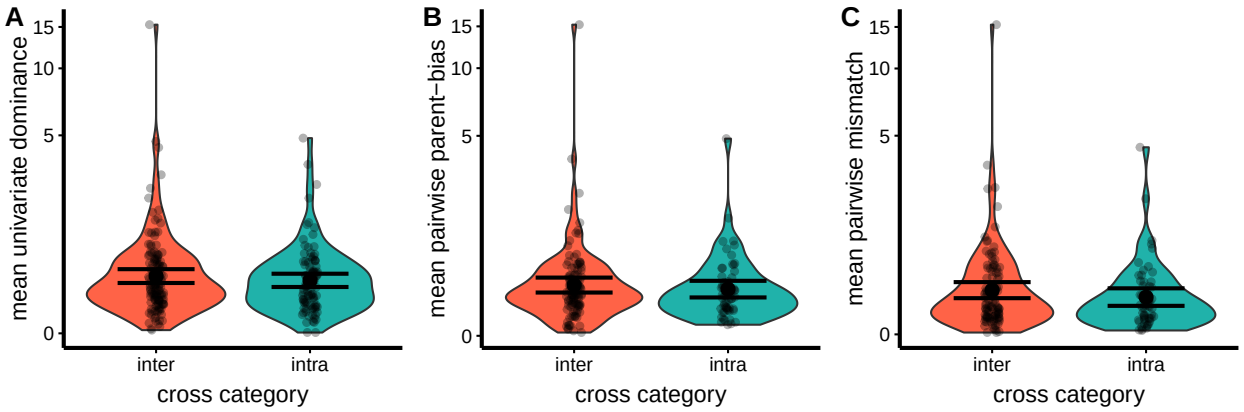


Figure B.7: **There are no differences (all $P > 0.25$) in any dominance metrics between intraspecific and interspecific crosses.** Each point is the cross mean dominance metric for $d_{univariate}$ (panel **a**), pairwise $d_{parent\text{-}bias}$ (panel **b**), and pairwise $d_{mismatch}$ (panel **c**). All $P > 0.5$. Note that the y-axis is ln-transformed.

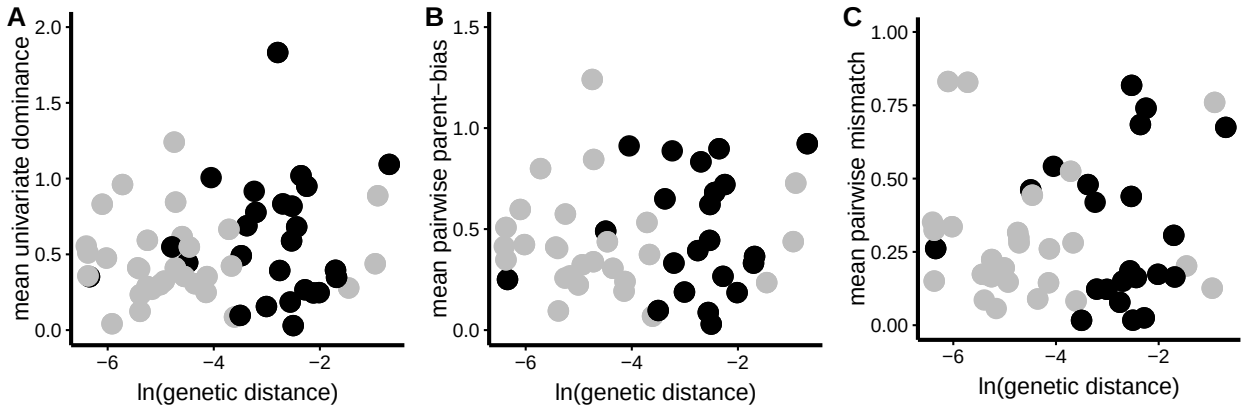


Figure B.8: **No association between any dominance metrics and genetic distance between the parents.** Divergence time was calculated from nucleotide sequences (see methods). Each point is the dominance metric for a cross (calculated following eqns 1–4 in the main text). Grey points are plants, and black points are animals. All $P > 0.5$.

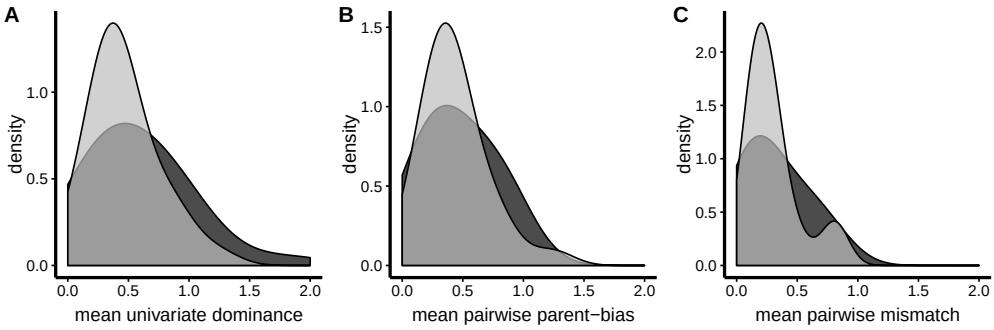


Figure B.9: **No differences in dominance between plants and animals.** Each point is the dominance metric for a cross (calculated following eqns 1–4 in the main text). All $P > 0.08$. Light grey density plot represents plants, and dark grey represents animals.

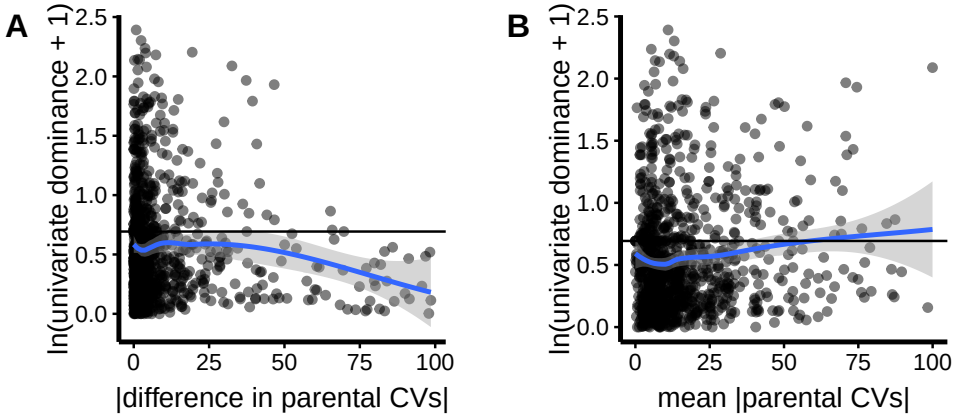


Figure B.10: **Parental phenotypic variation (CV) has significant effects on dominance in hybrids.** Panel (A) shows the relationship between the absolute difference in parent CVs (i.e., $|CV_1 - CV_2|$) and univariate dominance ($P = 0.035$). Panel (B) shows the relationship between the mean of parental CVs and univariate dominance ($P = 0.0092$). These analyses are simple linear models; only the pattern shown in (B) remains statistically significant after accounting for study as a random effect.

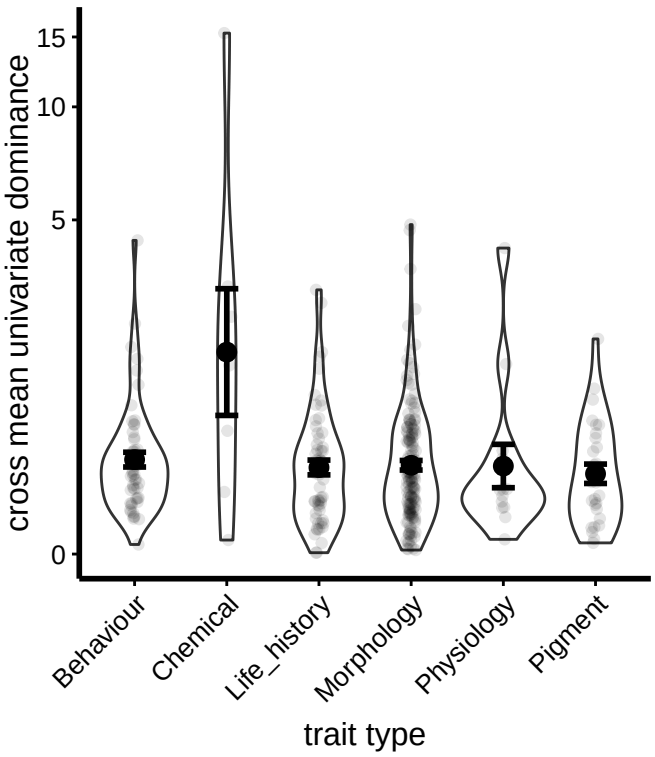


Figure B.11: **The relationship between trait type and univariate dominance ($d_{univariate}$).** Chemical traits have a significantly higher mean $d_{univariate}$ than all other trait categories (all $P < 0.05$) traits, which do not differ from one another ($P > 0.9$). This pattern is entirely caused by a substantial outlier.

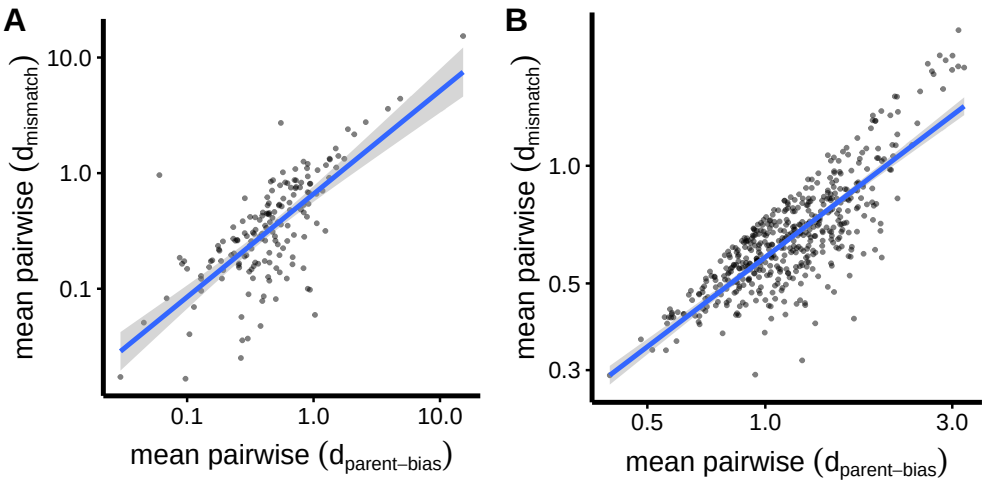


Figure B.12: **Relationship between mean pairwise parent-bias and mismatch dominance in systematic review and sunflower data.** Panel **a** shows the distribution of pairwise $d_{parent\text{-}bias}$ and $d_{mismatch}$ in the systematic review data (F_1s). Panel **b** shows the same metrics calculated at the individual level in the sunflower field experiment data (BC_1s). Both relationships are statistically significant. Note the $\log_{10}$-scale.

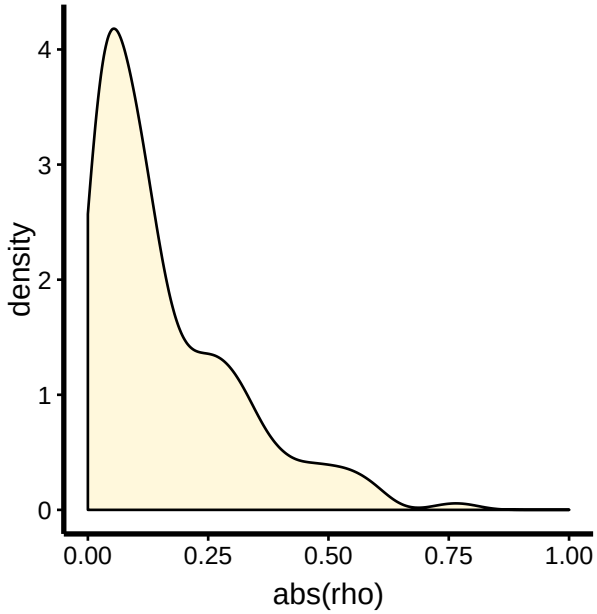


Figure B.13: **Distribution of pairwise trait correlations in the sunflower data (BC_1s only).** Since most correlations are weak, I conclude that it is unlikely issues caused by trait correlations undermine any of my conclusions.

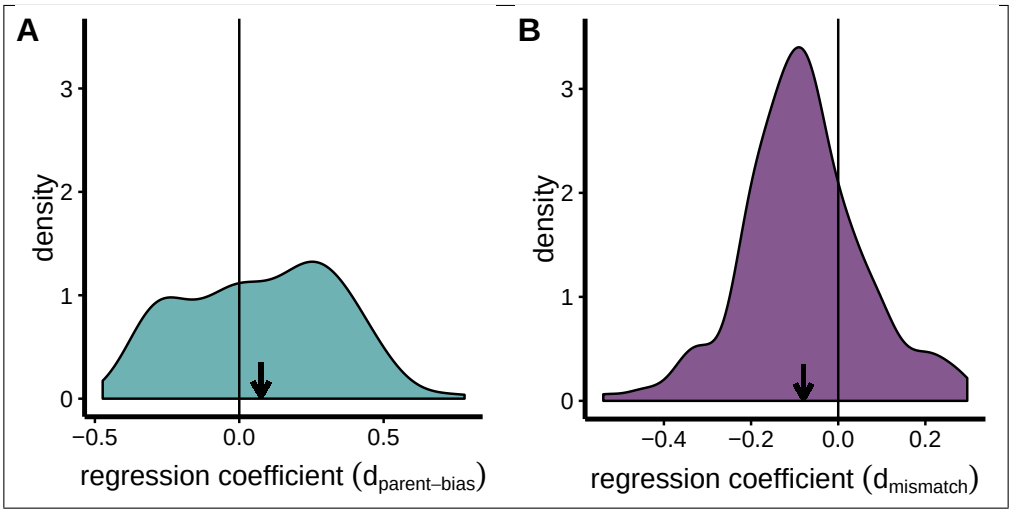


Figure B.14: **Distribution of partial regression coefficients in multiple regression analyses (see eqn. 5).** The thin vertical line demarcates a slope of zero. The black arrows show the mean of pairwise regression coefficients (plotted in the density plot). All coefficients are shown, not just those that are significant (*n* trait pairs = $\binom{19}{2} = 171$).

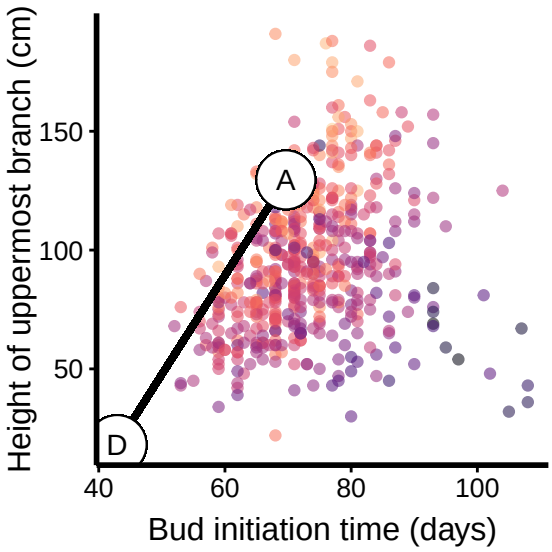


Figure B.15: **Phenotype and fitness distribution for a trait pair with substantial mismatch consequences in BC_1 hybrid sunflowers.** Darker values indicate lower $\log_{10}$(seed count). Large circles labelled *D* and *A* represent mean bivariate phenotypes of the parent species, *Helianthus debilis* and *H. annuus*, respectively. Individual plants that resemble the *H. debilis* parent in stature (i.e., are compact with branches clustered at the base of the plant) but resemble the *H. annuus* parent in phenology (i.e. are slow-developing) have particularly low fitness.

Figure B.16: **Photograph of sunflower experiment at the Lady Bird Johnson Wildflower Center, in Austin, Texas, USA.**

Appendix C

Appendix for Chapter 4

Please note that some of the information contained herein is redundant with that in Appendix B. However, I have elected to reproduce their supplementary materials in a format so that they can be consulted independently and so that each appendix contains all information that is relevant to each paper.

Search strategy

I searched the literature for studies that made measurements of traits in F_1 hybrids and their parents. To identify studies for possible inclusion, I conducted a systematic literature search using Web of Science (https://www.webofknowledge.com/). I included all papers that resulted from a general topic search of "Castle-Wright", and from a topic search of "F_1 OR hybrid OR inherit*" in articles published in *Evolution*, *Proceedings of the Royal Society B*, *Journal of Evolutionary Biology*, *Heredity*, or *Journal of Heredity*. These journals were selected because a preliminary search indicated that they contained nearly half of all suitable studies. These searches returned 106 studies deemed suitable after screening. To be more comprehensive, I conducted additional systematic searches by conducting similar topic searches among articles citing influential and highly-cited publications (Dobzhansky 1937; Hubbs 1955; Mayr 1963; Grant 1981; Lande 1981; Tave 1986; Churchill and Doerge 1994; Bradshaw et al. 1998; Lynch and Walsh 1998; Hatfield and Schluter 1999; Schluter 2000; Coyne and Orr 2004). The full literature search results are available in the archived data. My initial search returned 14048 studies, and after removing duplicates this left 11287 studies to be screened for possible inclusion. This literature search was primarily done for another unpublished study with the goal of understanding phenotype expression in F_1 hybrids.

Evaluation of studies

I required studies to meet several criteria to merit inclusion in my database. First, the study organisms had to originate recently from a natural (i.e., 'wild') population. This is because dominance patterns in domestic species differ substantially from non-domesticated species (Crnokrak and Roff 1995) and because I am explicitly interested in patterns as they occur in nature. I excluded studies using crops, domestic animals, laboratory populations that were > 10 (sexual) generations removed from the wild, or where populations were subject to artificial selection in the lab. If populations were maintained in a lab for more than 10 generations but were found by comparison to still strongly resemble the source population, I included the study ($n = 2$). I also excluded studies where the origin of the study populations was ambiguous. Hybrids had to be formed via the union of gametes from parental taxa, so I excluded studies using techniques like somatic fusion. Second, the ancestry of hybrids had to be clear. Many studies reported phenotypes of natural hybrids, for example in hybrid zones. I did not include these studies unless the hybrid category (i.e., F_1, F_2,

backcross) was confidently determined with molecular markers (typically over 95 % probability, unless the authors themselves used a different cut-off in which case I went with their cut-off) or knowledge that hybrids were sterile and thus could not be beyond the F_1).

Third, because I was interested in the inheritance of traits that are proximally related to organismal performance (McGee et al., 2015), I required studies to report measurements of at least one 'non-fitness' trait ('ordinary' traits [Orr 2001]) . Non-fitness traits (hereafter simply 'traits') are those that are likely under stabilizing selection at their optimum, whereas 'fitness' traits are those that are likely under directional selection and have no optimum (Merilä and Sheldon 1999; Schluter et al. 1991). In most cases it was possible to evaluate this distinction objectively because authors specifically referred to traits as components of fitness, reproductive isolating barriers, or as being affected by non-ecological hybrid incompatibilities. In some cases, however, I made the distinction myself. If particular trait values could be interpreted as resulting in universally low fitness, for example resistance to herbivores or pathogens, this trait was not included. The majority of cases were not difficult to assess, but I have included reasons for excluding particular studies or traits in the database screening notes (see Data accessibility).

Traits had to be measured in a quantitative manner to be included in the dataset. For example, if a trait was reported categorically (e.g., 'parent-like; or 'intermediate'), I did not include it. Some traits such as mate choice must often be scored discretely (in the absence of multiple trials per individual), even though the trait can vary on independent trials. Accordingly, I included discretely scored traits — like mate choice — when it was possible in principle to obtain a different outcome on independent trials. Such traits are recorded as 0s and 1s, but hybrids can be intermediate if both outcomes occurred with equal frequencies. I included traits where authors devised their own discrete scale for quantification. When suitable data were collected by the authors but not obtainable from the article, I wrote to the authors and requested the data. If the author cited a dissertation as containing the data, I attempted to locate the data therein because dissertations are not indexed by Web Of Science. I included multivariate trait summaries (e.g., PC axis scores) when reported. If traits reported both the raw trait values and the PC axis scores for a summary of those same traits, I collected both sets of data but omitted the PCs in my main analyses.

Using these criteria, I screened each article for suitability. As a first pass, I quickly assessed each article for suitability by reading the title and abstract and, if necessary, consulting the main text. After this initial search, I retained 407 studies. Since the previous steps were done by a team of five, I personally conducted an in-depth evaluation of each study flagged for possible inclusion. If deemed suitable, I next evaluated whether the necessary data could be obtained. After this second assessment, 198 studies remained. The reasons for exclusion of each study are documented the archived data (see Data accessibility).

Data collection

For each study, I recorded several types of data. First, I recorded the mean, sample size, and an estimate of uncertainty (if available) for each measured trait for each parental crosses and hybrid category. In most cases, these data were included in tables or could be extracted from figures. In some cases, I contacted authors for the raw data or summary data. Each study conducted a minimum of three records to the larger database: one trait measured in each parent and the F_1 generation. Traits were categorized as one of: behaviour, chemical,

life history, morphological, physiological, or pigmentation. If the same traits were measured over ontogeny, I used only the final data point. When data were measured in multiple 'trials' or 'sites' I pooled them within and then across sites. If data were reported for different cross directions and/or sexes I recorded data for each cross direction / sex combination separately. Data processing was immeasurably aided by the functions implemented in the `tidyverse` (Wickham 2017).

For each paper I recorded whether the phenotypes were measured in the lab or field, if in the lab the number of generations of captivity, and whether a correlation matrix (preferably in recombinant – F_2 or BC – hybrids; see below) was available or calculable from the raw data or figures. For the present study, specifically, each study would have had to contribute 8 or more datapoints – two traits from each of P_1, P_2, F_1, and F_2. Occasionally, different studies analysed different traits from individuals from the same crosses. In these cases, I simply grouped them as being the same study before analysis.

Comments on systematic nature of review

I attempted to follow PRISMA (Moher et al. 2009) guidelines to the best of my ability. Most of the criteria have been addressed above but a few other comments are warranted. I have no reason to suspect that any bias was introduced about estimates of parental divergence or segregation variance. This is because no studies seemed to have *a priori* hypotheses about such patterns. Accordingly, I do not believe that my estimates suffer from a file drawer problem, since detecting segregation variance in non-divergent traits was not the stated goal of any contributing studies. In addition, a formal meta-analytic framework here is not appropriate because I am not comparing studies that had any experimental treatment.

Estimating genetic divergence and divergence time

I estimated genetic distance for pairs of species where data were available for both parents. A preliminary screening revealed that the internal transcribed spacer (ITS I and II) was the most commonly available gene for plants and cytochrome b was the most available gene for animals in my dataset. I downloaded sequences in R using the `rentrez` package (Winter 2017), and retained up to 40 sequences per species. Sequences were then aligned with the profile hidden Markov models implemented in the `align` function in the package, `aphid` (Wilkinson 2018). After aligning sequences I calculated genetic distance by simply counting the number of sites that differed between two aligned sequences, implemented using the `raw` model option in the `dist.dna` function within `ape` (Paradis and Schliep 2018).

I also used timetree (http://timetree.org; Kumar et al. 2017) to obtain estimates of divergence time for each species pair in their database in years. After obtaining estimates of divergence time I regressed divergence time against the response and predictor variables used in the main analysis.

Phylogenetic signal

To determine whether patterns might be spurious and caused by differences among taxa, I wished to see if there was phylogenetic signal in the data. If either my response or predictor variables co-varied with

phylogeny this might indicate that phylogenetic independent contrasts or similar is necessary for analysis. I retrieved NCBI taxonomy IDs for my species using the `taxize R` package (Chamberlain and Szöcs 2013), and used these IDs (one arbitrarily chosen per cross) to generate a phylogeny using phyloT (https://phylot.biobyte.de/). Because branch lengths negligibly affect estimates of phylogenetic signal (Münkemüller et al. 2012), I assigned all branches equal lengths and used the `phylosig` function implemented in `phytools` (Revell 2012) to test for phylogenetic signal via Pagel's λ.

C.1 Supplementary figures

Although the analysis presented in the main text is, in my view, the most justifiable, there were several subjective decisions that I made when going from raw data to summary statistics. Accordingly, I wished to investigate the extent to which my findings were robust to making alternative choices. In this section I present supplementary figures that illustrate results described in the main text. In addition, I present a simple table showing the Spearman's ρ coefficient and P-value for the correlation between phenotypic divergence in divergent traits and segregation variance in non-divergent traits under many alternative data filtering/binning/transformation decisions. In all but one case, the pattern remains statistically significant. Accordingly, I conclude that the qualitative patterns observed are quite robust.

There were correlations between sample size and some of my response variables, including estimates of parental divergence. To determine if my results were robust to the exclusion of potentially low-power studies, I generated two new datasets with studies measuring fewer than (i) 20 or (ii) 70 parental individuals excluded. Sample size did not predict parental divergence for these datasets. These results are presented below in Table S1 and indicate that sample size is not responsible for the pattern. I also note that a multiple regression with parental divergence and sample size as predictors — flawed because of potential heteroskedasticity and low data - predictor ratio — indicated that only parental divergence was a significant predictor of segregation variance (divergence $P = 0.0195$; sample size $P = 0.755$).

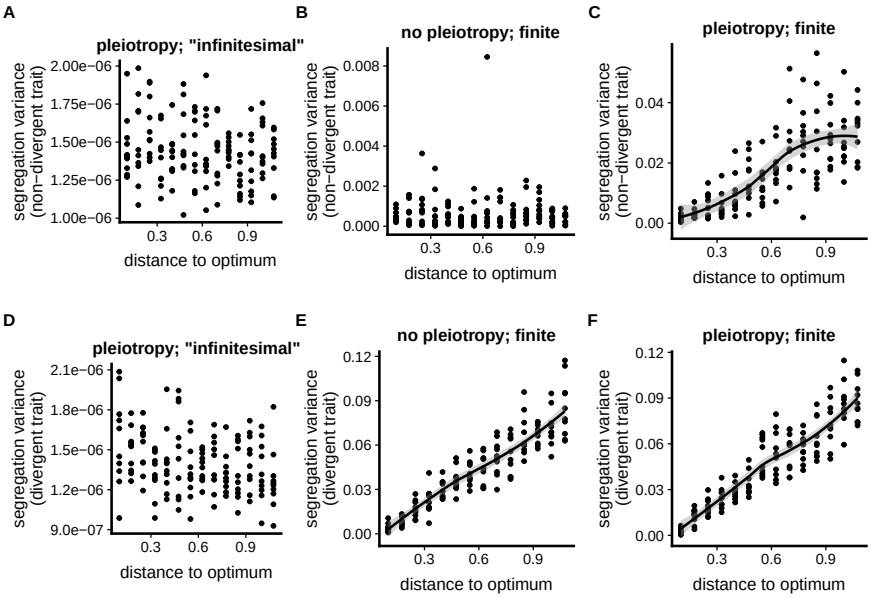


Figure C.1: **Simulation results to illustrate theoretical prediction.** I conducted individual based simulations using methods similar to those described fully in Thompson et al. (2019), which is open access. Briefly, two initially identical populations diverged without gene flow between them for 1000 generations. All mutations were unique to each population (no parallelism). Individuals had two traits and only trait 1 was under selection. Optima were defined as [d, 0] for population 1 and [-d, 0] for population 2, where d is the distance to the optimum from the origin. I vary d along the horizontal axis in all figures. After 1000 generations I made interpopulation hybrids and measured the variance in traits 1 (y-axis, top row) and 2 (y-axis, bottom row). Panels **A** and **D** show simulations where populations adapt from pleiotropic but very small alleles. Panels **B** and **E** show the case where mutations are appreciably large but not pleiotropic—only affecting trait 1 or trait 2 but never both. Panels **C** and **F** show the case where mutations are appreciably large-effect and can affect both traits simultaneously. Simulation code is archived online. Simulations are haploid and so the F_1 variance is the segregation variance. Here is what I would like you to get from the figure:(1) When mutations are small, there is no relationship between d and segregation variance in any trait, even with universal pleiotropy. (2) When mutations are large but there is no pleiotropy, a relationship between d and and segregation variance is observed only for the divergently-selected trait. And (3), with pleiotropic mutations of reasonably large effect, I see a relationship between d and and segregation variance for both traits. The only parameter that varies (other than the presence of pleiotropy) is mutation effect size; set to 10^{-6} in panels **A** and **D** and and 0.1 in the others.

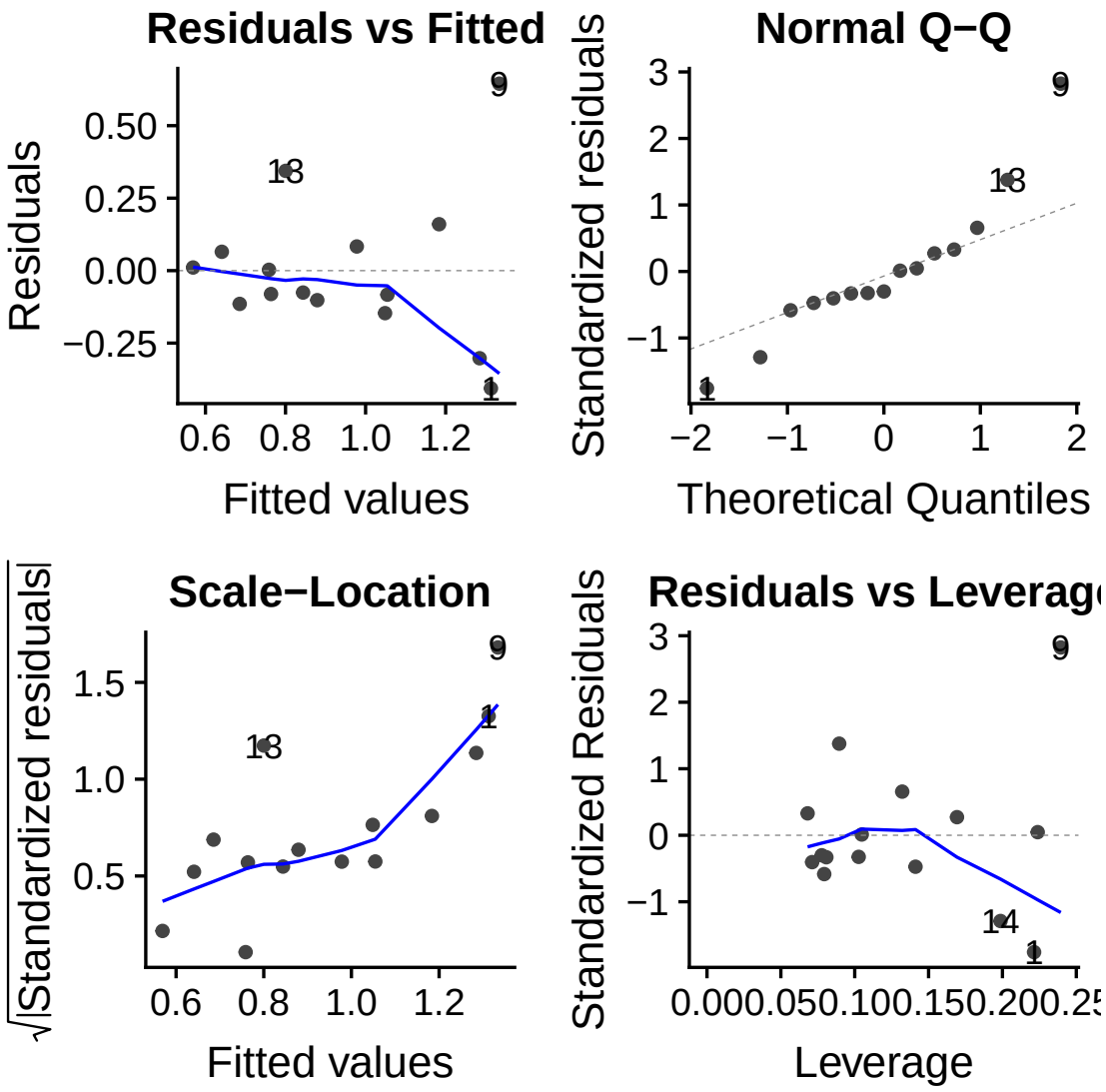


Figure C.2: **Diagnostics of the linear model testing the main hypothesis.** The figure was produced with the `autoplot` function in the `ggfortify` R package. Although the linear regression is significant ($F_{1,12}$ = 10.03, r^2 = 0.455, P = 0.00812), the diagnostic plots reveal significant heteroskedasticity (Breusch-Pagan test P = 0.00182). This plot is meant to convince the reader that a non-parametric approach is better-suited to test the hypothesis at hand.

Table C.1: **Results with alternative choices for data filtering and binning.**

Difference	Description	Spearman's ρ	P
Main text analysis	For reference	0.8	0.0005809
Less strict binning	Traits with ≥ 1 SD difference between parents deemed 'divergent'	0.714	0.00543
Same segvar fomula; difference	Eqn. 1 used but difference not ratio	0.729	0.00286
Alternative segvar fomula; ratio	segvar = var(F_2) / var(F_1)	0.607	0.0186
Alternative segvar fomula; difference	segvar = var(F_2) $-$ var(F_1)	0.607	0.0186
All traits included	No longer restricted to morphology	0.518	0.0523
Most traits included	All traits except for physiology & chemical (morphology, behaviour, pigment, life history)	0.451	0.0364
No transformation	No log transformation of variables	0.746	0.00202
Sample size filter #1	No studies with fewer than 20 parent individuals	0.736	0.00557
Sample size filter #2	No studies with fewer than 70 parent individuals	0.762	0.0368

Note: All analyses with morphological traits only and binning based on statistical tests except where noted.

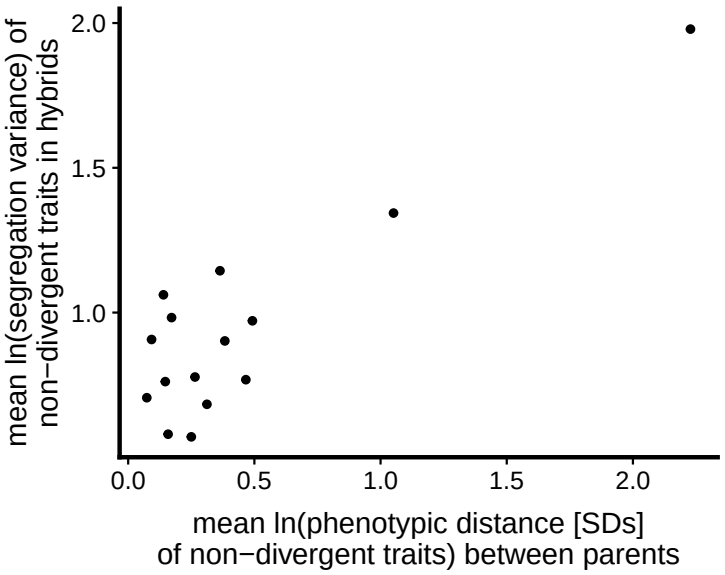


Figure C.3: **Segregation variance in non-divergent traits is not predicted by divergence in those traits.** This analysis complements the analysis in the main text. A Spearman's rank-order correlation is non significant ($\rho = 0.446$; $P = 0.0972$).

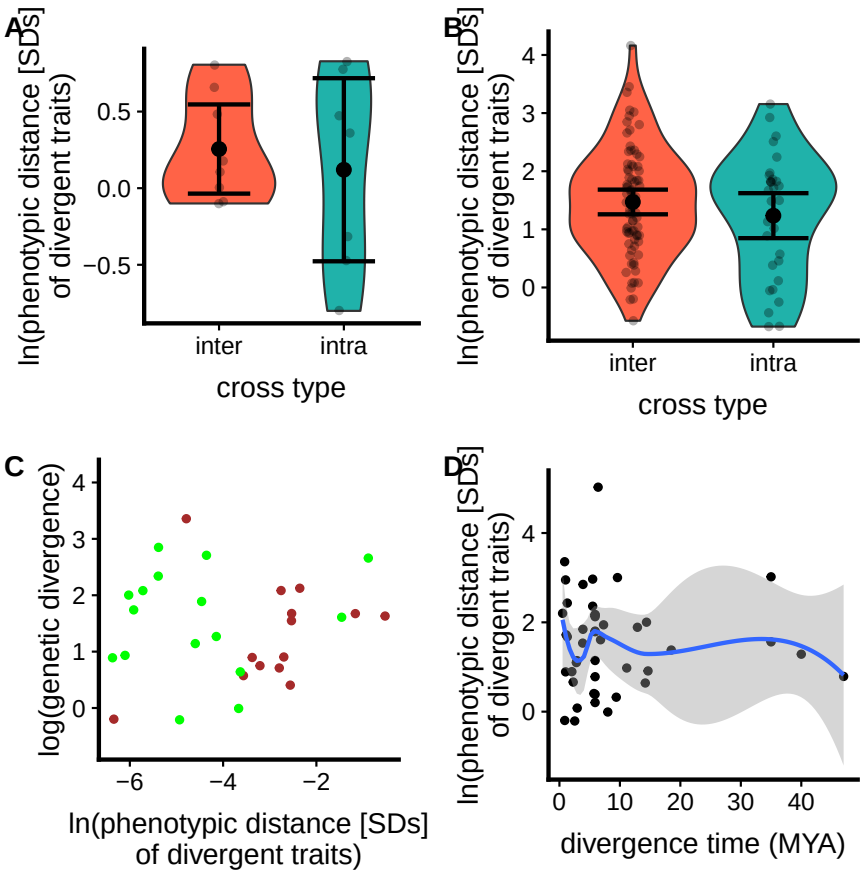


Figure C.4: **All available evidence suggests that phenotypic divergence between parents is uncorrelated with their divergence time.** Panels **A** and **B** use intra- vs interspecific crosses as a proxy for divergence time. Panel **A** compares the taxa analyzed in the main text ($P > 0.9$) and panel **B** uses a larger dataset of 198 studies ($P = 0.268$). Panel **C** uses continuous genetic distance between each pair of parents for which I could obtain DNA sequences in the larger database of 198 studies (green points = plants, brown points = animals); sequence divergence does not predict parental phenotypic divergence ($P = 0.746$). Panel **D** uses estimates of divergence time from timetree for all pairs of species for which data were available; there is no relationship ($P = 0.93$). I calculated a unique value for each unique pair of species—if two studies crossed the same two species (regardless of subspecies or population status) the species pair only provides a single datum. For example, a study not included in the subset analyzed in the main text crossed *Drosophila simulans* to multiple *D. melanogaster* populations, but only provided a single datapoint for panels **B** - **D**.

Appendix D

Appendix for Chapter 5

D.1.1 Supplementary methods

Details of crossing protocol and fish husbandry

Female stickleback were selected for spawning when their abdomens were sharply angled at the cloaca and the first egg was visible. I gently squeezed the sides of the female fish's body to release the eggs into a Petri dish containing water from her source habitat (tank, lake, or river water). Mature male stickleback were identified by their bright blue colouration and red throat. Males were euthanized with an overdose of MS-222, and then testes were extracted from the body cavity using fine forceps after making a small incision beginning at the cloaca. I used a small paintbrush to release sperm from testes and to ensure that sperm contacted all eggs. The live fish and fertilized clutches were transported to the InSEAS aquatic facility at the University of British Columbia, Vancouver, British Columbia, Canada.

All fish were hatched in 100 L aquaria with room temperature between 17 and 19 °C and a photoperiod that followed local dawn and dusk times. Instant Ocean® Sea Salt was added to maintain a salinity of 5 ppt in all tanks. Fry were fed live brine shrimp nauplii. Chopped frozen bloodworms were added to the diet when fish were large enough, and then finally adult-sized fish were fed full size frozen bloodworms and frozen mysis shrimp *ad libitum* (Hikari Bio-Pure®).

I sampled fish for phenotype measurements typically when the mean standard length of a family was approximately 40 mm. Sticklebacks have adult morphology at this stage and are not sexually reproducing. Due to occasional logistical constraints, some tanks were sampled at earlier or later mean standard length sizes. Also, due to logistical constraint, all populations except for Paxton benthic and Paxton limnetic were collected from lakes and rivers in 2017 (see Fig. 5.1).

Repeatability

I evaluated the repeatability of my measurements to determine the fraction of variance that could be attributable to measurement error. Repeat measurements were made on at least 25 fish. For linear measurements (including pectoral fin length), I took two separate photographs of each fish (or fin) and made the repeated measurements on these separate photographs. Second photographs were made after returning and then removing the fish (or fin) from its storage vial. Count and gill raker measurements were made on the original specimens. In all cases except pectoral fin length, first and second measurements were made more than one year apart.

Data diagnostics

I checked for outliers in the raw data and evaluated outlier individuals to ensure they were not caused by measurement or transcription error. If fish were inadvertently measured twice, I averaged trait values across measurements. Fish with broken second dorsal spines were removed from the dataset. One fish was removed because it had an unusual body shape—qualitatively appearing as if it had failed to inflate its swim bladder—and it was an extreme outlier in Normal Q-Q plots and in standardized residuals vs. leverage plots. Such phenotypes seem to be caused by environmental factors (e.g., a too-powerful air stone) rather than biological factors (e.g., hybrid incompatibilities).

D.2 Supplementary Tables & Figures

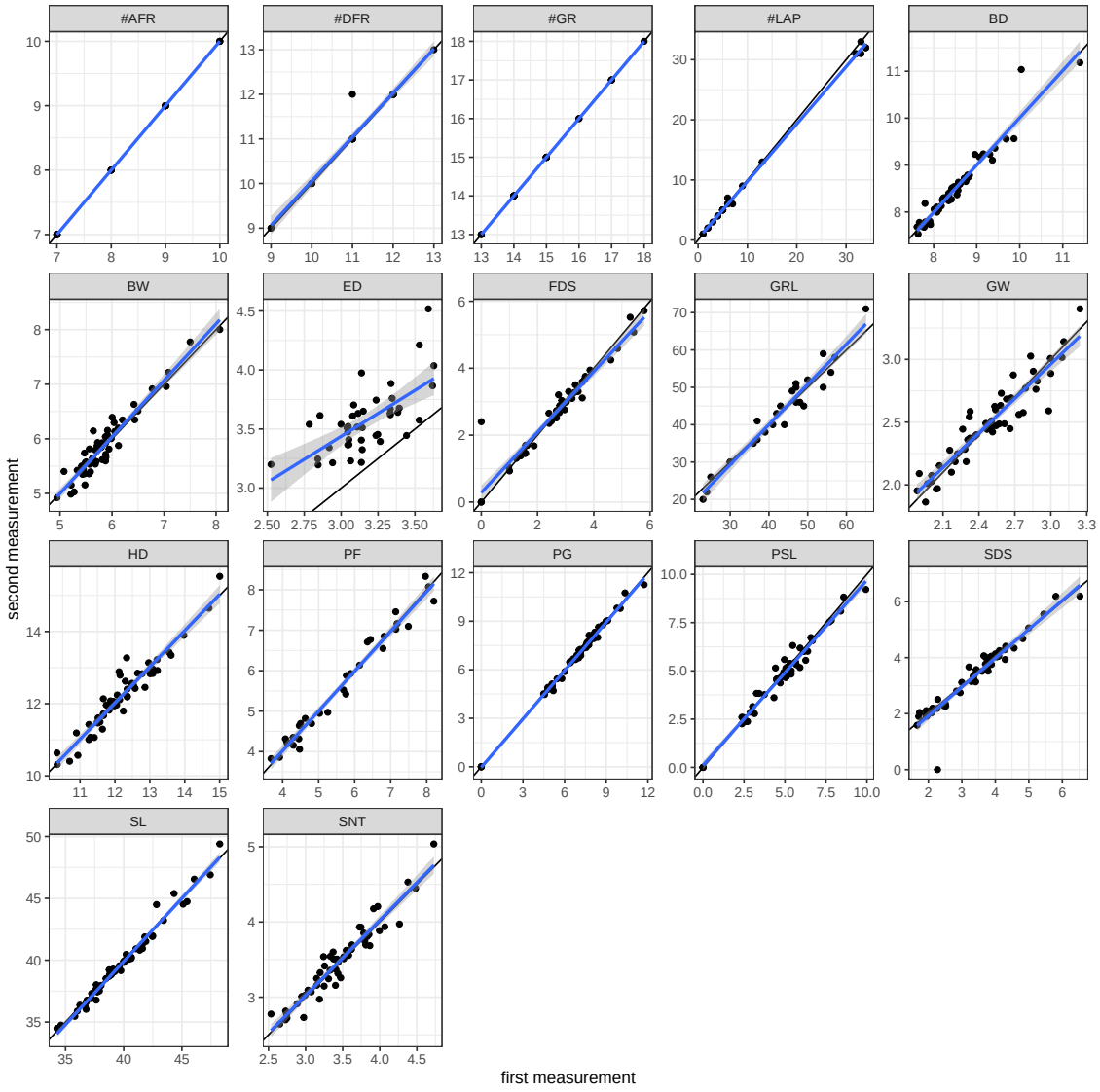


Figure D.1: **Repeatability data for all measured traits.** All plots show the first and second measurements made on all traits. Black lines are 1:1 lines, and blue lines are linear regressions. Trait codes are as in Fig. 5.1. The fish with a value of '0' for second dorsal spine (SDS) likely had its spine broken off during the time-frame between first and second measurements. All $r_{\text{Pearson}} > 0.9$, except for eye diameter (ED), which was dropped from the analysis.

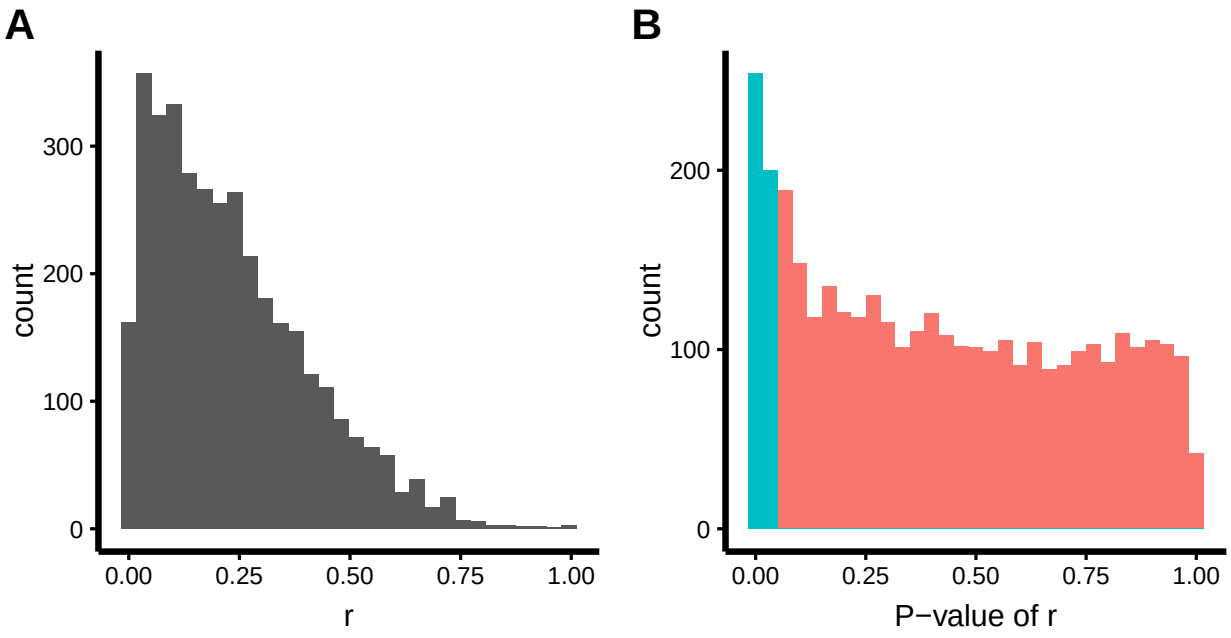


Figure D.2: **Summary of pairwise trait correlations in F_2 hybrids.** Correlations were run within each F_2 family where at least 10 individuals were measured. Panel (A) shows the distribution of Pearson's correlation coefficients (i.e. r) across the dataset, and panel (B) shows the distributions of P-values with blue indicating values of $P < 0.05$.

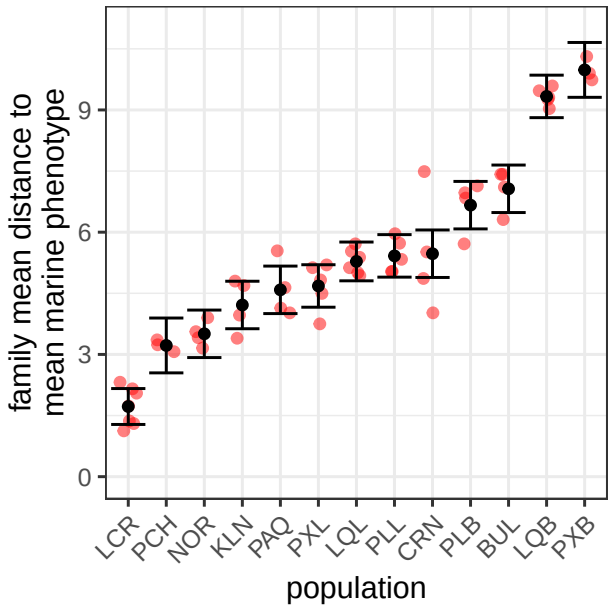


Figure D.3: **Freshwater divergence from the anadromous ancestor.** Boxplots show the distribution of partial residuals of individual phenotypic distance from the anadromous ancestor after accounting for the effect of 'family'. Population codes: PCH—Pachena Lake; PAX—Paxton Lake; CRN—Cranby Lake; PST—Priest Lake; LQU—Little Quarry Lake; PAQ—Paq (Lily) Lake; NOR—North Lake; KLN—Klein Lake; BUL—Bullock Lake.

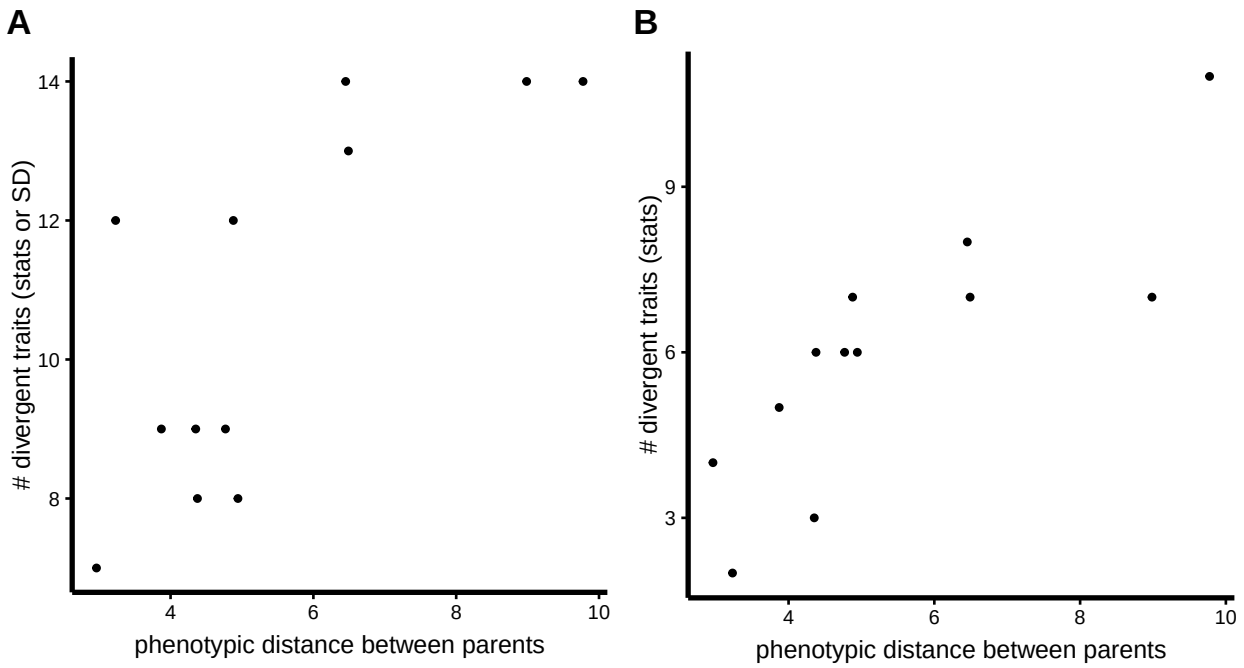


Figure D.4: **Phenotypic divergence between parents is positively associated with the number of traits that differ between them, as well as the number of possible pairwise mismatches.** There is one datum per freshwater population. Panels (A) and (B) show the number of traits that differ between the populations using both statistical and variance-based filtering ([A]; see Methods), or only statistics (B).

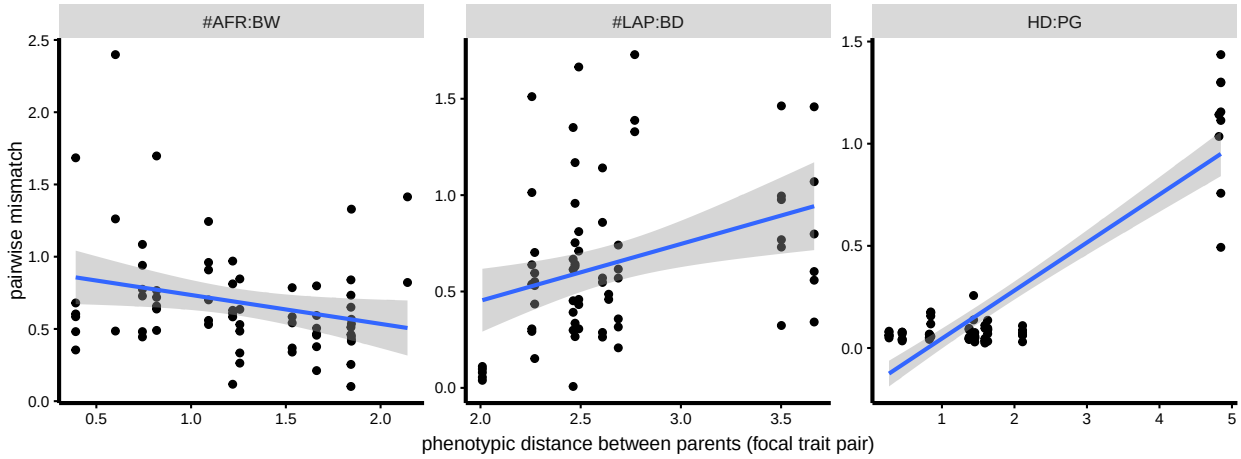


Figure D.5: **Examples of pairwise divergence:mismatch regressions.** The three plots show different regressions that generated coefficients in Fig. 5.2D. Trait labels are as in Fig. 5.1 (#AFR—number of anal fin rays; BW—body width; #LAP—number of lateral armour plates; BD—body depth; HD—head length; PG—pelvic girdle length). In the head length:pelvic girdle length regression, patterns are largely driven by the two girdle-less populations expressing a largely-marine phenotype for this trait but a largely freshwater head length.

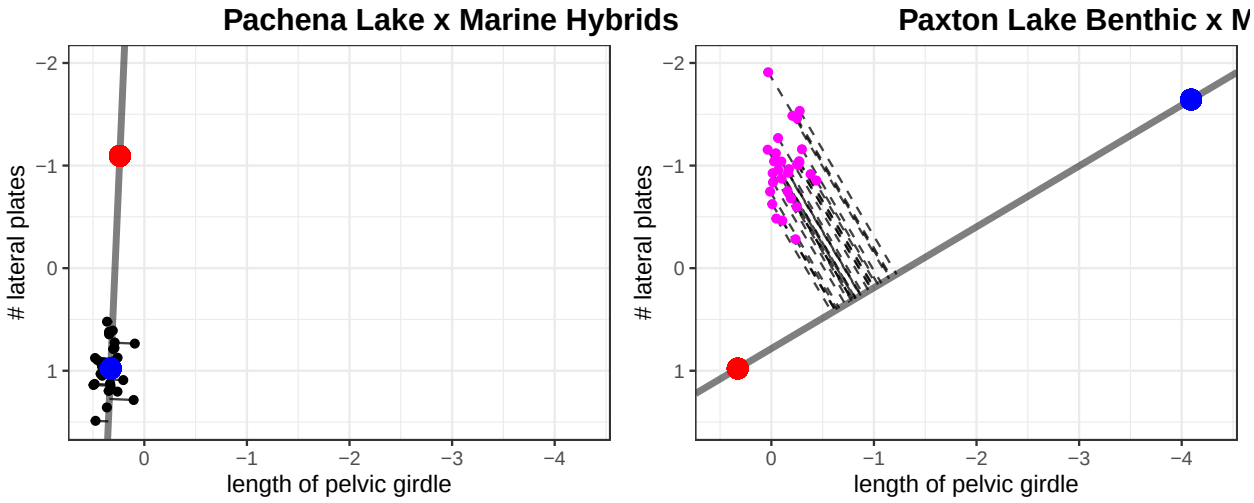


Figure D.6: **Visualization of pairwise mismatch in empirical data.** Each plot shows the scaled phenotype data used in all analyses. The red points indicate the freshwater parent, and the blue points indicate the anadromous parent. Black points are individual F_1 hybrids. Each hybrid's phenotype is connected to the line connecting parents by a perpendicular line—the length of this line is mismatch. For Pachena Lake F_1 hybrids, mean mismatch is 0.087. For Paxton Benthic F_1 hybrids, mean mismatch is 1.54. In this case, the high mismatch of the Paxton Benthics is due to the fact that the pelvic girdle phenotype resembles the anadromous ancestor whereas the plate number of biased toward the freshwater parent. The Pachena population is among the least divergent from the anadromous ancestor overall, whereas the Paxton Benthic population is the most.

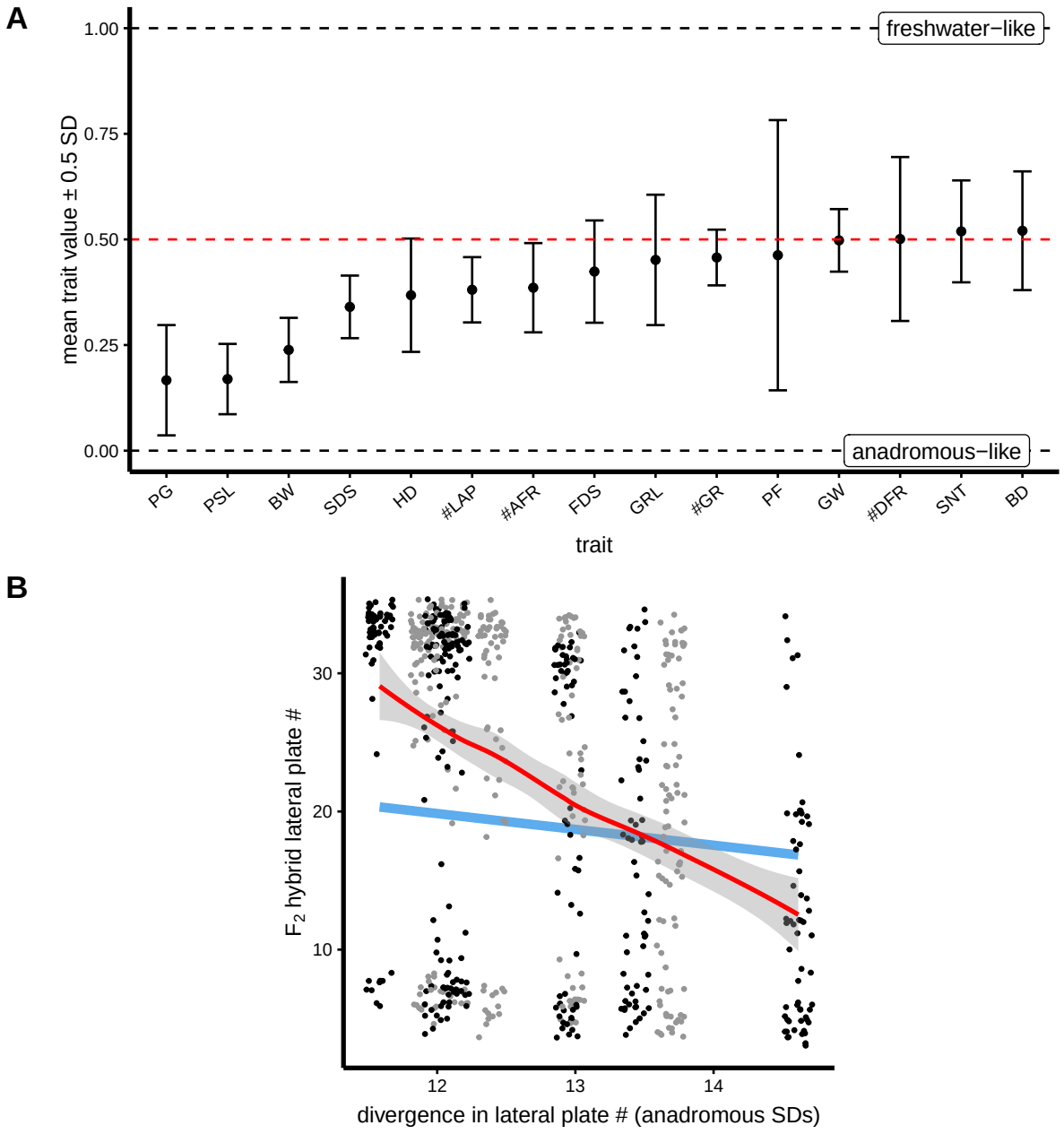


Figure D.7: **Dominance patterns in F_2 hybrids.** Patterns are largely similar to F_1s with a few notable exceptions. First, pelvic traits tend to be smaller than parents. Otherwise, trait values tend to be more intermediate among populations and dominance tends to be more consistent among populations. For the number of lateral plates, the general trend of increasing freshwater-dominance with divergence is apparent in F_2s, though they are much more variable than F_1s. Low values on the horizontal axis indicate that the parent population is less derived and larger values indicate that it is more derived. The red line is a loess-smooth fit to the data, and the blue line shows the parental midpoint.

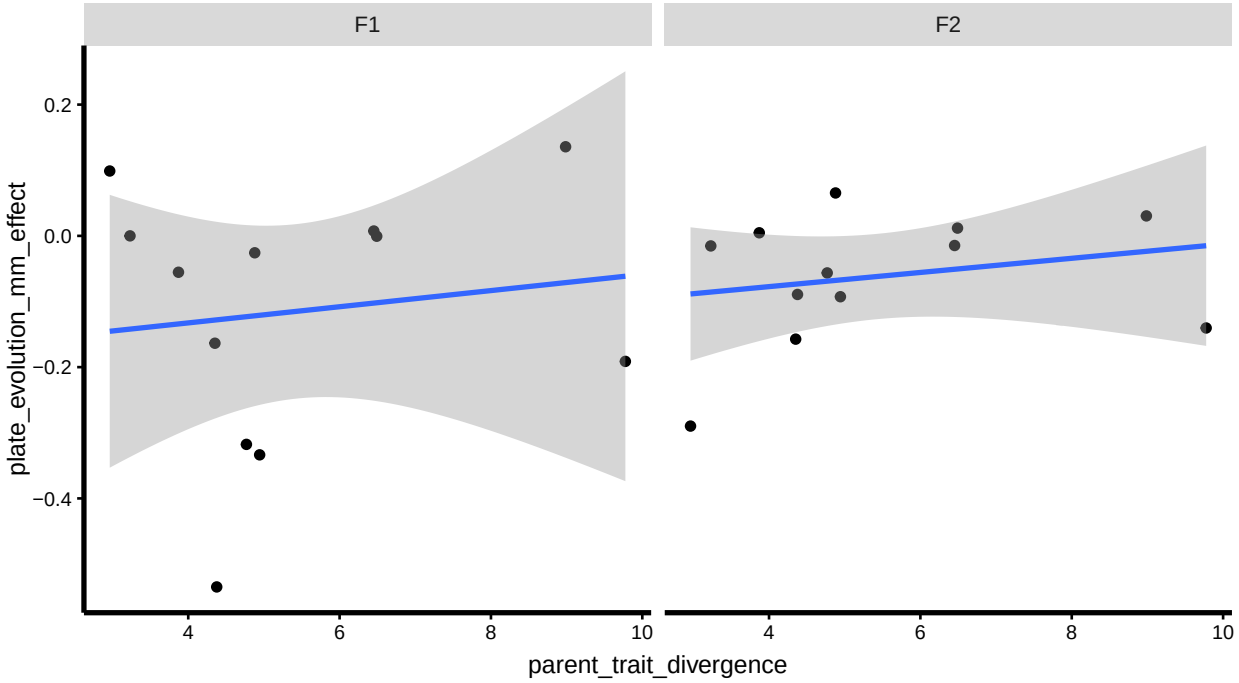


Figure D.8: **Variation in dominance for lateral plate count does not impact the mismatch-divergence relationship.** Both plots show the difference between mismatch if dominance were intermediate between parents minus the mismatch of the median hybrid—on this scale positive values indicate that dominance of this trait increases mismatch. Regressions for the F_1 and F_2 are both non-significant.

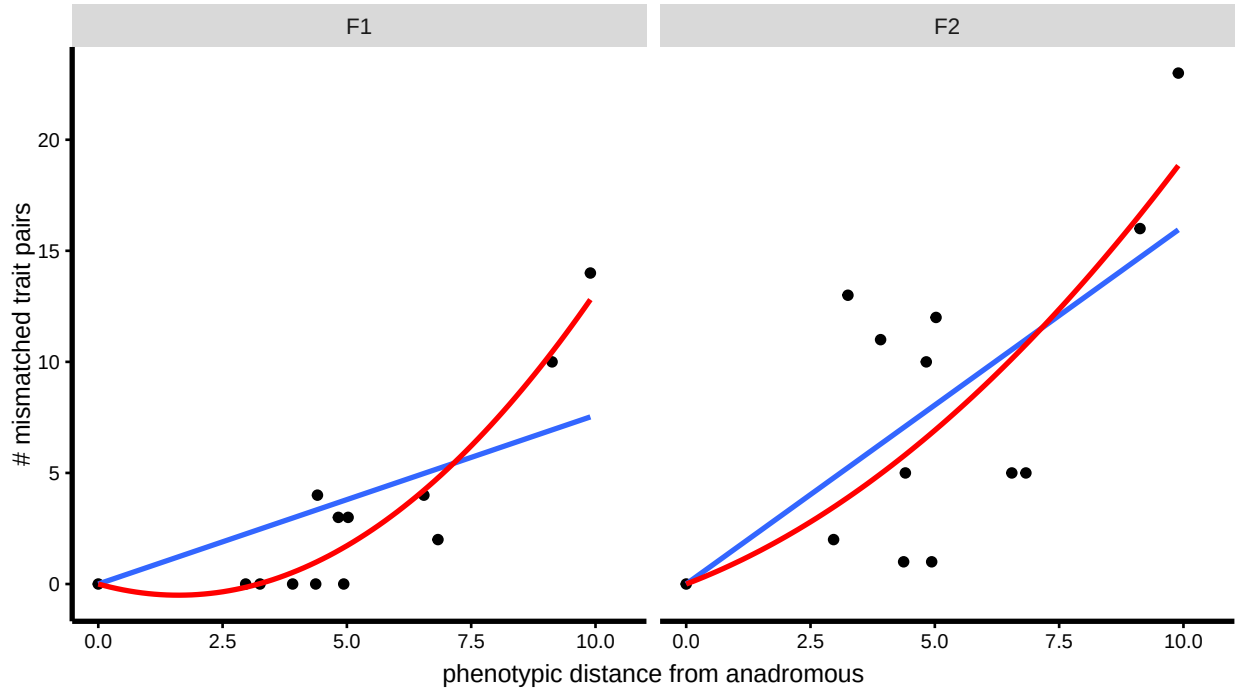


Figure D.9: **The number of mismatched trait pairs 'snowballs' with the magnitude of phenotypic divergence between parents, but only in F_1 hybrids.** The y-axis shows the number of trait pairs with significant mismatch, determined using *t*-tests which tested the null hypothesis that the difference between hybrid mismatch (with weighted pooling by families) and the pooled 'mismatch' across pure freshwater populations was 0. *P*-values were Bonferroni-corrected. The plot and regression lines are modelled after the 'snowball' studies of Moyle and Nakazato (2010) and Matute et al. (2010). The blue lines are linear regressions and the red lines are quadratics. Results hold if the intercept is not forced through zero, and if the 'origin' datum is omitted.

Appendix E

Appendix for Chapter 6

E.1 Supplementary methods

E.1.1 Experimental animals

Wild fish were collected from Priest and Paxton Lakes (Texada Island, BC, Canada) and Little Quarry Lake (Nelson Island, BC) from 2017–2019. Two Paxton benthic males were collected from a pond population on UBC campus founded with wild Paxton Lake benthic fish in 2016. All fish in the experiment were therefore 2–5 generations removed from the wild. All benthics were captured using minnow traps as were most limnetic males. Gravid limnetic females were caught almost invariably by dip-netting. One Paxton limnetic family and one Little Quarry limnetic family was raised from a nest-collected clutch. In these cases I closely examined resulting fish to ensure none were benthic $\times$ limnetic hybrids.

Wild fish were crossed either at the lakeside or at the lab by gently squeezing the eggs from a gravid female fish into a small Petri dish filled with lake or aquarium water. Male parents were euthanized with an overdose of MS-222, and their testes were removed and placed in the Petri dish. A fine paintbrush was then used to release sperm from the testes and ensure it was well mixed among the eggs in the clutch. Crosses for the present experiment were all made in the lab in much the same way. Males were occasionally used to fertilize multiple clutches. Due to logistical constraints, I made crosses as females became gravid—this led fish from different treatments to have different mean ages. In total, fish from 124 crosses were used in this study.

Crosses were made from 4 March to 9 April, 2020, and raised in 5 ppt saltwater (Instant Ocean) 110 L aquaria with a small amount of methylene blue added as a fungicide. These two additions reduce the loss of clutches to fungus and also reduce labour required to raise healthy fish (because *Artemia* nauplii can live for a couple of days in 5 ppt saltwater). Fish were fed *Artemia* nauplii daily. Due to logistical constraints I could not monitor hatching success but I note that I observed very little mortality. In general, there is very limited evidence that F_1 or F_2 hybrid stickleback exhibit 'intrinsic' deficiencies (Lackey and Boughman, 2017).

After fish were large enough to handle without risk of mortality (approx. 3–4 weeks), I split large families into multiple tanks and culled excess fish. At this time I began feeding fish chopped frozen bloodworms and conducting weekly 50 % water changes with fresh dechlorinated water. After splitting and culling I kept the density of fish to fewer than 30 individuals per aquarium. When fish were large enough, I added unchopped bloodworms, spirulina adult brine shrimp, and chopped mysis shrimp to their diets. Eventually, the mysis shrimp were fed whole.

E.1.2 Coded wire tagging

Here I fully document my methodology for using sequential coded wire tags (CWTs). I ordered sufficient quantities of CWTs from Northwest Marine Technology (NMT, https://www.nmt.us/; Anacortes, WA, USA). The CWTs come on sheets with two columns—one 'fish' column and one 'reference' column. Although CWTs appear in sequential order, the numbers do not increase in perfect 1:1 association with tag position. Because of this, the unique sequences on CWTs cannot be inferred without some possibility of error. It is therefore neccessary to pre-read some tags. I never used adjacent (i.e., on the same row) tags for fish bound for the same pond. This allowed me to pre-read fewer tags without risking error. Specifically, I pre-read every third tag in the 'reference' column using a Magniviewer (NMT) and could reliably infer other tags because I know which pond each fish was retrieved from.

Before tagging, fish were anaesthetised with MS-222. I then used a single-shot CWT injector (Northwest Marine Technology) to inject CWTs beneath the skin on the fish's dorsal musculature. Fish were injected while laying on their side atop a large sponge, and their head was covered with a paper towel soaked in water from their original tank. Light pressure was applied to the head and caudal peduncle to stabilize the fish during injection. Injection was easiest if the lateral plates were used as a leverage point to implant the head of the needle beneath the skin. (Limnetics, with their denser musculature and invariable presence of lateral plates, were easier to inject). I found maximal success when the push rod was not pushed to be maximally extended—in fact this can cause the CWT to emerge from the fish, rather an extension of about ¾ worked best. When injecting, I took care that the pectoral fin was either oriented toward the head or down toward the ventral area, so as to not be pierced by the tagging needle. Fish were kept temporarily in aerated water from their original tank for recovery and then moved back to their original tank before introduction into the ponds three days later. Methylene blue was added to their original tank to prevent infection of the tagging wound. I estimate that each fish took approximately 30 seconds to anaesthetize, weigh, and tag, and it took about 10 minutes to pre-read a page of CWTs.

At the end of the experiment I retrieved and read tags from each fish. I located tags under a dissecting microscope. Tags were always handled delicately because metal forceps can easily damage them. Most often, a scalpel was used to cut away skin and visually identify the tag. After cutting away muscle, the magnetic tag is attracted to the scalpel, where it can be gently placed onto a magnetic pencil. Removed tags were read with a MagniViewer (NMT) and then stored for later reference if necessary.

Ambiguities and possible transcription errors were identified by ensuring a match between species (benthic or limnetic) and pond, checking for duplicates and implausible values, and ensuring each recovered tag was assigned to a fish.

E.1.3 Estimating fitness via fecundity and overwinter survival

In the main text I use a conservative fitness estimate that incorporates only survival and growth. A more complicated and possibly more accurate estimate of fitness would account for overwinter survival and fecundity. This estimate of relative fitness considers three components. The first, survival during the experiment (i.e., 'summer survival), was directly measured. The next two fitness components consider the fitness effects of body size (directly measured) on overwinter survival and fecundity (both estimated). Absolute fitness was

calculated as summer survival × (estimated) winter survival × (estimated) fecundity. Relative fitness was calculated as the absolute fitness of each cross type divided by the cross type with the highest absolute fitness (within species).

I estimated relative overwinter survival and fecundity using previously published data. These previous studies used standard length to predict fitness components, and I measured approximately 50 individuals of both species to estimate the mass–length relationship for both. Quadratic models had a high explanatory pattern for both species ($r^2_{\text{limnetic}} = 0.88$; $r^2_{\text{benthic}} = 0.79$) (Fig. E.4).

To estimate relative overwinter survival, I conservatively interpret the analysis of Carlson et al. (2010), who found that standard length was positively associated with overwinter survival in Alaskan stickleback. The specific relationship between length and survival was either quadratic (concave) or linear, depending on the year (Carlson et al., 2010). I conservatively assume that relative overwinter survival is a linear function of standard length. I estimated fecundity using two previously published datasets from stickleback experiments in the UBC Experimental Ponds (Fig. E.5). Schluter et al. (2021) estimated fecundity in F_2 marine × freshwater stickleback hybrids. Specifically, the fitness of over 200 F_2 hybrid females was quantified as the tally of her surviving F_3 offspring (from a sample of 500 F_3s). Male fitness was not estimated by Schluter et al. (2021), but the authors speculate that selection acted similarly on males and females because the evolutionary response observed was highly similar to what was expected from estimates of female fitness. Bay et al. (2017) estimated the number of mating events for Paxton Lake F_2 benthic × limnetic hybrid females, and pure (i.e., non-hybrid) benthic and limentic males. Analysis of the data suggested that length–mating relationships were the same across groups (i.e., no significant interaction) so all data were grouped for analysis. Data were analyzed using simple linear models from which intercepts and slopes were extracted. Relationships were significant (in a Poisson generalized linear model) and positive in both datasets, but I chose to base fecundity estimates on the estimates from Schluter et al. (2021) because differences in relative fitness among groups were smaller (i.e., it is more conservative).

E.2 Supplementary Tables & Figures

Table E.1: Summary of fish numbers and survival rates for the 2020 pond experiment

pond	origin	species	type	*n* released	*n* recaptured	survival proportion
4	Little Quarry	b	Parent	102	46	0.45
4	Paxton	b	Parent	107	70	0.65
4	hybrid	b	F_1	206	123	0.60
4	hybrid	b	F_2	210	130	0.62
4	Little Quarry	l	Parent	106	31	0.29
4	Paxton	l	Parent	107	39	0.36
4	hybrid	l	F_1	207	101	0.49
4	hybrid	l	F_2	208	62	0.30
9	Little Quarry	b	Parent	104	84	0.81
9	Priest	b	Parent	106	93	0.88
9	hybrid	b	F_1	203	185	0.91
9	hybrid	b	F_2	205	169	0.82
9	Little Quarry	l	Parent	105	36	0.34
9	Priest	l	Parent	105	57	0.54
9	hybrid	l	F_1	204	86	0.42
9	hybrid	l	F_2	205	100	0.49
19	Priest	b	Parent	103	95	0.92
19	Paxton	b	Parent	103	87	0.84
19	hybrid	b	F_1	204	174	0.85
19	hybrid	b	F_2	205	161	0.79
19	Priest	l	Parent	105	32	0.30
19	Paxton	l	Parent	105	42	0.40
19	hybrid	l	F_1	206	88	0.43
19	hybrid	l	F_2	205	72	0.35

Table E.2: Fitness components and relative fitness estimates for cross types within species.

species	cross type	summer surv.	mass (g)	std. lgth. (cm)*	rel. winter surv.*	fecundity*	rel. fit.†
limnetic	F_1	0.454	1.09	4.59	1.000	1.74	1.000
	F_2	0.386	0.98	4.40	0.959	1.45	0.678
	pure	0.387	1.01	4.46	0.971	1.54	0.731
benthic	F_1	0.818	2.14	5.91	1.000	3.80	1.000
	F_2	0.797	2.01	5.73	0.970	3.53	0.877
	pure	0.799	2.06	5.81	0.982	3.64	0.918

*estimate; †w = summer surv $\times$ rel. winter surv. $\times$ fecundity

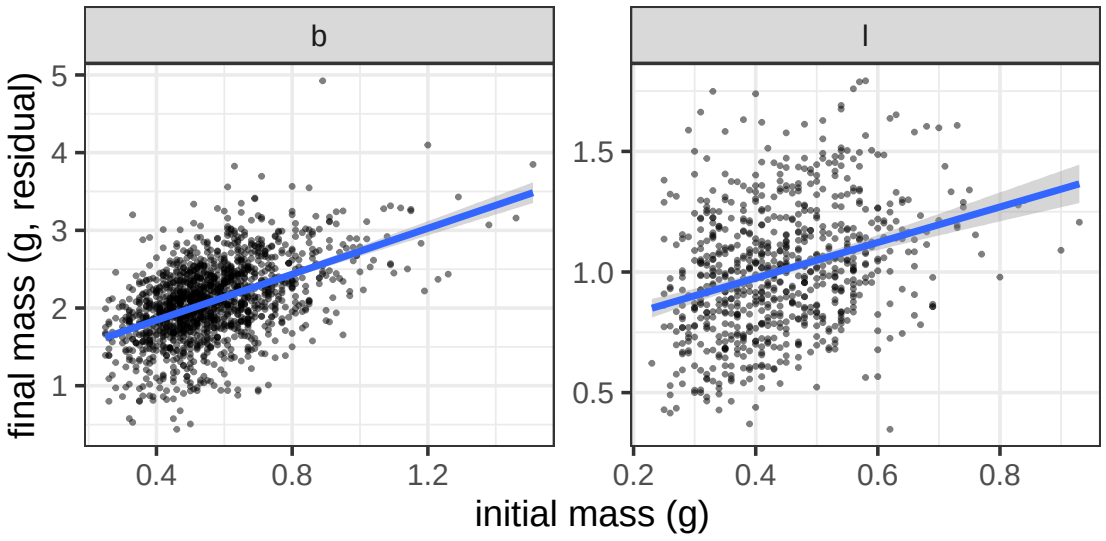


Figure E.1: **Relationship between initial and final mass in the 2020 pond experiment.** The left plot shows benthics and the right shows limnetics. Points are partial residuals from a mixed model with pond as a random effect.

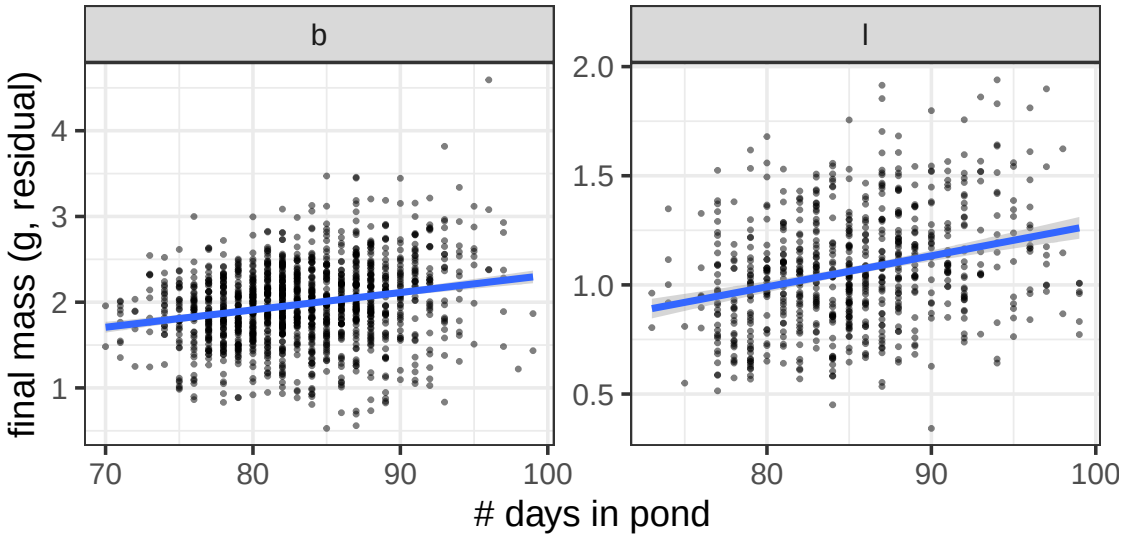


Figure E.2: **Relationship between the number of days in pond and final mass in the 2020 pond experiment.** Both relationships are significant and positive.

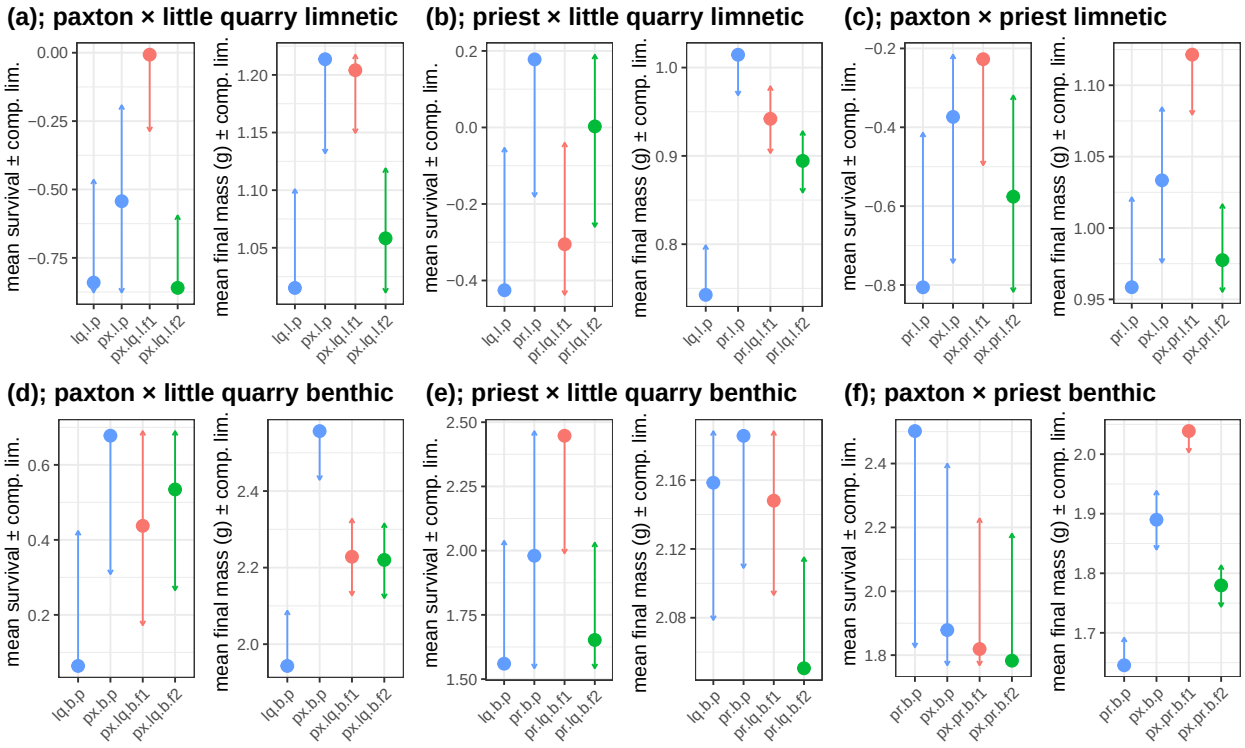


Figure E.3: **Fitness components with the two parent populations plotted separately.** Points are estimated marginal means and arrows are comparison limits. Note the differences in scale among all plots.

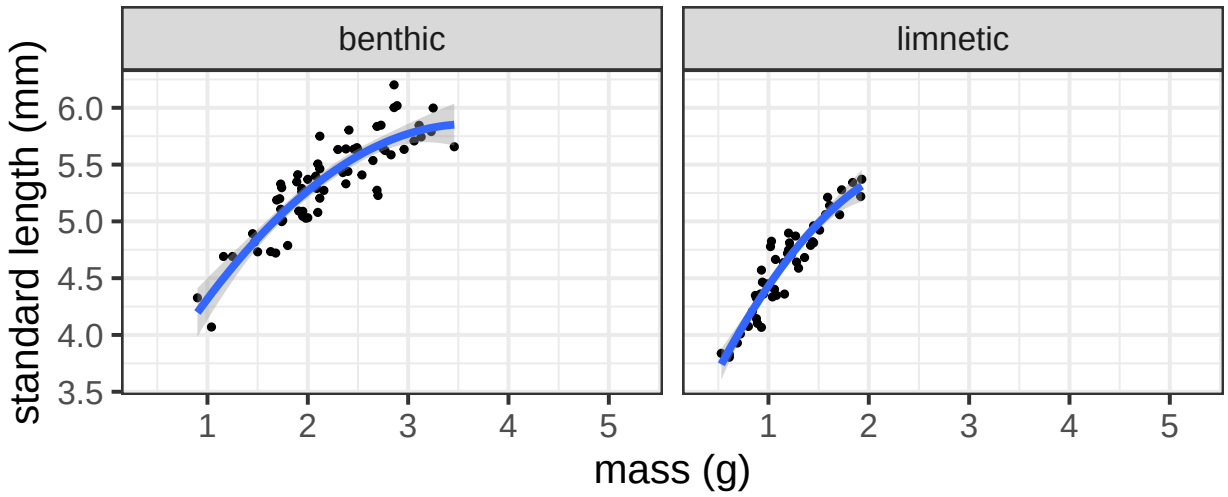


Figure E.4: **Relationship between mass and standard length for fish collected from ponds.** Relationships are shown separately for benthics and limnetics. Both are highly significant and positive. These data were used to estimate standard length across the dataset.

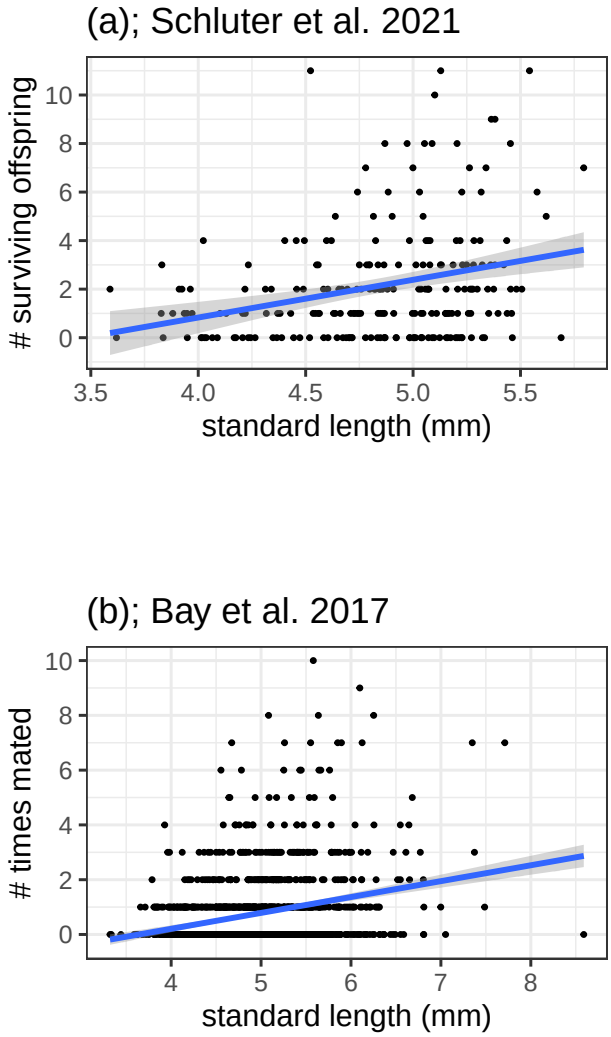


Figure E.5: **Relationship between standard length and fecundity in previously published stickleback pond experiments.** Schluter et al. (2021) considered F_2 marine-freshwater hybrid females and measured the number of F_3 hybrids to which they could be confidently assigned parentage. Bay et al. (2017) considered F_2 benthic-limnetic (Paxton) hybrid females and non-hybrid males of both species and recorded the number of successful mating events (inferred from genotyped eggs and offspring). A generalized linear model did not reject the hypothesis that the slope differed by group (F_2 female, pure benthic male, or pure limnetic male), so I plot them together. Both relationships are significant and positive as evaluated with a generalized linear model.

Appendix F

Appendix for Chapter 7

F.1 Supplementary Figures

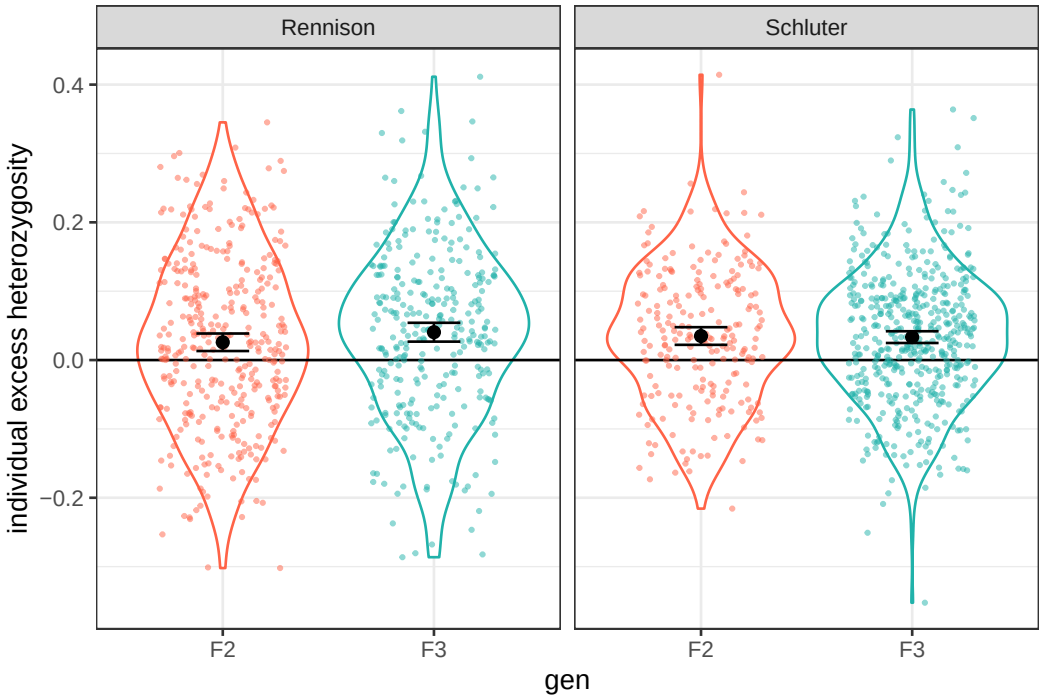


Figure F.1: **Heterozygosity does not differ between F_2 and F_3 hybrids.** The plots show individual excess heterozygosity from the two studies that genotyped both the F_2 and F_3 generations (Schluter et al., 2021; Rennison et al., 2019). The means do not differ between generations in either study.

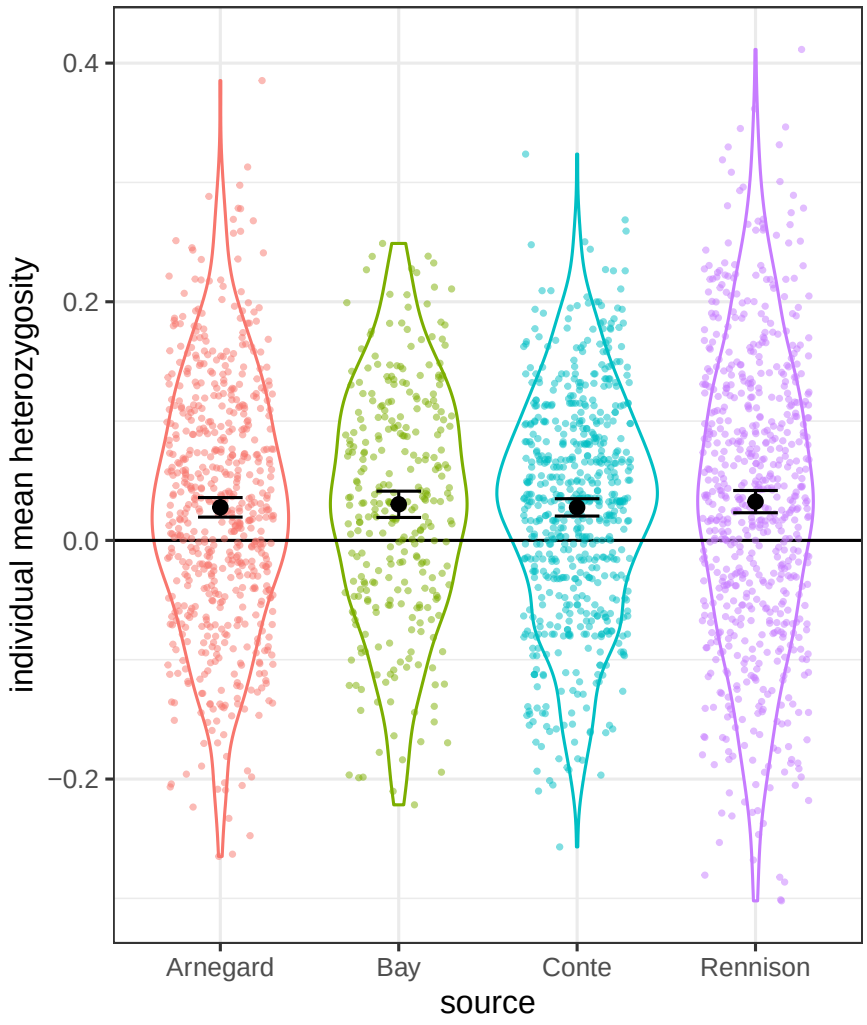


Figure F.2: **Heterozygosity does not differ between studies involving Paxton Lake benthic × limnetic hybrids.** The means of all three studies are statistically indistinguishable.

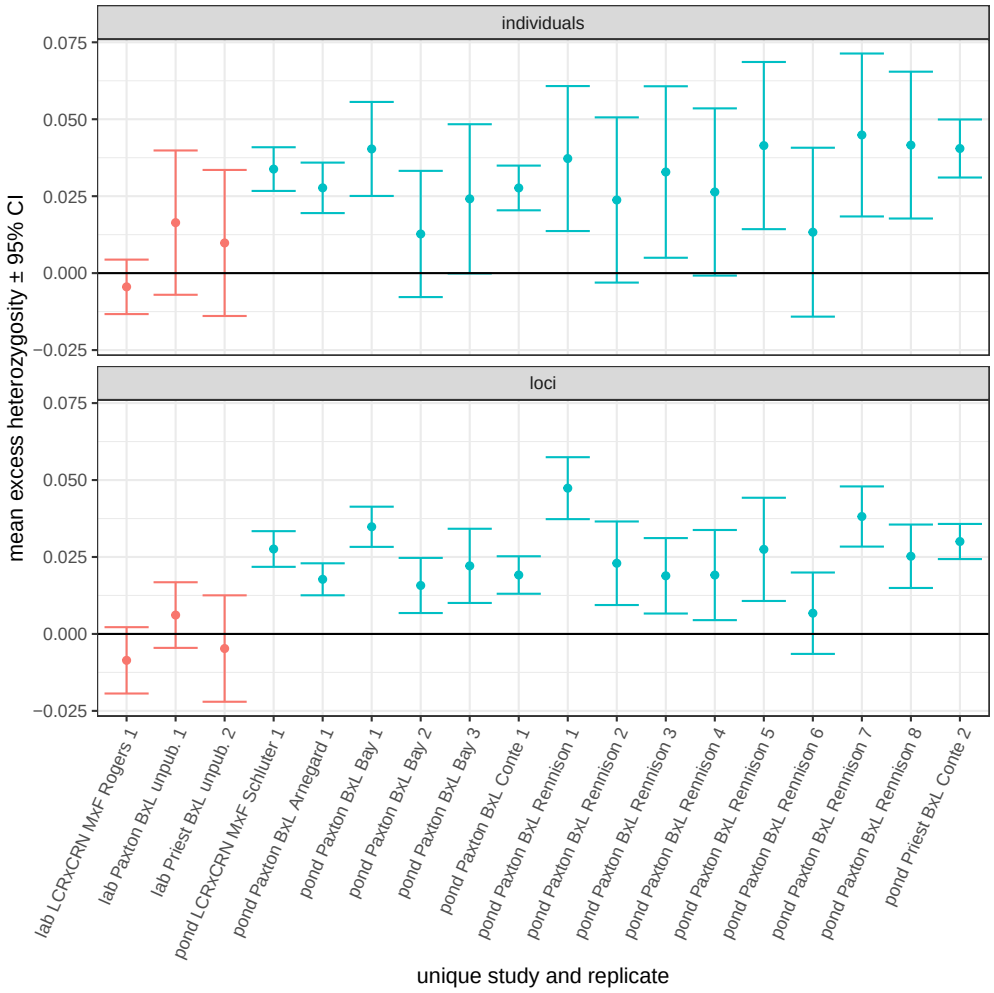
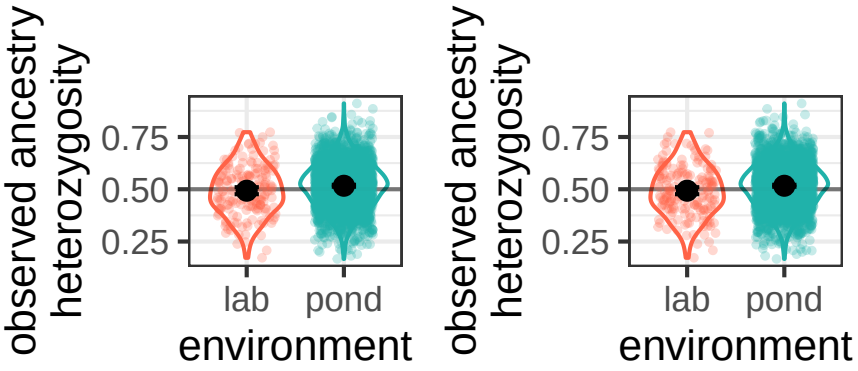


Figure F.3: **Estimates of excess heterozygosity for individuals and loci across 'replicates'.** We consider a replicate to be a unique bi-parental cross for aquarium studies, and a unique pond for pond studies. Mean excess heterozygosity is shown for each such replicate for both individuals (upper) and loci (lower). In each panel the horizontal line indicates no excess heterozygosity. Red points are 'lab' replicates, and blue points are 'pond' replicates.


observed ancestry heterozygosity
0.75
0.50
0.25
lab
pond
environment
observed ancestry heterozygosity
0.75
0.50
0.25
lab
pond
environment

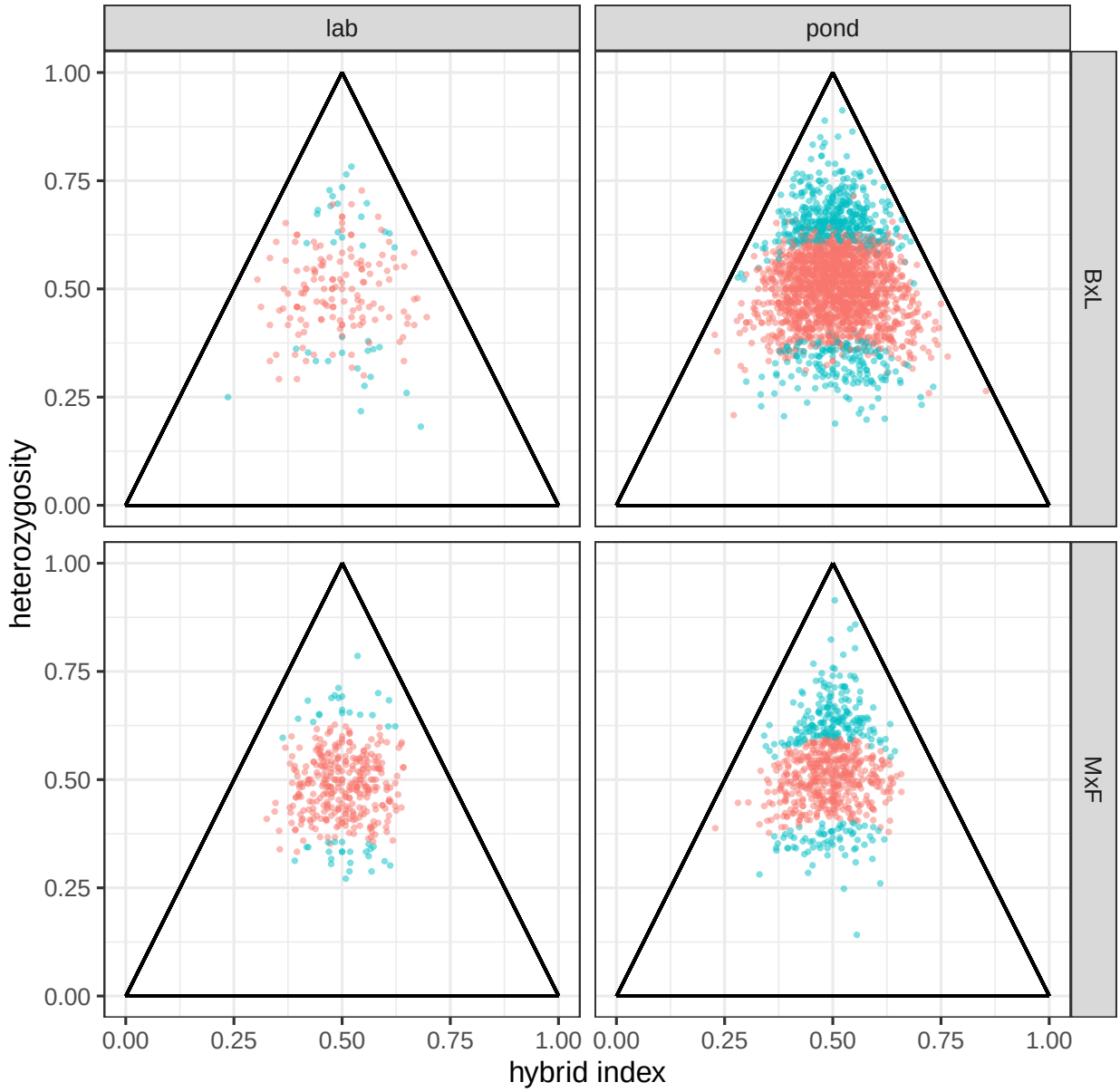


Figure F.5: **de Finetti ternary diagrams for genotyped *individuals* in each of the bi-parental crosses.** Each point represents an individual F_2 hybrid shows each individual's hybrid index (frequency of benthic alleles in its genome) and its mean heterozygosity. Loci with significant deviations from Hardy-Weinberg equilibrium (χ^2-test) are shown in blue.

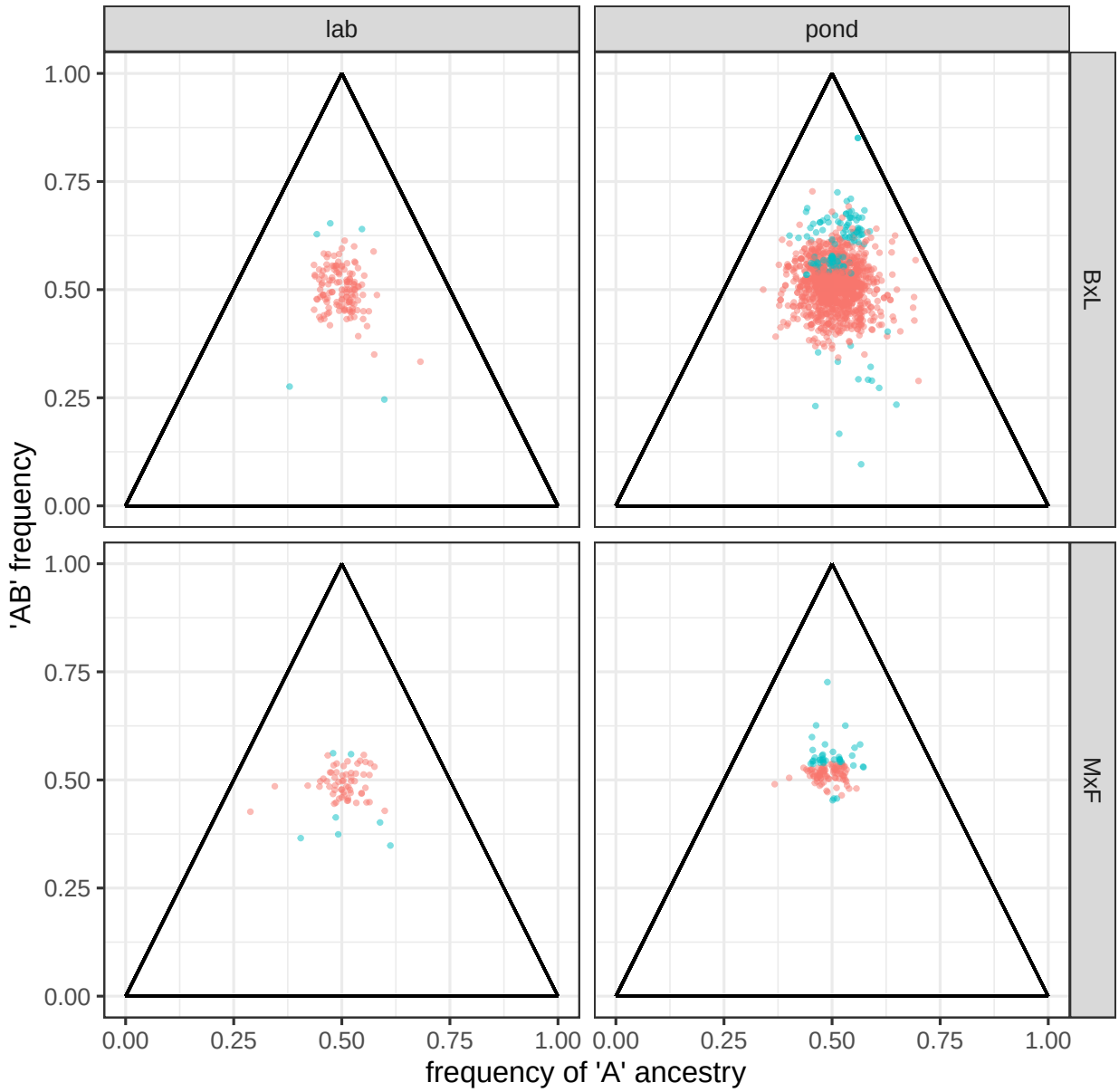


Figure F.6: **de Finetti ternary diagrams for genotyped *loci* in each of the bi-parental crosses.** Each plot shows relative frequency of benthic and limnetic alleles along the *x*-axis and of heterozygotes on the *y*-axis. Loci with significant deviations from Hardy-Weinberg equilibrium (χ^2-test) are shown in blue.

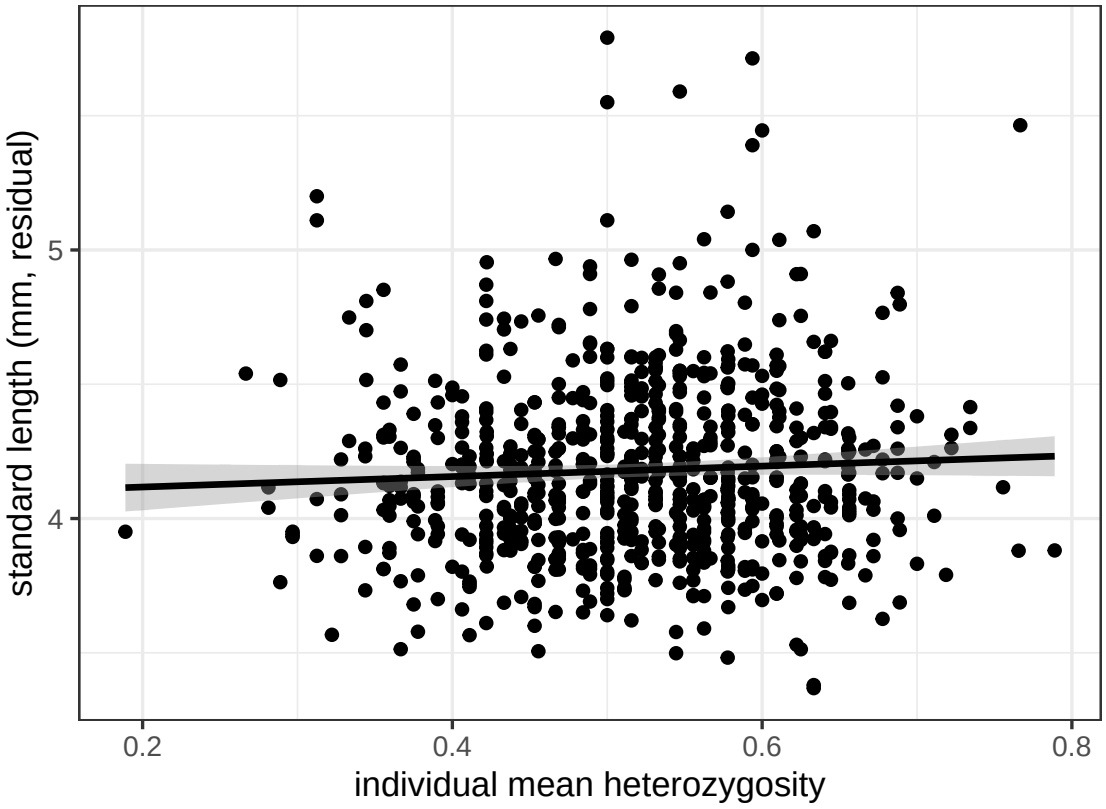


Figure F.7: **No relationship between individual mean heterozygosity and growth (standard length) in the aquarium-raised biparental benthic-limnetic F_2 hybrids.** Results are residuals from `visreg` (Breheny and Burchett, 2017). Each point is an individual F_2 hybrid. The interaction between lake-of-origin mean heterozygosity was non-significant so I plot the main effect across both lakes-of-origin (Paxton and Priest). Mean heterozygosity was not significantly associated with standard length (β = 0.46 $\pm$ 2.75 [SE], $F_{1,175}$ = 0.028, P = 0.86).

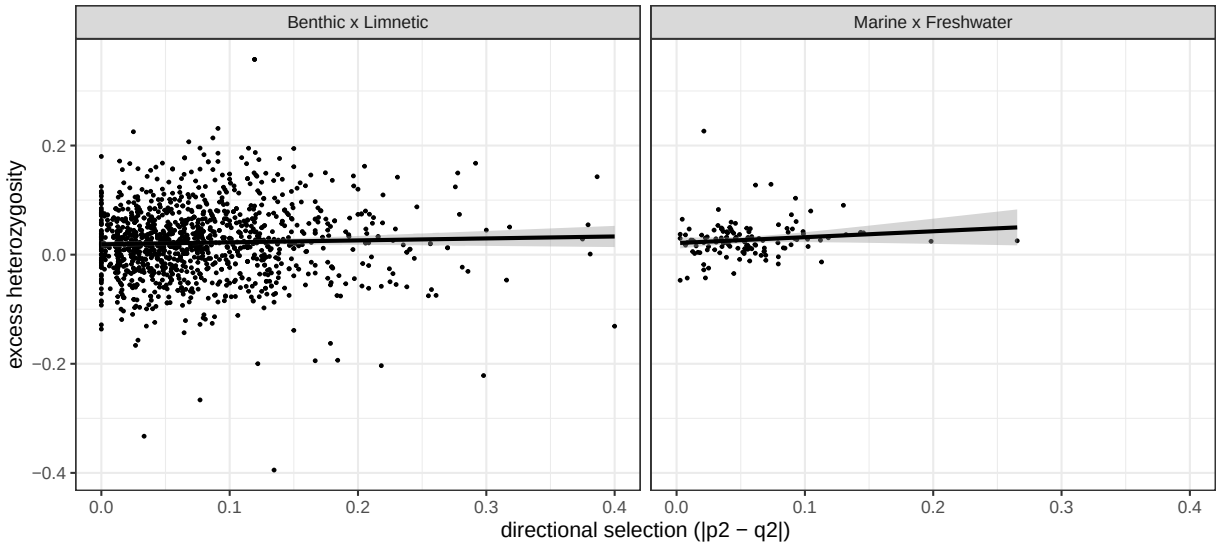


Figure F.8: **No relationship between estimated directional selection at each locus and its mean excess heterozygosity in ponds.** Results are residuals from `visreg` (Breheny and Burchett, 2017). Each point is an individual locus genotyped in F_2 hybrids in ponds. The left panel shows benthic-limnetic hybrids and the right panel shows marine-freshwater hybrids. Directional selection was calculated as the absolute difference in the frequency of both homozygotes. Both relationships are non-significant.