ALIGNMENT OF GENETIC VARIATION, PLASTICITY, AND SELECTION, AND THE EFFECTS OF COST OF PLASTICITY

A Dissertation
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Monica Anderson Berdal

In Partial Fulfillment of the Requirements
for the Degree of
DOCTOR OF PHILOSOPHY

Major Department:
Biological Sciences

March 2021

Fargo, North Dakota

North Dakota State University
Graduate School

Title

ALIGNMENT OF GENETIC VARIATION, PLASTICITY, AND SELECTION, AND THE EFFECTS OF COST OF PLASTICITY

By

Monica Anderson Berdal

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

DOCTOR OF PHILOSOPHY

SUPERVISORY COMMITTEE:

Ned Dochtermann
Chair

Julia Bowsher

Steven Travers

Lauren Hanna

Approved:

4/12/2021
Date

Kendra Greenlee
Department Chair

ABSTRACT

Phenotypic expression depends on both the underlying genetics and the environment the phenotype is expressed in, i.e., plasticity. Adaptive theory predicts that selection should align with the dimensions of most genetic variation and plasticity because this will increase the evolutionary rate of a population, meaning that a population would reach its fitness optimum faster than if they were misaligned. Alignment with selection is only predicted if there is directional selection, and not under stabilizing selection. In addition, only adaptive plasticity is predicted to align with both selection and genetic variation, with the proportion of the plastic variation consisting of adaptive plasticity determining how well aligned plasticity should be. In the first chapter of this dissertation, I outline the evolutionary consequences of the relationship between selection, genetic variation, and plasticity, as well as what the predictions are for their alignments and how to estimate them. In my second chapter I empirically test the alignment between selection, among- and within-individual variation (used as proxies for genetic variation and plasticity respectively) for three behaviors in a wild population of deer mice (*Peromyscus maniculatus*). I found that selection, among- and within-individual variation were all misaligned, and that there was very little variation in all three behaviors. This could indicate that the behaviors have already reached their fitness optimum due to previous selection pressure. Consequently, this population might not be able to adapt to environmental change. In my last chapter I investigate the cost of plasticity in response to a predatory cue on reproductive outputs in isogenic lines of the banded cricket (*Gryllodes sigillatus)*. Plasticity is assumed to have associated costs which would affect its alignment with selection and genetic variation. I found no evidence for cost of plasticity in *G. sigillatus*, and in addition there was no genetic variation in plasticity among the lines. Again, previous selection might drive the population's mean plasticity

to its fitness optimum, reducing the variation and the costs of plasticity, making it harder to detect.

ACKNOWLEDGMENTS

First and foremost, I want to thank my supervisor Ned Dochtermann for all his help and guidance throughout my PhD studies. Without his patience and support I would never have been able to finish my dissertation, and for that I am very grateful. I feel extremely lucky to have a supervisor who weights mental health heavier than a timely submitted PhD. I would also like to thank my committee, Julia Bowsher, Steven Travers, and Lauren Hanna, for their understanding and flexibility in letting me complete my dissertation abroad. Next, I would like to thank Raphaël Royauté for many helpful theory discussions, help with statistical analyses, as well as a good friendship full of many motivational speeches. I would also like to thank Mark Clark for his help with mark-recapture analysis. My thanks also go to my family for supporting my move to a different country to continue my studies in biology, and for all their help throughout this time. A major part of motivating me through all the ups and downs of this work was played by all my friends who never stopped supporting me, so a big thank you to Diana Eckert, Ane Negard, Thomas Haaland, Gunhild Engen, Beate Bjorli Dahle, Anne Nagelhus, Pamela Block, Maria Guixe, Aleix Valls, Kevin Cortes, Liz Cambron, Courtney Grula, Lauren Dennhardt, Dacotah Melicher, and Cody Hellquist. Lastly, I want to thank my boyfriend Julian Evans who has always been there for me. His encouragement and help with everything from coding to pep-talking has been a major factor in finishing this thesis.

TABLE OF CONTENTS

ABSTRACT iii

ACKNOWLEDGMENTS v

LIST OF TABLES ix

LIST OF FIGURES x

LIST OF ABBREVIATIONS xii

LIST OF SYMBOLS xiii

LIST OF APPENDIX TABLES xiv

LIST OF APPENDIX FIGURES xv

1. GENERAL INTRODUCTION 1

2. ADAPTIVE ALIGNMENT OF PLASTICITY WITH GENETIC VARIATION AND SELECTION 5

2.1. Abstract 5

2.2. Introduction 5

2.3. Testing for alignment 13

2.4. Expectations for the alignment of E 17

2.5. Conclusions 24

3. MISALIGNMENT OF SELECTION, PLASTICITY, AND AMONG-INDIVIDUAL VARIATION: A TEST OF THEORETICAL PREDICTIONS WITH *PEROMYSCUS MANICULATUS* 25

3.1. Abstract 25

3.2. Introduction 26

3.3. Methods 29

3.3.1. Study species 29

3.3.2. Trapping and tagging 29

3.3.3. Behavioural assays 30

3.3.4. Statistical analyses 32
3.4. Results 37
3.4.1. Assay validity 37
3.4.2. Average behavior 38
3.4.3. Behavioral variation 38
3.4.4. Correlation matrices 39
3.4.5. Mark-recapture results 40
3.4.6. Vector correlations and angles 41
3.5. Discussion 42
4. NO COST OR GENETIC VARIANCE FOR BEHAVIORAL PLASTICITY IN CRICKETS 49
4.1. Abstract 49
4.2. Introduction 50
4.3. Methods 55
4.3.1. Breeding 55
4.3.2. Rearing 56
4.3.3. Behavioral tests 56
4.3.4. Reproductive measurements 57
4.3.5. Statistical analysis 58
4.4. Results 60
4.4.1. Trait correlations within and between sexes 60
4.4.2. Among- and within-line variation 62
4.5. Discussion 65
REFERENCES 70
APPENDIX 77
A.1. Testing for historical and stabilizing selection 78

A.2. Mark-recapture model 81

A.3. Appendix references 85

LIST OF TABLES

Table — Page

2.1. Relevant evolutionary alignments and some of their biological significance. 12

2.2. Components contribution to phenotypic variation with example expected values and corresponding correlations 21

3.1. Among-individual correlations (above the diagonal) and within-individual correlations (below the diagonal) between activity, aggression, and predator cue response 40

3.2. ψ_θ and proportion of $\psi_\theta > 0$ for the alignment between among – and within-individual variation (i_{max} and w_{max} respectively) and selection gradient (β), and proportion of estimates (out of 1000) that are below three different null expectations (correlation of 0.975. 0.95, and 0.90) for the vector correlation between i_{max}, w_{max}, and β. Because the proportion of ψ_θ estimates > 0 was less than 0.95 none of the vectors were clearly misaligned.. 42

4.1. Model comparisons to investigate whether the among-line correlations between plasticity, mass, and fitness between the sexes were substantial. 62

4.2. Model comparisons to investigate whether the among- and within-line correlations between plasticity, mass, and fitness within the sexes were substantial. 62

4.3. Model comparison to investigate whether the among- and within-line variation was substantial. 63

LIST OF FIGURES

Figure — Page

2.1. A. Response of a population under stabilizing selection where the population's mean is at the optimum (θ). Over successive generations, if the optimum does not change, variation around the mean is lost. B. An example of a selection surface (ω) for which fitness increases as traits approach intermediate values for both traits, i.e. when trait combinations are along the adaptive ridge. C. If a population's multivariate mean is at an optimum, variation will be depleted over successive generations so as to mirror the selection surface (B)—note that under stabilizing selection the distribution of variation is visualized as being similar to the topography of the selection surface and that, as a result, most variation is present in the direction of weakest selection (i.e. under stabilizing selection G is orthogonal to ω). 7

2.2. Relationship among G, E, P, and β in two populations (A and B). In both populations the main axes of genetic and environmental variation (i.e. the dominant eigenvector of G and E) are closely aligned with a minimal angle (*r*°) between them. 11

2.3. A. Relationship between a vector correlation and the degree of the angle between two vectors in multivariate space. The shaded area represents a 95% confidence interval around the degree of the angle for constant uncertainty of the vector correlation. B. The breadth of the 95% CI (A) around the degrees of the angle relative to the magnitude of a vector correlation. Uncertainty around r° was held constant but because of the non-linear relationship (A), the uncertainty in the angle is greatest at high values of r°. 16

3.1. The mean and standard deviation for A) repeatability, B) among-individual variation, and C) within-individual variation for activity level (red, Act), aggression (blue, Agg), and predator cue response (yellow, PR). 39

3.2. Angles between A) i_{max} (dashed blue line) and w_{max} (dotted black line), posterior HPD-inteval = 29.32 – 90.00, B) i_{max} and β (solid red line), posterior HPD-inteval = 26.50 – 89.98, and C) w_{max} and β, posterior HPD-inteval = 20.32 – 89.75. 41

4.1. Differences in cost of phenotype and cost of plasticity A) G1 (blue line) and G2 (red line) are fixed genotypes and the expressed phenotype is the same regardless of the environment (absence vs. presence of a predator). G3 (black line) is a plastic genotype and the phenotypic expression changes according to the environment the genotype is expressed in. B) The fitness of G1 and G2 depends on which environment they are expressed in, where fitness will be higher in an environment where the phenotype matches. 51

4.2. Correlations between plasticity, mass, and fitness in G. sigillatus among and within the two sexes. 61

4.3. Raincloud plots showing the inter- and intra-genetic variation for A) fitness, B) plasticity, and C) mass for males (blue) and females (red) for each line..... 64

LIST OF ABBREVIATIONS

AI .....Active irreversible plasticity.

AR.....Active reversible plasticity.

PI.....Passive irreversible plasticity.

PR.....Passive reversible plasticity.

OI .....Irreversible organismal error.

OR.....Reversible organismal error.

PE.....Permanent environmental effects.

TE.....Temporary environmental effects.

OR.....Reversible organismal error.

Act.....Activity.

Agg.....Aggression.

PR.....Predator cue response.

GxE .....Genotype by environment interaction.

IGV .....Intra-genotypic variation.

LIST OF SYMBOLS

θ .....Population optimum

G .....Genetic variance-covariance matrix

β .....Selection gradient

ω.....Selection surface

$\mathbf{g}_{max}$.....Dominant eigenvector of **G**

E.....Environmental variance-covariance matrix

$r°$ .....Vector correlation

P .....Phenotypic variance-covariance matrix

I.....Among-individual variance-covariance matrix

W.....Within-individual variance-covariance matrix

$\mathbf{w}_{max}$.....Dominant eigenvector of **W**

$\mathbf{i}_{max}$.....Dominant eigenvector of **I**

ψ.....Probability of similarity between matrices

LIST OF APPENDIX TABLES

Table — Page

A.1. The eight mark-recapture models fit to investigate the effect of behavior on survival. ... 78

A.2. The eight mark-recapture models fit to investigate stabilizing selection on the three different behaviors. The directional selection terms for the three behaviors are included in all models. ... 79

A.3. Number of trials for the different sexes and developmental stages. ... 79

A.4. Coefficients as standard deviation for the fixed effect from the multi-response generalized mixed effect model. Values in parenthesis shows 95% HDP-intervals. ... 80

A.5. Survival and recapture coefficients estimated for sex, developmental stage, mass, and the three behavioral measurements based on the full mark-recapture model (Table A.1., Model 1). ... 81

A.6. AICc values for the 8 mark-recapture models. The reduced model has the best fit, indicating that the behavioral traits had very little impact on survival. ... 82

A.7. Effect of sex, developmental stage, mass, and the linear and quadratic terms for the three behavioral measurements on survival and recapture probability based on the post-hoc mark-recapture model. ... 83

A.8. AICc comparison for the mark-recapture models with quadratic terms. The reduced model has the best fit, but not significantly different from the model with the quadratic term for aggression. ... 84

A.9. Vector correlations (above the diagonal) and angles (below the diagonal) between the dominant eigen vector of the among- and within-individual covariance matrices and the selection gradient (β). ... 85

LIST OF APPENDIX FIGURES

Figure Page

A.1. Map of the Cassel wood, Minnesota (46°56'34.6"N, 96°47'03.2"W). The trapping grid's location is indicated by the white square. 77

A.2. Arena set up. The black line indicates the position of the mirror in the mirror arena, while the circle indicates the position of the predator cue in the anti-predator arena 77

A.3. Mean differences in A) distance from predator cue in the predator response test (PRT) and the open field (OF), and B) time spent on the mirror side of the arena in the mirror test (MT) and open field. The distance from the predator cue was substantially larger in the predator response test (pMCMC = 0.00), and the time spent in front of the mirror was substantially longer in the mirror test compared to the open field (pMCMC = 0.008). 80

1. GENERAL INTRODUCTION

Traits under selection should have a phenotypic distribution reflecting the selection pressure within a population (Phillips and Arnold 1989, Armbruster and Schwaegerle 1996, Arnold et al. 2008). A phenotype consists of both genetic and environmental variation, and selection can act upon both to shape the phenotype. To understand the distribution of a phenotype it is therefore necessary to understand the relationship between selection, genetic variation, and adaptive phenotypic plasticity (i.e., the phenotypic changes based on the environment the phenotype is expressed in). Theory predicts that both the dimension with most genetic variation and most plastic variation should align with each other and with the directional selection gradient (**β**), i.e., they should all be collinear (Jones et al. 2007, Arnold et al. 2008, Draghi and Whitlock 2012, Berdal and Dochtermann 2019, Noble et al. 2019, Wright et al. 2019).

Because traits often correlate with other traits, selection on one trait will have an effect on correlated traits, and vice versa (Lande 1979, Lande and Arnold 1983, Phillips and Arnold 1989), and when investigating how selection shapes a phenotypic distribution it should be done from a multi-trait point of view. For genetic variation, this multivariate structure is captured in the genetic variance-covariance matrix, i.e., the **G**-matrix (Lande and Arnold 1983), where the dominant eigenvector from this matrix ($\mathbf{g}_{max}$) corresponds to the dimension of most genetic variation (Lande and Arnold 1983). Theoretical models have shown that $\mathbf{g}_{max}$ and **β** should align over time, especially when the fitness optimum is stable (Jones et al. 2007, Arnold et al. 2008). Populations where $\mathbf{g}_{max}$ and **β** are aligned will respond faster to the selection pressure, reaching their fitness optimum faster than if there was a misalignment (Schluter 1996). A misalignment might indicate that either the traits being investigated have reached their fitness optimum and are

no longer under directional selection, or that the selection pressure has changed, and the underlying genetics have not had enough time to realign with the new selection gradient.

Because the phenotype also consists of environmental variation, adaptive phenotypic plasticity could increase the alignment between phenotypic variation and selection, even though the genetic variation is misaligned. Adaptive plasticity allows an individual to match its phenotype to the environment, and plasticity should therefore be aligned with selection (Gavrilets and Scheiner 1993). Plasticity should also align with genetic variation because plasticity masks the genetic variation in this dimension, decreasing the selection pressure and thus increasing the genetic variation in this dimension (Draghi and Whitlock 2012). While genetic variation takes several generations to realign with the selection gradient, plasticity can respond much quicker to a change in selection pressure and might be more important for a population establishing in a new environment (Price et al. 2003, Ghalambor et al. 2007, Draghi and Whitlock 2012), giving genetic variation time to realign with the new selection gradients over time (Jones et al. 2007, Arnold et al. 2008).

The alignment between plasticity and both selection and genetic variation is affected by two factors, namely what proportion of phenotypic plasticity consist of adaptive plasticity and the cost of plasticity. Phenotypic plasticity consists of other types of plasticity than adaptive plasticity, e.g., passive plasticity and organismal error (Westneat et al. 2015). Only adaptive phenotypic plasticity is predicted to align with selection and genetic variation, and if adaptive plasticity only contributes a small proportion of the phenotypic plasticity the alignment will be reduced (Westneat et al. 2015, Berdal and Dochtermann 2019). High cost of plasticity will also reduce the alignment between plasticity and selection and genetic variation because it will reduce

the benefit of the plastic response and not be as strongly favored by selection (DeWitt et al. 1998, Murren et al. 2015).

In my first chapter I explain in depth the adaptive theory predicting the alignment between genetic variation, plasticity, and selection, as well as factors that could lead to misalignment between them, as summarized above. I also outline a method for how to estimate alignment. In my second chapter I use a field experiment on a wild population of deer mice (*Peromyscus maniculatus*) where I empirically test for alignment between selection, genetic variation, and plasticity based on the prediction and suggested methods described in my first chapter. The use of a wild population allows me to calculate selection gradients for individuals under natural selection, an important factor for testing the predicted alignments. I repeatedly captured individuals using live traps and tagged them at first capture for identification at subsequent recaptures. I recorded repeated measures of three behaviors for each individual: activity, aggression, and response to a predator cue. In this study I used among- and within-individual covariance matrices as proxies for genetic variation and plasticity, respectively. The among-individual covariance matrix consists of both genetic and permanent environmental (**PE**) variation, where about 50 % of the variation is contributed by genetic variation (Dingemanse and Dochtermann 2014, Dochtermann et al. 2015). The within-individual covariance matrix consists of temporal environmental variation, including adaptive reversible plasticity (Whitman and Agrawal 2009, Westneat et al. 2015). Selection gradients were estimated from survival probability based on recapture data. Estimating selection gradients, among-individual variation (genetic variation), and within-individual variation (plasticity) allowed me to empirically test my predictions laid out in my first chapter, something that, to my knowledge, has not been done in a wild population before.

In my third and final chapter I investigate the cost of plasticity in isogenic lines of banded crickets (*Gryllodes sigullatus*), as cost of plasticity is one of the main predicted causes for misalignment between plasticity and both genetic variation and selection. Isogenic lines allow me to look at the expression of the same genotype in different environment, which is necessary for investigating cost of plasticity (Van Tienderen 1991, Van Buskirk and Steiner 2009). Plasticity was measured as absolute change in activity level in the absence vs. presence of a predator cue, while fitness was measured as number of eggs laid in females and spermatophore mass in males. A cost of plasticity would be detected if there was a substantial negative correlation between plasticity and fitness, i.e., genotypes that invest more in plastic responses have reduced fitness. I also investigated the inter- and intra-genetic variation among the lines. Inter-genetic variation (i.e., genetic variation among the lines) is necessary for any potential evolutionary change in a trait and estimating this allowed me to investigate whether there was any potential for plasticity to be under selection in this population. Intra-genetic variation is also interesting to investigate as this shows that some genotype varies more in their phenotypic expression than other, which could be a form of bet-hedging (Stamps et al. 2013, Ayroles et al. 2015).

My dissertation lays out an in-depth explanation for the prediction for alignment between genetic variation, plasticity, and selection based on evolutionary theory, as well as a method for how to estimate these alignments. I also empirically tested these predictions in a wild population of deer mice, something that has not been tested in wild population before. Lastly, I tested the assumption that there are costs associated with plasticity, which is one of the main factors causing a misalignment between plasticity and both genetic variation and selection.

2. ADAPTIVE ALIGNMENT OF PLASTICITY WITH GENETIC VARIATION AND SELECTION

2.1. Abstract

Theoretical research has outlined how selection may shape both genetic variation and the expression of phenotypic plasticity in multivariate trait space. Specifically, research regarding the evolution of patterns of additive genetic variances and covariances (summarized in matrix form as **G**) and complimentary research into how selection may shape adaptive plasticity lead to the general prediction that **G**, plasticity, and selection surfaces are all expected to align with each other. However, less well discussed is how this prediction might be assessed and how the modelled theoretical processes are expected to manifest in actual populations. Here, we discuss the theoretical foundations of the overarching prediction of alignment, what alignment mathematically means, how researchers might test for alignment, and important caveats to this testing. The most important caveat concerns the fact that, for plasticity, the prediction of alignment only applies to cases where plasticity is adaptive, whereas organisms express considerable plasticity that may be neutral or even maladaptive. We detail the ramifications of these alternative expressions of plasticity vis-à-vis predictions of alignment. Finally, we briefly highlight some important interpretations of deviations from the prediction of alignment and what alignment might mean for populations experiencing environmental and selective changes.

2.2. Introduction

Understanding the distribution of phenotypes within a population requires understanding three major and interacting contributors: genetic variation, phenotypic plasticity, and how each is shaped by natural selection. Selection shapes the distribution of phenotypes in many ways but, most simply, it is predicted to drive a population's mean phenotype towards a fitness optimum

and deplete genetic variation around that optimum (Figure 2.1.A.; Fisher 1930). In particular, under stabilizing selection, a population which has its mean at the fitness optimum is expected to lose variation around this mean, with the distribution of phenotypes narrowing around the optimum (Figure 2.1.A.). However, selection does not usually operate on single traits (Lande 1979, Lande and Arnold 1983), but instead affects multiple traits simultaneously. For example, particular combinations of traits might have a higher fitness compared to others combinations, i.e. correlational selection (Figure 2.1.B.; Endler 1986), and thus a population's distribution of phenotypes in multivariate trait space is expected to narrow around a multivariate optimum (Figure 2.1.C.) and can give rise to covariances among traits at the genetic level (Phillips and Arnold 1989, Armbruster and Schwaegerle 1996, Roff 1997).

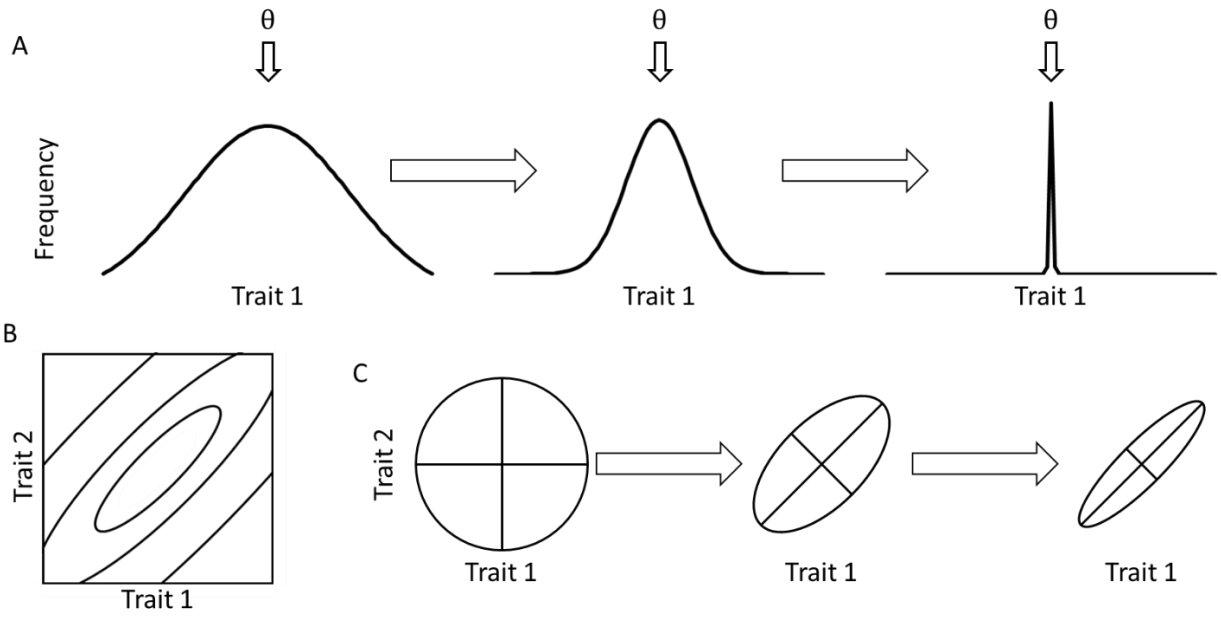

Figure 2.1. A. Response of a population under stabilizing selection where the population's mean is at the optimum (θ). Over successive generations, if the optimum does not change, variation around the mean is lost. **B**. An example of a selection surface (ω) for which fitness increases as traits approach intermediate values for both traits, i.e. when trait combinations are along the adaptive ridge. **C**. If a population's multivariate mean is at an optimum, variation will be depleted over successive generations so as to mirror the selection surface (B)—note that under stabilizing selection the distribution of variation is visualized as being similar to the topography of the selection surface and that, as a result, most variation is present in the direction of weakest selection (i.e. under stabilizing selection **G** is orthogonal to **ω**).

This narrowing of variation around an optimum changes the distribution of phenotypes in a population by changing the amount of genetic variation for those traits under selection (direct or otherwise) and by changing the magnitude and direction of trait covariances. Put another way, selection can affect the multidimensional geometric structure of genetic variances and covariances, as captured by the "**G** matrix" (following conventional notation bold values denote matrices and vectors rather than single values or effects). Phillips and Arnold (1989) provided the means by which the response of **G** to selection could be calculated (equation 2 therein) and simulation experiments have subsequently explored the long-term dynamics of how **G** is expected to track fitness optima (Jones et al. 2003, Jones et al. 2004, Jones et al. 2007). One

general finding from both analytical theory and simulations is that, over time and under directional selection, **G** is expected to *align* with the direction of selection, as measured either by selection gradients, **β** (Lande 1980), or Gaussian selection surfaces **ω**, (Jones et al. 2007, Arnold et al. 2008). Alignment in this regard can be defined as the dominant eigenvector of the **G** matrix (**g_{max}**, Schluter 1996) being collinear with a selection gradient or the dominant eigenvector of a static selection surface (Arnold et al. 2008) and can more generally be defined as geometric collinearity. Note that alignment only occurs when traits are under directional selection, where the fitness optimum is outside the existing genetic variation. Under stabilizing selection genetic variation will be depleted, and $\mathbf{g}_{max}$ will be in the direction of weakest selection (Figure 2.1.B. and 2.1.C.).

Two related factors affect the alignment between **G** and the selection surface: i) the stability of **G** over evolutionary time, and ii) whether the fitness optimum is constant, fluctuating, or changing in a constant direction. The stability of **G** will be determined by the relative contribution of pleiotropy versus selection-induced linkage disequilibrium (Roff 1997, Conner 2002, Conner et al. 2011), the contribution of mutations to covariances (e.g. Jones et al. 2007, Arnold et al. 2008), and the orientation of **G** relative to selection. For a linearly moving optimum, Jones et al. (2004) showed that **G** was more stable if the movement of the optimum was in the same direction as the $\mathbf{g}_{max}$. In contrast, an optimum moving in a direction different from $\mathbf{g}_{max}$, will decrease the stability of **G** (Jones et al. 2004). **g_{max}** will therefore be better aligned with the selection surface if the optimum is stable or moving in a constant direction**.**

While the above discussion provides general expectations as to the relationship between genetic covariation and selection, it ignores the fact that the distribution of phenotypes within a population is only partially due to genetic variation. In fact, the average heritability of traits is

estimated to be around 0.3 – 0.4 for behavior, physiological, and life-history traits and 0.55 for morphological traits (Mousseau and Roff 1987, Stirling et al. 2002, Dochtermann et al. 2019). The remaining ~70 – 45% of phenotypic variation stems from the influence of environmental factors on phenotypes, i.e. phenotypic plasticity—as well as developmental noise, e.g. stochastic gene expression—raising the question of whether we can similarly predict how environmental contributions to phenotypic variation and covariation should be oriented relative to selection.

Differences among individuals in phenotype due to differences in the environment experienced by these individuals fall under a broad operational definition of phenotypic plasticity (Lynch and Walsh 1998). If we partition phenotypic variation solely into genetic variation (i.e. **G**) and environmental variation (i.e. **E**), **E** necessarily captures the effects of plasticity on phenotypic variation (Whitman and Agrawal 2009). **E** also captures the effects of correlated plasticity on the expression of phenotypic traits: the off-diagonal elements of **E** are trait covariances due to environmental effects. These off-diagonal elements represent "multivariate trait plasticity" (Whitman and Agrawal 2009) which may stem from common developmental pathways, trade-offs in allocation of energy, or optimal combinations of phenotypic expression. When there is among-genotype variation in plasticity within a population—i.e., gene-by-environment interactions—plasticity can also respond to selection. Specifically, plasticity is expected to evolve such that the fitness of individuals increases when plasticity is expressed. Thus, if adaptive, plasticity should be aligned with selection surfaces (Gavrilets and Scheiner 1993, Draghi and Whitlock 2012). Put another way, adaptive plasticity is expected to alter an individual's phenotype to be closer to a fitness optima (Gotthard and Nylin 1995) and so the expression of variation due to environmental effects will be oriented in the direction of selection.

The multivariate expression of plasticity, as summarized by the geometric properties of **E**, will thus be expected to align with directional selection surfaces.

Interestingly, plasticity itself might also contribute to alignment between genetic covariation and the selection surface (Draghi and Whitlock 2012). Draghi and Whitlock (2012) argued that because the mechanisms necessary for phenotypic plasticity are expected to be more strongly expressed along the main axis of a selection surface, this would also increase the genetic variance found in this dimension. Phenotypic plasticity therefore contributes to the accumulation of cryptic genetic variation (Gibson and Dworkin 2004). Because adaptive plasticity masks this genetic variation, it is under weakened selection and genetic variation will therefore be greatest in the direction of plasticity, except when depleted under very strong selective pressure (Gavrilets and Scheiner 1993). If the new fitness optimum is stable over evolutionary time, this adaptive facilitation may ultimately contribute to evolution via genetic assimilation (Lande 2009), and $\mathbf{g}_{max}$ and the fitness surface will again be aligned.

The relevance of the above areas of research can be synthesized as: *i*) the distribution of genetic variation should align with selection surfaces (Jones et al. 2003, Jones et al. 2004, Jones et al. 2007, Arnold et al. 2008), *ii*) if plasticity is adaptive, it should be expressed such that variation aligns with selection surfaces (Gotthard and Nylin 1995, Draghi and Whitlock 2012), and *iii*) genetic variation and variation due to plasticity should be aligned if both are aligned to selective surfaces and this may be reinforced by the influence of plasticity on adaptation (Draghi and Whitlock 2012). Taken together this leads to a general prediction: adaptive plasticity, genetic variation (**G**), and selection surfaces should all be aligned in multidimensional trait space. Understanding the alignment among these three components (Figure 2.2.) can also help explain

why genetic variation sometimes acts as an evolutionary constraint and emphasizes the importance of plasticity for adaptation.

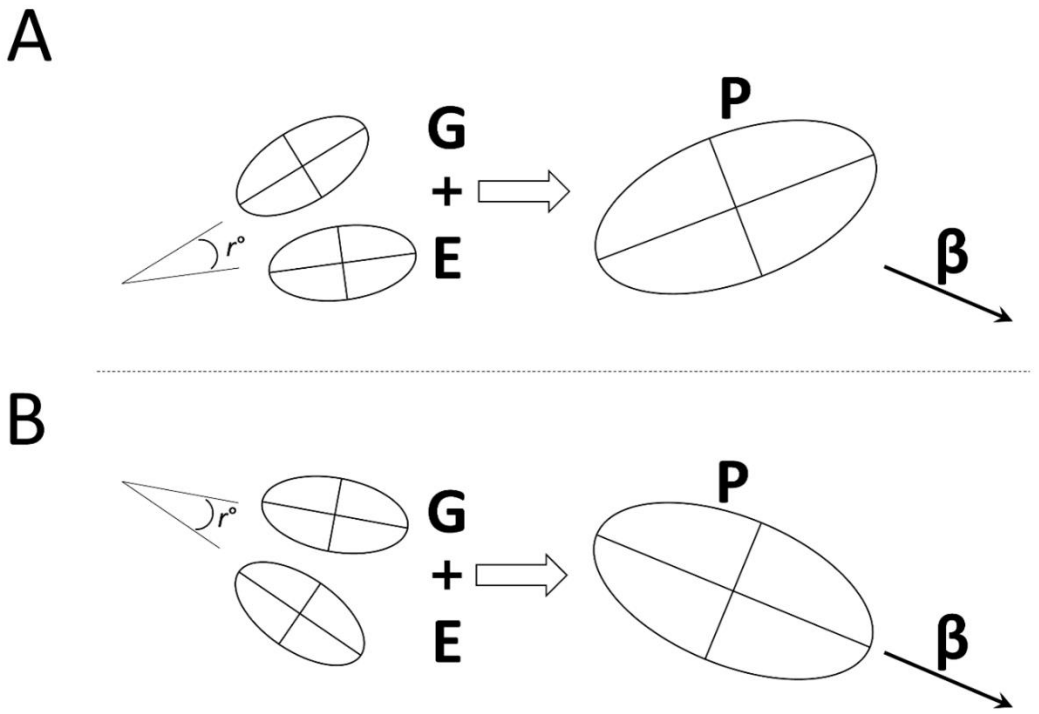

Figure 2.2. Relationship among **G**, **E**, **P**, and **β** in two populations (A and B). In both populations the main axes of genetic and environmental variation (i.e. the dominant eigenvector of **G** and **E**) are closely aligned with a minimal angle (*r*°) between them. Consequently, phenotypic variation is similarly oriented (**P**). However, in A, variation at all levels (**G**, **E** and **P**) is misaligned with the direction of selection (**β**) by ~45°. In contrast, in B, **G**, **E**, **P**, and **β** are all approximately aligned. Assuming the same amount of variation is present in each population and that the strength of selection is the same, this difference in alignment with **β** between the two populations means that population B will more rapidly respond to selection than will population A.

The only empirical test of the general prediction of alignment among **G**, plasticity, and selection that we are aware of was with *Daphnia pulex* that were exposed to cues of one of two predators, after which **G**, selection responses, and plasticity were estimated (Lind et al. 2015). **G** matrices and the expression of plasticity differed between predatory regimes but the difference

between the expressions of plasticity was greater than that observed for **G** matrices. Plasticity and **G** were also only aligned in one of the two predator exposure treatments.

While further tests of this general prediction are needed, two important questions about such testing remain: *i*) how do we test for this alignment; and *ii*) to what degree should we expect plasticity—manifested as environmental variation (**E**)—to be aligned with either **G** or selection? Here we describe statistical methods for empirically testing alignment among genetic variation, plasticity, and the selection surface, as well as an overview of contributors to **G** and **E**, and what implications these contributors have for the general prediction of alignment. These predictions unfortunately lack testing, and our motivation for discussing these issues here is because alignment, or lack thereof, is informative as to whether, how, and how quickly populations might respond to selection. The relative alignments of **E** and **G** with **β** or **ω** might also lead to hypotheses about current versus past selective pressures (Table 2.1.).

Table 2.1. Relevant evolutionary alignments and some of their biological significance.

	Biological Significance	**Relevant references**
G ∥ **ω** or **β**	Alignment between the dominant eigenvector of **G** ($\mathbf{g}_{max}$) and the direction or shape of selection. This alignment results in populations more rapidly responding to selection and suggests the possibility that the geometry of **G** has responded to selection and is itself adaptive.	Schluter (1996), Arnold et al. (2008)
E ∥ **ω** or **β**	Alignment between the dominant eigenvector of **E** ($\mathbf{e}_{max}$) and the direction or shape of selection. With this alignment plasticity is adaptive in that the expression of plasticity produces phenotypes approaching the optima.	this paper and Draghi and Whitlock (2012), Lind et al. (2015)
G ∥ **E**	Alignment between the dominant eigenvectors of **G** ($\mathbf{g}_{max}$) and of **E** ($\mathbf{e}_{max}$). If these are misaligned, and we assume plasticity is adaptive, it might suggest that current selective pressures are different than those in the past.	this paper and Draghi and Whitlock (2012), Lind et al. (2015)

2.3. Testing for alignment

Alignment, as operationally defined by Arnold et al. (2008, defined therein as shared eigenvectors), can be mathematically defined in terms of orientation in space, specifically as the angle between two vectors in *n* dimensional space (see also Lind et al. 2015). This can be calculated as the inner dot product of two vectors, (e.g. the dominant eigenvectors of **G** and the dominant eigenvectors of a matrix describing the shape of the selection surface (**ω**, Table 2.1.), i.e. the vector correlation:

$$r° = \frac{\boldsymbol{V_A^T V_B}}{\|\boldsymbol{V_A}\| \, \|\boldsymbol{V_B}\|} \qquad \text{(Equation 2.1.)}$$

where $\mathbf{V_A}$ and $\mathbf{V_B}$ are the two vectors being compared. For this application the absolute value of the vector correlation, $|r°|$, is of interest and can also be converted to degrees:

$$arccos(|r°|) \times \frac{180}{\pi} \qquad \text{(Equation 2.2)}$$

An $|r°|$ statistically indistinguishable from 1, i.e. the angle is indistinguishable from 0°, would then mean that matrices or matrices and the selection surface are aligned, or that there is insufficient power to detect differences. Alternatively, vector correlations significantly different from 1 would demonstrate misalignment.

Correlations are typically evaluated against a null expectation of 0 (corresponding to an angle of 90° between vectors), but here we are interested in whether alignment differs from a vector correlation of 1. As such, conventional approaches to calculating p-values are not appropriate. Moreover, because we are comparing vectors and matrices *within* the same population, randomization approaches described by Roff et al. (2012) and Aguirre et al. (2014) to generate null expectations for vector correlations between groups are not appropriate either as there are no groups over which randomization could be conducted. Given these limitations, and until better alternatives can be developed, two approaches exist for testing the significance of

alignment. First, when only point estimates of covariance matrices and/or vectors are available, standard approaches to comparing correlation coefficients can be used. Specifically, $|r°|$ can be converted to a *Z* value by taking its inverse hyperbolic tangent and testing its difference from null expectations (e.g. $r^{o}_{null} = 0.975$; if set to 1, equation 2.3 goes to -∞). The difference between the *Z* values is then divided by the pooled standard deviation and compared to a normal distribution (mean = 0, standard deviation = 1) to determine significance:

$$z = \frac{arctanh(|r°|) - arctanh(\mathring{r}_{null})}{\sqrt{\frac{2}{n-3}}} \qquad \text{(Equation 2.3)}$$

where *n* is an estimate of the sample size after controlling for non-independence (see also Noble et al. 2017). *n* can be estimated as the sample size at the highest level of the data's hierarchical structure (e.g. number of families or number of parent-offspring pairs rather than the total number of individuals sampled) or via adjusting for trait repeatability (Noble et al. 2017). Since r^{o}_{null} does not have an associated sample size it is here assumed to have the same uncertainty as $r°$ leading to the specified denominator (Zar 1999).

As an example of the calculation of vector correlations and the *Z*-value based testing approach, consider two matrices:

$$\text{M1} = \begin{bmatrix} 30 & -15 & 0 \\ -15 & 40 & 0 \\ 0 & 0 & 50 \end{bmatrix} \text{ and M2} = \begin{bmatrix} 25 & 0 & 0 \\ 0 & 40 & 20 \\ 0 & 20 & 55 \end{bmatrix}$$

The dominant eigenvector for M1 is $\begin{bmatrix} 0.58 \\ -0.81 \\ 0 \end{bmatrix}$ and the dominant eigenvector for M2 is $\begin{bmatrix} 0 \\ 0.57 \\ 0.82 \end{bmatrix}$.

Based on equation 2.1 $|r°|$ is 0.46. Following equation 2.3, with an *n* of 25 *Z* is then -5.60. From this we would then conclude that M1 and M2 have an angle of 62.48 and that this angle significantly differs from 1 ($p = 2.19 \times 10^{-8}$).

Two main caveats to these approaches should be noted. First, the relationship between the vector correlation $r°$ and the angle in degrees is not linear (Figure 2.3.A.). Since $r°$ is estimated with uncertainty, this non-linearity results in uncertainty that is not uniform over the 0 : 90 degrees by which vectors might differ (Figure 2.3.B.). Because this uncertainty is highest around $r° = 0.7$, estimation of intermediate angles and angles approaching 0 will be more imprecise than larger angles. Second, this approach is dependent on the degree to which $\sqrt{\frac{2}{n-3}}$ accurately estimates the uncertainty in $r°$. Given that estimates of **G** are typically made with considerable uncertainty, more uncertainty than is expected for correlation coefficients, equation 2.3 will be anticonservative for **G**. When possible, more appropriate standard errors should be used in the denominator. Given these concerns, significance—or lack thereof—of $r°$ should be interpreted with caution despite its biological importance.

A second approach is possible if posterior distributions from Bayesian analyses are available instead of just point estimates of matrices and vectors. Following Ovaskainen et al. (2008) and Lind et al. (2015), the posterior distributions of two matrices or vectors (**A** and **B**) can be compared based on the distribution of the observed vector correlation between **A** and **B** versus a null expectation of this angle (**A** versus **A** and **B** versus **B** across two estimates within the posterior):

$$\psi_{r°}(\lambda^A, \lambda^B) = \left[r°\left(\lambda_i^A, \lambda_j^A\right) + r°\left(\lambda_i^B, \lambda_j^B\right)\right] - \left[r°\left(\lambda_i^A, \lambda_j^B\right) + r°\left(\lambda_i^B, \lambda_j^A\right)\right] \quad \text{(Equation 2.4)}$$

where λ^A and λ^B are the posterior distributions of the dominant eigenvectors of matrices **A** and **B** (i.e. an estimate of **G**, plasticity, or selection surfaces) or the posterior estimate of a selection gradient. *i* and *j* correspond to two different posterior samples from the posterior distribution of either *A* or *B*. The distribution of $\psi_{r°}$ can then be assessed versus 0 and, if 95% of $\psi_{r°}$ estimates are greater than 0, **A** and **B** can be deemed significantly misaligned.

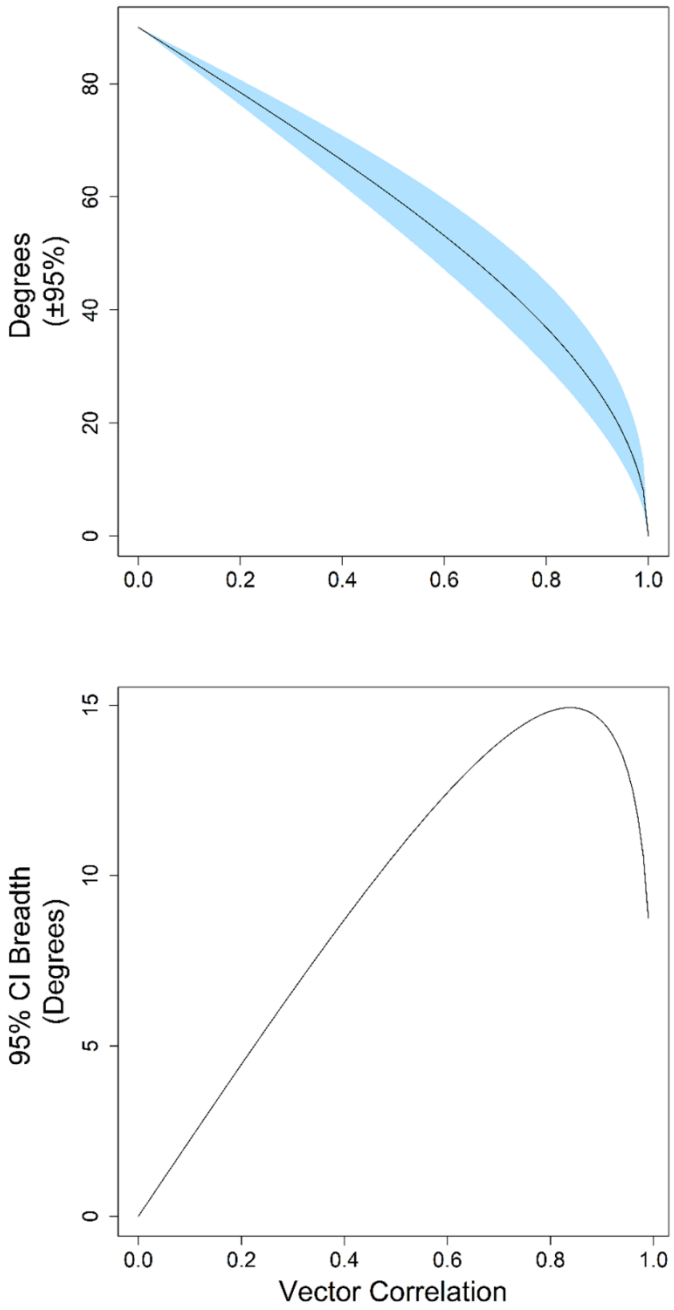

Figure 2.3. A. Relationship between a vector correlation and the degree of the angle between two vectors in multivariate space. The shaded area represents a 95% confidence interval around the degree of the angle for constant uncertainty of the vector correlation. **B**. The breadth of the 95% CI (A) around the degrees of the angle relative to the magnitude of a vector correlation. Uncertainty around r° was held constant but because of the non-linear relationship (A), the uncertainty in the angle is greatest at high values of r°.

2.4. Expectations for the alignment of E

As previously discussed, several lines of theoretical research lead to the prediction that selection surfaces, **G**, and the direction in phenotypic space in which plasticity manifests will be aligned. However, it is important to ask what this means for natural systems and how alignment will be expressed in populations. From the perspective of quantitative genetics, another way to ask this question is: how do we expect particular variance-covariance matrices and vectors to be aligned?

Starting from the simplest conceptualization of variances and covariances, we can model the (co)variances of quantitative traits of an organism as stemming from the additive contribution of genetic and environmental (co)variances:

$$\mathbf{P} = \mathbf{G} + \mathbf{E} \qquad \textbf{(}\text{Equation 2.5)}$$

where **P**, **G**, and **E** are symmetrical matrices containing phenotypic, genetic, and environmental variances, respectively, along the diagonal and covariances off the diagonal. While **E** is often dismissed as random or residual variation, it is necessarily phenotypic plasticity as, under equation 2.5, all variation expressed among genotypes due to experiencing different environments will be partitioned into the **E** matrix (Whitman and Agrawal 2009). While plasticity expressed in response to known environmental influences may be directly estimated by estimating equation 2.5 using mixed-effect animal models (Kruuk 2004), plasticity—adaptive or otherwise—in response to unknown or otherwise unmodeled environmental influences will still be captured by **E**.

Given this, we can predict that **G** and **E** should be "aligned" insofar as **E** captures adaptive plasticity and, following equations 2.1 – 2.4, the significance of departure from this alignment can be assessed. However, in real organisms, both **G** and **E** subsume considerable and

important biology. **G** can be further decomposed to the unique effects of additive, dominance, and other epistatic effects. All the contributors to **G** (e.g. dominance and other epistatic effects) might be predicted to be aligned with a selection surface or gradient, although this has been examined primarily for additive effects alone (Arnold et al. 2008). Likewise, **E** contains numerous sources of environmental influences each of which can be considered "plasticity" according to the operational definition given earlier. More specific contributors to **E** can be defined by collapsing categories from Westneat et al. (2015) and include:

Active irreversible plasticity (AI; West-Eberhard 2003): the phenotype of an organism changes in response to an environmental cue indicative of selective pressures. These changes are persistent—at least over the entire time period of measurement, i.e. permanent environmental effects (Lynch and Walsh 1998).

Active reversible plasticity (AR): reversible changes in an individual's phenotype expressed in response to environmental cues indicative of selective pressures and for which variability can occur within an individual (also termed phenotypic flexibility; Piersma and Drent 2003, Piersma and Van Gils 2011).

Passive plasticity: phenotypic changes in response to environmental conditions rather than cues of selective pressures. This includes passive responses to abiotic conditions, such as hypoxic conditions (Whitman and Agrawal 2009), and thus includes developmental instability. This passive plasticity can take the form of either *passive irreversible* (PI) or *passive reversible plasticity* (PR).

Organismal error: changes to an organism's phenotype due to a failure to process a cue correctly. This error can take the form of either *irreversible* (OI) or *reversible organismal error* (OR).

While this list of six contributors to **E** is hardly comprehensive (see also Whitman and Agrawal 2009, Forsman 2015, Westneat et al. 2015), it illustrates the point that the prediction that **E** will align with selection surfaces and **G** is imprecise: not all of these sources of environmental variation are expected to align with selection surfaces or **G**. Specifically, we would only predict that active irreversible (**AI**) and active reversible plasticity (**AR**) are aligned with selection surfaces or **G**. Passive plasticity and organismal error are expected to be largely independent of selection and thus to produce variation orthogonal or otherwise misaligned with selection.

We can include these concerns within equation 2.5 by incorporating the relevant contributors to variances and covariances:

$$\mathbf{P} = \mathbf{G} + \underbrace{\mathbf{AI} + \mathbf{PI} + \mathbf{OI}}_{\substack{\text{Permanent Environmental Effects (}\mathbf{PE}\text{)}}} + \underbrace{\mathbf{AR} + \mathbf{PR} + \mathbf{OR}}_{\substack{\text{Temporary Environmental Effects (}\mathbf{TE}\text{)}}} \qquad \text{(Equation 2.6)}$$

If individual organisms are measured across environmental conditions or individuals are measured multiple times specific contributors to **PE** can be estimated (Taylor et al. 2012, discussed below, Thomson et al. 2018), potentially distinguishing among some of the matrices of equation 2.6. Unfortunately for our prediction of matrix alignment, it will often be impossible to distinguish all active plasticity from passive plasticity and organismal error, as they will be occurring simultaneously. Thus **AI**, **PI**, and **OI** will be conflated as general "permanent

environmental effects" (**PE**) and AR, PR, and OR effects will typically be conflated as general "temporary environmental effects" (**TE**). As a result, matrix alignment of **G** and **E** will be reduced by the degree of misalignment of passive plasticity and organismal error and their relative contributions to trait correlations.

Following Roff (1997) and Dingemanse and Dochtermann (2014) the combined contributions of these contributors to permanent and temporary environmental effects and to environmental correlations (i.e. the pair-wise correlation coefficients corresponding to the off-diagonal elements of **E** and its components) are:

$$r_{PE} = r_{AI}\sqrt{\frac{V_{AI_y}}{V_{PE_y}}\frac{V_{AI_z}}{V_{PE_z}}} + r_{PI}\sqrt{\frac{V_{PI_y}}{V_{PE_y}}\frac{V_{PI_z}}{V_{PE_z}}} + r_{IO}\sqrt{\frac{V_{IO_y}}{V_{PE_y}}\frac{V_{IO_z}}{V_{PE_z}}} \qquad \text{(Equation 2.7)}$$

$$r_{TE} = r_{AR}\sqrt{\frac{V_{AR_y}}{V_{TE_y}}\frac{V_{AR_z}}{V_{PE_z}}} + r_{PR}\sqrt{\frac{V_{PR_y}}{V_{TE_y}}\frac{V_{PR_z}}{V_{TE_z}}} + r_{RO}\sqrt{\frac{V_{RO_y}}{V_{TE_y}}\frac{V_{RO_z}}{V_{TE_z}}} \qquad \text{(Equation 2.8)}$$

$$r_{E} = r_{PE}\sqrt{\frac{V_{PE_y}}{V_{E_y}}\frac{V_{PE_z}}{V_{E_z}}} + r_{TE}\sqrt{\frac{V_{TE_y}}{V_{E_y}}\frac{V_{TE_z}}{V_{E_z}}} \qquad \text{(Equation 2.9)}$$

where $V_{\_y}$ and $V_{\_z}$ correspond to the variance of traits *y* and *z* due to a particular factor (_, above). Likewise, the correlation coefficients on the right hand side of equations 2.7 – 1.9 are the correlations between traits *y* and *z* due to each of the factors described above. These correlation coefficients are calculated by variance-standardizing the covariances in each matrix in equation 2.6. With this framework we can begin to understand how the prediction of alignment will be manifested. Importantly, given equations 2.6 – 2.9, the alignment between **G** and **E** can be quite different than expected, even if active plasticity (**AI** and **AR)** and genetic variation (**G**) are perfectly aligned.

As a simple worked example, assume that three traits each have a heritability of 0.3 (Mousseau and Roff 1987, Stirling et al. 2002, Dochtermann et al. 2019), that the phenotypic variance of the three traits differ (here 0.60, 4.27, and 4.13, randomly drawn from a uniform distribution), that the genetic correlation between traits is 0.57 (Dochtermann 2011), and that the magnitude of alignment ($|r°|$) is 1 among active irreversible plasticity (**AI**), active reversible plasticity (**AR**), and genetic variation (**G**, Table 2.1.). For demonstration purposes we will also assume that organismal error results in an environmental correlation that is orthogonal to that observed for active plasticity and that passive plasticity is uncorrelated among traits (Table 2.2.). If all the components contributing to V_{TE} and V_{PE} do so equally (Table 2.2.), the end result is that $r_E = 0$ because active plasticity and organismal error offset each other. Given $r_G = 0.57$, $r°$ is then 0.70 (45.4°) and **G** and **E** are partially misaligned.

Table 2.2. Components contribution to phenotypic variation with example expected values and corresponding correlations. Equation 2.5. – 2.7. can also be represented as covariances. Correlations are presented to simplify equations.

Variances and correlations	Expected value	Rationale	
V_P			
V_G	$0.3 \times V_P$	Average heritability of life-history, physiological, and behavioral traits is approximately 0.3	(Mousseau and Roff 1987, Stirling et al. 2002)
V_{PE}	$0.3 \times V_P$	V_G and V_{PE} are of similar magnitude for behavior, although this relationship is unknown for other traits	(Dochtermann et al. 2015)
V_{AI}	$0.1 \times V_P$	Assumes these three components contribute equally to V_{PE} .	
V_{PI}	$0.1 \times V_P$		
V_{OI}	$0.1 \times V_P$		

Table 2.2. Components contribution to phenotypic variation with example expected values and corresponding correlations (continued). Equation 2.5. – 2.7. can also be represented as covariances. Correlations are presented to simplify equations.

Variances and correlations	Expected value	Rationale	
V_{TE}	$0.4 \times V_P$	$V_{TE} = V_P - V_G - V_{PE}$	
V_{AR}	$0.1\bar{3} \times V_P$	Assumes all three components contribute equally to V_{TE}	
V_{PR}	$0.1\bar{3} \times V_P$		
V_{OR}	$0.1\bar{3} \times V_P$		
V_E	$0.7 \times V_P$	Equation 5	
r_G	0.57	Absolute average of genetic correlations for behaviors	(Dochtermann 2011)
r_{PE}	Equation 6		
r_{AI}	0.57	Assumes perfect alignment between $\mathbf{g_{max}}$ and plasticity	
r_{PI}	0	Assumes that passive plasticity is agnostic to selection	(Whitman and Agrawal 2009)
r_{OI}	-0.57	Assumes that errors result in phenotypes orthogonal to $\mathbf{g_{max}}$	
r_{TE}	Equation 7		
r_{AI}	0.57	Assuming perfect alignment between $\mathbf{g_{max}}$ and plasticity	
r_{PI}	0	Assumes that passive plasticity is agnostic to selection	(Whitman and Agrawal 2009)
r_{OR}	-0.57	Assumes that errors result in phenotypes orthogonal to $\mathbf{g_{max}}$	

As the relative contribution of active plasticity to **E** increases, so too will alignment between **G** and **E**, if active plasticity and **G** are both aligned with the population's selection surface. For example, under the same conditions as above except without passive plasticity and with active plasticity contributing twice as much to permanent and temporary environmental variation as organismal error, r_E now equals 0.19, and $r°$ is 0.99 (6.3°), indicating alignment. The ability to detect alignment therefore depends, in part, on the relative contribution of each

component in equation 2.6 and the contribution to **E** from factors besides active plasticity will lead to an underestimate of r°.

Because environmental (co)variances are frequently ignored or moved to the denominator of effect sizes and test statistics, the relative contributions of each source of variation described in equation 2.6 is currently unknown. Fortunately, common approaches to estimating **G** also allow estimation of permanent environmental effects on phenotypic covariances. When groupings of individuals that may have experienced similar environmental effects are known, a multiple-matrix animal model approach allows estimation of permanent environmental covariances (Thomson et al. 2018). Using a multiple-matrix animal model, Taylor et al. (2012) estimated the contribution of additive genetic effects and three sources of permanent environmental (co)variation to phenotypic covariance in North American red squirrels (*Tamiasciurus hudsonicus*). The three, possibly adaptive, sources of permanent environmental effects considered were: *i*) environmental effects unique to individuals (based on repeated measures), *ii*) cohort determined permanent environmental effects (based on birth year), and *iii*) maternal effects, although maternal effects also have a heritable component (Räsänen and Kruuk 2007). In these red squirrels Taylor et al. (2012) found that the magnitude of these permanent environmental effects on phenotypic variation were of similar magnitude to the magnitudes estimated for additive genetic effects. More relevant to our discussion here, permanent environmental and maternal covariances were also of a similar magnitude and sign to the estimated additive genetic covariances (Taylor et al. 2012). This suggests that aspects of **AI** (equation 2.6) were indeed aligned with **G**.

2.5. Conclusions

Although previous research has suggested that **G** (e.g. Arnold et al. 2008) and **E** (e.g. Draghi and Whitlock 2012, Lind et al. 2015) should align with selection surfaces, this has not previously been detailed in the context of *both* theoretical models of the evolution of **G** and vis-à-vis plasticity. Moreover, previous research did not consider this issue specifically in terms of the **E** matrix. Considering the general prediction of alignment in terms of **E** emphasizes the generality of the prediction of alignment. While alignment of **G** and relevant contributors to **E** will be obscured due to the considerable biology each of these matrices subsumes (equation 2.6), we still consider the estimation of alignment to be an interesting and important pursuit. Alignment informs us of the ability of populations to respond to selection and the relative contribution of active plasticity to fitness. When selection surfaces have been estimated for natural populations, alignment also informs us of the effects of the structure of **G** and **E** on fitness and possible rates of adaptation. Moreover, detection of misalignment generates evolutionarily interesting hypotheses. For example, if **E** and **β** are aligned with each other but **G** is misaligned with both, one possible reason would be that current selection pressures (**β**) differ from those that shaped **G**. The genetic correlations will in this case be quantitively constraining evolutionary responses, but phenotypic expression may still match a fitness optimum through plasticity, i.e. **E** still aligns with the new fitness surface. Finally, estimation of alignment with current selective surfaces versus projected selective surfaces may provide indications of the ability of populations to respond to the increasing effects of humans on natural populations.

3. MISALIGNMENT OF SELECTION, PLASTICITY, AND AMONG-INDIVIDUAL VARIATION: A TEST OF THEORETICAL PREDICTIONS WITH *PEROMYSCUS MANICULATUS*

3.1. Abstract

Genetic variation and phenotypic plasticity are predicted to align with selection surfaces, a prediction that has rarely been empirically tested. Understanding the relationship between sources of phenotypic variation (primarily genetic variation and plasticity) and selection surfaces improves our ability to predict a population's ability to adapt to a changing environment and our understanding of how selection has shaped phenotypes. Here, we estimated the (co)variances among three different behaviors (activity, aggression, and predator cue response) in a natural population of deer mice (*Peromyscus maniculatus*). Using multi-response generalized mixed effects models, we divided the phenotypic covariance matrix into among- and within-individual matrices. The among-individual covariances include genetic and permanent environmental covariances and the within-individual (co)variances include reversible phenotypic plasticity. To determine whether genetic variation, plasticity and selection align in multivariate space we calculated the dimensions containing the greatest among-individual variation and the dimension in which most plasticity was expressed (i.e. the dominant eigenvector for the among- and within-individual covariance matrices respectively), and estimated selection coefficients based on survival estimates from a mark-recapture model. Alignment between the dominant eigenvectors of behavioural variation and the selection gradient was estimated by calculating the angle between them, with an angle of 0 indicating perfect alignment. The angles between vectors ranged from 68° to 89°, indicating that genetic variation, phenotypic plasticity, and selection are misaligned in this population. This misalignment could be due to the behaviors being close to

their fitness optima, which is supported by low evolvabilities, or because of low selection pressure on these behaviors.

3.2. Introduction

Populations under selection should have phenotypic distributions reflecting this selection (Phillips and Arnold 1989, Armbruster and Schwaegerle 1996, Arnold et al. 2008), and to maximize fitness the majority of phenotypic variation should align with selection gradients (β, Berdal and Dochtermann 2019, Noble et al. 2019, Wright et al. 2019). Put another way, the axes of most phenotypic variation should be collinear with selection gradients. Both genetic and environmental variation contribute to a population's phenotypic variation and both sources of variation are expected to align with selection (Jones et al. 2007, Arnold et al. 2008, Draghi and Whitlock 2012, Berdal and Dochtermann 2019).

Because traits often covary, selection pressures on one trait affect the mean and variance of other traits and vice versa (Lande 1979, Lande and Arnold 1983, Phillips and Arnold 1989). It is therefore necessary to investigate alignment between variation and selection from a multi-trait perspective. In quantitative genetics, the **G**-matrix summarizes additive genetic variation and covariation across traits (Lande and Arnold 1983), with the greatest variation present in the matrix's dominant eigenvector, $\mathbf{g}_{max}$. A population where **β** and $\mathbf{g}_{max}$ are aligned will reach a fitness optimum faster than if there was a misalignment (Schluter 1996). Simultaneously, selection is expected to have shaped **G** and models have shown that $\mathbf{g}_{max}$ will align with selection surfaces given enough time and stable selection pressures (Jones et al. 2007, Arnold et al. 2008).

Misalignment between $\mathbf{g}_{max}$ and **β** can be ameliorated by adaptive phenotypic plasticity, contingent on costs of being plastic. That is, phenotypic plasticity allows an individual to change their expression of a trait based on the environment and increase their fitness, i.e. adaptive

phenotypic plasticity (Ghalambor et al. 2007). Adaptive phenotypic plasticity is predicted to evolve in more heterogeneous environments, allowing populations to move closer towards trait optimum within environments (Van Tienderen 1991, Scheiner 1998, Sultan and Spencer 2002). Consequently, adaptive plasticity should align with the selection gradient (Gavrilets and Scheiner 1993).

More recently, Draghi and Whitlock (2012) showed that phenotypic plasticity should be expressed so as to align with $\mathbf{g}_{max}$, because it maintains, and possibly increases, genetic variation (Draghi and Whitlock 2012). Consistent with this, Noble et al. (2019) found via meta-analysis a significant alignment between $\mathbf{g}_{max}$ and direction of plasticity in novel environments. While genetic adaptation can only occur over several generations, phenotypic plasticity can change within the same generation and might therefore be the major contributor to alignment between phenotypic variation and selection in new environments (Price et al. 2003, Ghalambor et al. 2007, Draghi and Whitlock 2012).

However, because individuals do not show the full repertoire of phenotypes present in a population, fitness costs and/or limits to plasticity are expected to frequently be substantial (DeWitt et al. 1998, Van Buskirk and Steiner 2009, Murren et al. 2015). As costs to plasticity increase, $\mathbf{g}_{max}$ is predicted to more quickly align with selection, reducing the need for plasticity and its costs (DeWitt et al. 1998, Schneider and Meyer 2017). Consistent with this, Johansson et al. (2020) showed that divergence among populations of damselflies (*Lestes sponsa*) was aligned with both genetic variation and plasticity and that it had been shaped by selection.

In natural populations **G** is difficult to estimate because it requires both a pedigree and trait measurements. Consequently, alignment among **G**, plasticity, and selection gradients has rarely been assessed and existing predictions about their interplay have not been tested.

Fortunately, the among-individual variance-covariance matrix (**I**) can be estimated for a single generation by repeatedly sampling individuals from a population. The **I**-matrix consists of the joint effects of **G** and permanent environmental (**PE**) correlations (Dingemanse and Dochtermann 2014), each explaining approximately 50% of repeatable variance in behaviors (Dochtermann et al. 2015). Moreover, consistent with Cheverud's conjecture, phenotypic correlations are concordant with both genetic correlations (Cheverud 1988, Dochtermann 2011) and among-individual correlations (Brommer and Class 2017). **I** can therefore be used as a proxy for the **G**-matrix, albeit with caveats. Similarly, plasticity can be difficult to measure in wild populations. However, the within-individual variance-covariance matrix (**W**-matrix) contains temporary environmental (TE) correlations, i.e. correlation among traits due to changes in the environment within the timeframe of trait measurement. This includes adaptive reversible plasticity (Whitman and Agrawal 2009, Westneat et al. 2015) and $\mathbf{w}_{max}$, the dominant eigenvector of **W**, is the direction in which most plasticity is expressed. Thus, using among- and within-individual variance-covariance matrices as proxies for genetic variation and phenotypic plasticity respectively allows us to estimate their alignment with each other and selection surfaces without the need of a pedigree.

Here we examined alignment between selection, among-individual covariances, and phenotypic plasticity in a wild population of deer mice, *Peromyscus maniculatus*, by estimating the angles between the selection gradient (**β**) and the dimensions in phenotypic space containing the greatest amount of among- and within-individual behavioral variation. If fitness optima are stable both within and across generations for our population of deer mice, both among- and within-individual variation are expected to be aligned with the selection gradient and each other. However, if the selection pressure has changed recently, only plasticity is predicted to align with

the selection gradient, as genetic variation has not yet had enough time to realign. In addition, if the traits are close to the fitness optimum or under stabilizing selection genetic variation and selection gradients are predicted to be misaligned.

3.3. Methods

3.3.1. Study species

A wild population of deer mice was sampled at Cassel wood, Minnesota, USA (Figure A.1.). Deer mice are highly suited for our questions because repeatability (Brehm and Mortelliti 2018, Brehm et al. 2020) and appreciable additive genetic variation (Careau et al. 2011) have been previously demonstrated for behaviors similar to those measured here, making among-individual correlations more likely to be a suitable proxy for **G.** In addition, behavioral covariances have previously been estimated in the closely related congener *P. leucopus* (Bester-Meredith and Marler 2007). All research was conducted in accordance with institutional guidelines (NDSU IACUC **A17055)** and the guidelines of the Animal Behavior Society (Animal Behaviour Society 2020) and the American Society of Mammalogists (Sikes et al. 2016).

3.3.2. Trapping and tagging

Individuals for whom phenotypes and selection were estimated were repeatedly captured using Sherman live traps (5.2 × 6.4 × 22.9cm). Traps were set in a 9 × 11 grid, with traps 12.5 m apart, totaling 99 traps. Mixed birdseed and rolled oats were mixed with peanut oil and used to bait traps, and cotton was added to provide insulation. Trapping was conducted between May 30th and October 13th in 2017, with trapping occurring around three times a week. Traps were set between 3-6 pm and checked starting at 6 am the following morning.

Individual deer mice were tagged with metal ear tags in both ears at first capture for identification at recaptures. All captured individuals were identified by individual ID, had their

mass and sex recorded, and identified as either juvenile or adult (developmental stage). Individual ID was recorded to allow subsequent mark-recapture analysis, the estimation of survival, and for repeated behavioral trials.

3.3.3. Behavioural assays

To investigate the relationships between behavioral syndromes, plasticity, and selection in this population, captured deer mice were tested in three behavioral assays: an open field test, a mirror image stimulation test, and a predator cue response test. All behavioral tests were conducted in arenas (60 × 60 × 40 cm) made of 2 cm thick plywood. Transparent plexiglass was used as lids and a video camera was mounted above each arena (Figure A.2.). Deer mice were always put through the assays in the same order: 1) open field test, 2) mirror image stimulation test, and 3) predator cue response test. This was done to avoid any carry-over effects from interacting with the mirror on open field tests or from the exposure of the predator cue affecting open field or mirror image response. Individuals that had been through all three assays at least three times were given a shelter (cardboard cup with a small hole in the bottom) in their subsequent trials to see how this would affect their behavior. After the deer mice had been run through all three assays they were released at the same location as they were caught.

3.3.3.1. Open field test

In the open field assay, the arena was empty and was used to measure general activity level (Reale et al. 2007). Activity in an open field arena has been shown to give repeatable measures for general movement in several species of rodents (Herde and Eccard 2013, Hewes et al. 2017) including *Peromyscus maniculatus* (Brehm and Mortelliti 2018, Brehm et al. 2020). At the start of the open field assay deer mice were placed under a cup (11.5 cm in diameter) and given one minute to acclimate. The shelter was then removed, and the deer mouse had 6 minutes

to explore the arena. We walked at least 20 m away from the recording area to reduce any disturbance caused by our presence. After the deer mouse completed the open field assay, the cup was placed over the deer mouse again, and a cardboard plate was pushed under it to enable us to transport it to the mirror arena. Individual activity and location in the open field and other assays was tracked using Ethovision X (Noldus Information Technology, Wageningen, The Netherlands). Tracking was started 2 minutes into the video to eliminate disturbance from setting up assays that were simultaneously run for other individuals (see below) and to move away from the arena area. Distance moved (in cm) in the open field was used as a measure of activity.

3.3.3.2. Mirror image stimulation test

The arena for the mirror image stimulation test had a mirror attached along one wall (Figure A.2.). Mirror tests have been shown to be an appropriate measure of agonistic behavior in several species of rodents (Svendsen and Armitage 1973, Boon et al. 2008, Uchida et al. 2020), and to correlate with aggression towards conspecifics in other rodents (Dochtermann et al. 2012). Mirror tests also have the advantage of standardizing size differences between the opponents as well as avoiding any injury to individuals being tested. Here, deer mice were moved from the open field arena into the mirror arena, where they were again given one minute to acclimate under a cup. Video analysis started 1.5 minutes after removal of the cup (again, to avoid any disturbance caused by starting the predator cue response trial and then move 20m away) and the time spent on the mirror side of the arena was recorded in seconds and used as a proxy for aggression. If deer mice perceived their reflection as a conspecific, spending more time in front of the mirror would indicate interaction with conspecifics, while time spent away from the mirror indicates avoidance.

3.3.3.3. Predator cue response test

For the predator cue response assay a circular filter paper (11 cm in diameter) soaked in coyote urine was placed in the upper left corner of the behavioral arena (Figure A.2.). Small rodents have been shown to have aversive responses towards predator cues (Nolte et al. 1994, Hegab et al. 2014), including the closely related species *P. gossypinus* (Brinkerhoff et al. 2005). Both coyotes (*Lupus latrans*) and red foxes (*Vulpes vulpes*) have been observed in the study area, making canid urine generally and coyote urine specifically a suitable cue of predator presence for this deer mouse population. The filter papers were prepared elsewhere and stored at -20˚C until the morning of behavioral trials. After completing the mirror test, a deer mouse was moved to the predator cue response arena and put under a cup for a 1-minute acclimation period. In subsequent video analysis, tracking started 30 seconds into the video, and predator cue response was measured as the mean distance from the predator cue (in cm). More responsive individuals were predicted to increase their distance from the predator cue, indicating shy behavior, while bolder and less responsive individuals would stay closer to the cue.

3.3.4. Statistical analyses

Statistical analyses were conducted using the packages RMark (Laake 2013) and MCMCglmm (Hadfield 2010) in R (R Development Core Team 2015).

3.3.4.1. Assay validity

Mixed effect models were used to investigate the validity of the behavioral assays by comparing the average responses towards the mirror and the predator cue to behavior in the open field arena using the MCMCglmm package in R (Hadfield 2010). For the mirror assays, the time spent on the mirror side of the arena was compared with time spent on the equivalent side in the open field trial. This allowed us to investigate the behavioral response of individuals towards the

mirror stimuli, where a difference in time spent on the mirror side of the arena in the mirror test compared to the open field assay would indicate that individuals did respond to their own reflection. Similarly, in the predator cue response assay the distance from the predator cue was compared to the distance from the equivalent location in the open field assay, and an increase in distance from the predatory cue in the predator cue assay compared to the open field would indicate that individuals are avoiding the cue. Arena type (open field, mirror stimulation assay, and predator cue assay) was used as a fixed effect and animal ID was added as a random effect. Models were run with an inverse-Wishart prior and for 650,0000 iterations, with a thinning interval of 5,000, and a 150,0000 iteration burn-in. pMCMC values (used similarly to p-values for maximum likelihood statistics) were used to determine whether there was a substantive effect of arena type. pMCMC was calculated here as the proportion of estimates of the fixed effects from the posterior distribution that were below 0. A high proportion (>0.95) indicates that most estimates are below 0, while a low proportion (<0.05) indicates that most estimates are above 0. Both high and low proportion means that the posterior distribution has a very low overlap with 0 and is therefore considered to have a substantive effect on the behaviors.

3.3.4.2. Repeatabilities and correlation matrices

Repeatability and among- and within-individual variance-covariance matrices (**I** and **W**) were estimated for distance moved (activity), time spent in front of the mirror (aggression), and mean distance from predator cue (predator cue response) using a multi-response mixed effect model (Dingemanse et al. 2012, Dingemanse and Dochtermann 2013) also using the MCMCglmm package (Hadfield 2010). Sex, developmental stage, mass, and the presence of a shelter were included as fixed effects, where mass was within-subject centered (Van de Pol and Wright 2009). Individual ID was included as a random effect. Because distance moved had a

left-skewed distribution, this variable was square root transformed prior to all statistical analyses. All three behaviors were mean centered and scaled by their standard deviations to facilitate model fit. The number of iterations, thinning intervals, and burn-in was the same as above. Priors were minimally informative for (co)variances and flat for correlations. Repeatabilities, correlation matrices, and estimates of individual behaviour (i.e. best linear unbiased predictors, BLUPs) were estimated from the posterior distributions. pMCMC values were used to determine if the fixed effects had any substantial effect on behaviors (calculated as above).

3.3.4.3. Selection gradients

Directional selection gradients (**β**) for the three behaviors (activity, aggression, and predator cue response) were estimated from the effect of individual behavior (BLUPs from the multi-response generalized mixed effect model) on survival probability from mark-recapture models fitted using Program MARK and RMark (Laake 2013). Sex, developmental stage, mass, and the BLUPS for distance moved, time spent in front of the mirror, and mean distance from predator cue were included as covariates. We assumed a closed population and recapture probability was set as constant throughout the season. Because some of the individuals were too young to identify sex, sex was recorded as female (1 0), male (0 1), or unidentified (1 1). The coefficients for survival for the three behaviors were converted to a traditional Lande and Arnold (1983) directional selection gradient (**β**) using the method detailed in Waller and Svensson (2016).

To investigate whether behaviors had a substantive effect on survival, seven additional mark-recapture models were fit with 1-3 of the behavioral covariates to allow for comparison (Table A.1.). If any of the behaviors had a substantial effect on survival, models 1 – 7 would

have a better fit than the model with no behavior covariates (model 8). Relative model fit was determined based on AICc values.

3.3.4.4. Alignment

We used the among- and within-individual covariance matrices (**I**- and **W**-matrices) as proxies for genetic covariation and multivariate expression of adaptive plasticity respectively, to test predictions regarding alignment. $\mathbf{i}_{max}$ was the dominant eigenvector for the **I**-matrix and was the dimension with the most among-individual variation while $\mathbf{w}_{max}$ was the same for the **W**-matrix. The dominant eigenvectors for both the among- and within-individual covariance matrices were estimated from their posterior distributions estimated by the multi-response mixed effects model. These were then used to calculate the vector correlation and the angle between (i) $\mathbf{i}_{max}$ and $\mathbf{w}_{max}$, (ii) $\mathbf{i}_{max}$ and the selection gradient ($\boldsymbol{\beta}$), and (iii) $\mathbf{w}_{max}$ and $\boldsymbol{\beta}$. Vector (V) correlations were calculated using equation 2.1 and converted to degrees using equation 2.2. A vector correlation that does not differ from 1 (i.e., an angle that does not differ from 0) would mean that the vectors are aligned.

Because we were interested in whether a vector correlation was statistically distinguishable from 1, rather than common null expectation of 0, standard statistical analyses were not appropriate. We therefore used a Bayesian approach developed by Ovaskainen et al. (2008), which uses posterior distributions. Using the posterior distribution from the multi-response mixed effects model, we obtained 1000 estimates of the dominant eigenvectors for the among- and within-individual correlation matrices. To get similar estimates for the selection gradients, the full mark-recapture model described in Table A.1. (Model 1) was re-fit using the 1000 posterior estimates of the BLUPs for the three behaviors from the multi-response mixed effects model. Probability of alignment between among- and within-individual eigenvectors,

among-individual eigenvectors and selection gradients, and within-individual eigenvectors and selection gradients was then estimated using equation 2.4 (Ovaskainen et al. 2008, Berdal and Dochtermann 2019). The correlation between samples of the same vectors should be 1, thus the first part of the equation ≈ 2. The inclusion of this term incorporates estimation uncertainty present in the posterior distribution. The second part of the equation is the correlation between estimates of different vectors. If two vectors are highly correlated, the second part of the equation is also ≈ 2, and $\psi_{r\circ} \approx 0$, with sampling error allowing $\psi_{r\circ}$ to be below zero. As the correlation between Λ^A and Λ^B decreases, $\psi_{r\circ}$ approaches 2. If 95 % of the estimated $\psi_{r\circ}$ are above 0, the two matrices are considered misaligned. We therefore calculated the proportion of $\psi_{r\circ}$ values that were positive, with a higher value meaning that more estimates exclude 0 and will in this case indicate a substantial misalignment.

Unfortunately, the method of Ovaskainen et al. (2008) can be sensitive to low power so to further understand the alignment between vectors we also compared the estimated vector correlation to three different null expectations of the correlation following equation 2.3 (Berdal and Dochtermann 2019). The null expectations where set to 0.975, 0.95, and 0.9, which are considered highly correlated and indicates that the vectors are aligned. The differences between observed and null correlations were then transformed to z-values. Z-values larger than 0 would indicate that the estimated vector correlation was higher than the null expectation, which would mean that the vectors were aligned. One thousand estimates of vector correlations between $\mathbf{i}_{max}$ and $\mathbf{w}_{max}$, between $\mathbf{i}_{max}$ and $\boldsymbol{\beta}$, and between $\mathbf{w}_{max}$ and $\boldsymbol{\beta}$ were compared to the null expectations. We then calculated how many of the Z-values were above 0, i.e., how many of the estimated vector correlations were above null expectations. As before, if more than 95% of the estimates are above this value the misalignment was considered to be substantial.

3.3.4.5. Testing for historical and stabilizing selection

After comparing the eight mark-recapture models (Table A.1.) we found that the model without behavioral traits had the best fit, indicating that the behaviors measured here were under weak or no directional selection. Two post-hoc analyses were therefore carried out to estimate evolvability and the possibility of stabilizing selection (Table A.2.) for the three behavioral traits (see Appendix A), as both low variation and stabilizing selection will reduce alignment with the behavioral variation and selection. Evolvability was estimated as:

$$I_A = \frac{V_A}{\mu^2} \qquad \text{(Equation 3.1)}$$

where V_A is the trait variance and μ is the trait mean. V_A was scaled based on number of individuals in the different categories of sex and developmental stage.

3.4. Results

We sampled the population for 30 nights and captured 92 individuals (including individuals that escaped prior to being tagged or identified) with a mean recapture rate of 3. Of these, 72 individuals were behaviorally assayed and used in analyses. This included 32 females, 36 males, and 4 individuals of unidentified sex, where 48 were adults and 24 were juveniles. The other 20 individuals managed to escape either while being tagged or during the trials and were never caught again, meaning they had insufficient behavioral data collected to be used in our analyses. In total, we conducted 641 behavioral assays (Table A.3.), with a mean number of trials per individual of 3.13±2.65, 2.65±2.23, and 2.79±2.56 for activity, aggression, and predator cue response respectively.

3.4.1. Assay validity

We found that deer mice responded as predicted to the predator cue, increasing their distance from the cue in the predator response test compared to the corresponding corner during

the open field test (pMCMC = 0.00, Figure A.3.A.). Deer mice also spent more time on the mirror side of the arena compared to the controlled open field test (pMCMC = 0.008, Figure A.3.B.). In addition, repeatability for staying on the mirror side of the arena was 0.36 for the mirror-trial, while for the open field trial side preference was indistinguishable from 0, indicating that the deer mice stayed on one side of the arena consistently when the mirror was present but not when absent, providing further validation of the assay.

3.4.2. Average behavior

From the multi-response generalized mixed effect model we found that males had a higher activity level compared to females (pMCMC = 0.007), and the presence of a shelter reduced the distance moved in the open field, (pMCMC = 1,) and predator cue response assay, (pMCMC = 0.96). No other fixed effect substantively affected the assayed behaviors (Table A.4.).

3.4.3. Behavioral variation

All behaviors were moderately repeatable: activity, aggression, and predator cue response had estimated repeatabilities of 0.47 (0.35-0.60), 0.36 (0.24-0.50), and 0.33 (0.21 – 0.46) respectively (Figure 3.1.).

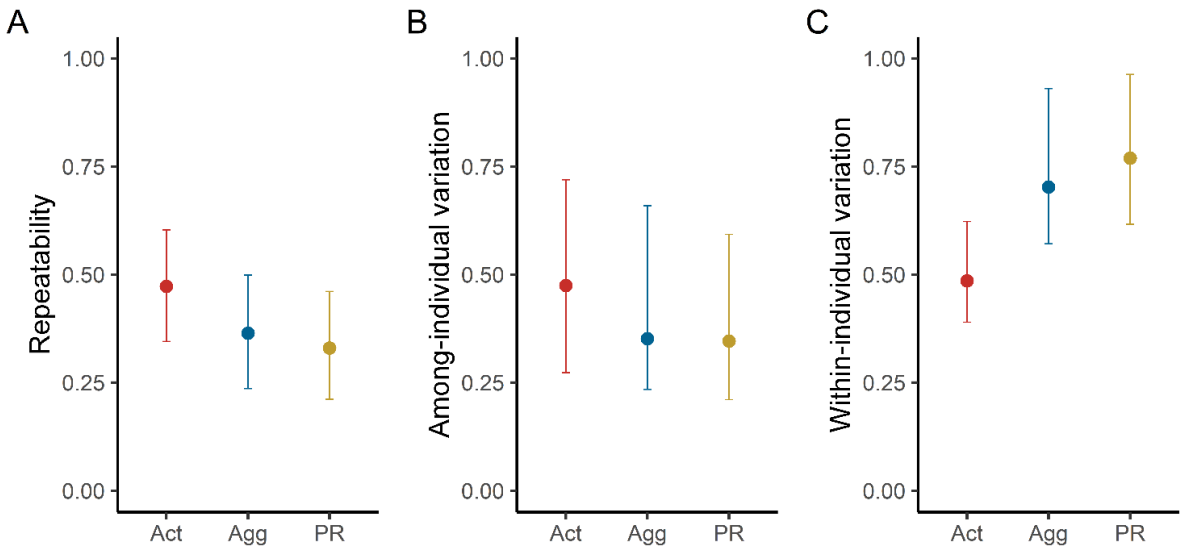

Figure 3.1. The mean and standard deviation for A) repeatability, B) among-individual variation, and C) within-individual variation for activity level (red, Act), aggression (blue, Agg), and predator cue response (yellow, PR). Behaviors were mean centered which is why the among- and within-variances are scaled between 0 and 1.

3.4.4. Correlation matrices

Predator cue response and activity in the open field were negatively correlated at both the among- and within-individual levels. Put another way, less active individuals stayed further away from the predator cue and individuals that reduced their activity level also increased their distance from the predator cue (Table 3.1.). The among-individual correlation between predator cue response and aggression was around 0, while the within-individual correlation was slightly negative (Table 3.1.). This negative correlation suggests that when individuals increased their time in front of the mirror, they also reduced their distance to the predator cue. The largest difference between the among- and within-individual correlation was for the relationship between activity and aggression (Table 3.1.). Here, the among-individual correlation was negative while the within-individual correlation was positive. This means that while individuals that are more active spend less time in front of the mirror, an individual that increases its activity

level will also increase its time spent in front of the mirror. However, only the within-individual correlations between activity and aggression and between activity and predator cue response were substantively different and statistically distinguishable from zero (Table 3.1.)

Table 3.1. Among-individual correlations (above the diagonal) and within-individual correlations (below the diagonal) between activity, aggression, and predator cue response. Substantive correlations (pMCMC > 0.95 or pMCMC < 0.05) are bolded. Evolvabilities are shown on the diagonal (shaded). All estimates are presented with their corresponding 95% credibility intervals in parentheses.

	Activity (Act)	**Aggression (Agg)**	**Predator cue response (PR)**
Act	0.22 (0.08 – 0.36)	-0.25 (-0.54 – 0.07) pMCMC = 0.92	-0.25 (-0.59 – 0.05) pMCMC = 0.91
Agg	**0.17 (0.03 – 0.36)** **pMCMC = 0.01**	0.03 (0.003 – 0.08)	-0.06 (-0.32 – 0.40) pMCMC = 0.50
PR	**-0.14 (-0.31 – 0.02)** **pMCMC = 0.96**	-0.16 (-0.30 - 0.04) pMCMC = 0.94	0.01 (0.001 – 0.03)

3.4.5. Mark-recapture results

None of the three behaviors had a significant effect on survival coefficients (Table A.5.), and the model with no behavioral terms had the lowest AICc score (Table A.6.). The same was found when investigating the presence of stabilizing selection on the behaviors, where the model with no quadratic terms had the best fit (Table A.7., Table A.8.). However, the model without behaviors did not statistically differ from one including a quadratic term for aggression (AICc < 2 points away from the reduced model), indicating the possibility of stabilizing selection on aggressive behavior.

Despite the lack of clear effects on survival, we converted the selection coefficients for the effects of behaviors on survival to Lande and Arnold's directional selection gradients (**β,** Lande and Arnold 1983) which were used to estimate vector correlations and angles with among- and within-individual eigen vectors. The **β**s were -0.67 (-1.23 – 0.45), 0.23 (-1.10– 1.19), and -

0.62 (-1.22 – 1.19) for activity, aggression, and predator cue response, respectively. This indicates that selection favored lower level of activity, but a higher level of aggression and for individuals to keep a longer distance from a predator cue.

3.4.6. Vector correlations and angles

The vector correlations between the dominant eigenvectors for the among- and within-individual covariance matrices, and with selection gradients, were low, leading to angles around 70˚ to 90˚ (Figure 3.2., Table A.9.).

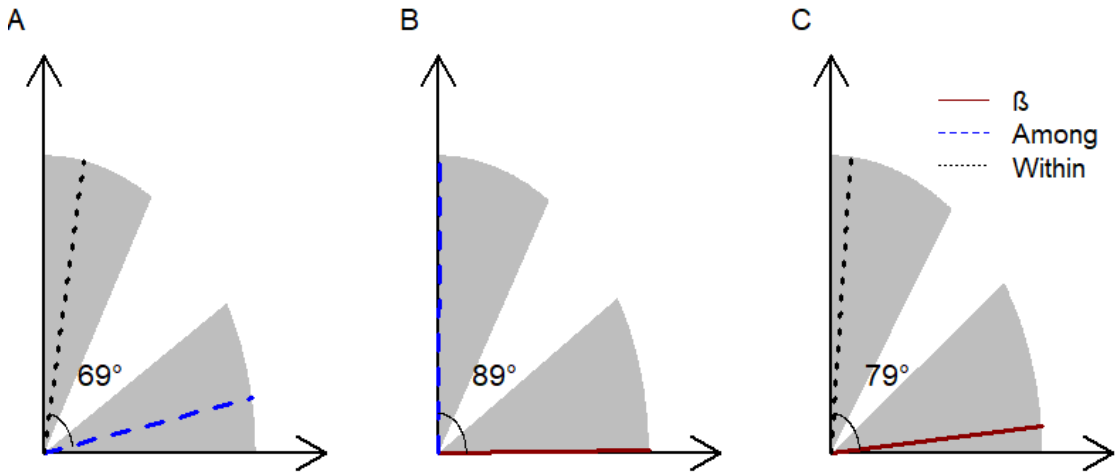

Figure 3.2. Angles between A) $\mathbf{i}_{max}$ (dashed blue line) and $\mathbf{w}_{max}$ (dotted black line), posterior HPD-inteval = 29.32 – 90.00, B) $\mathbf{i}_{max}$ and **β** (solid red line), posterior HPD-inteval = 26.50 – 89.98, and C) $\mathbf{w}_{max}$ and **β,** posterior HPD-inteval = 20.32 – 89.75. The angles between lines are based on the posterior distribution from the multi-response mixed effects model, and the shaded areas show the 95% HPD interval, where the non-shaded area indicates minimal angles between the vectors.

Using the Bayesian method for matrix comparison described by Ovaskainen et al. (2008), the three vectors were not clearly misaligned, but the probability of misalignment (i.e. proportion of estimates > 0) was quite high, particularly between $\mathbf{i}_{max}$ and $\mathbf{w}_{max}$ (Table 3.2.). Because this method has low power, we also investigated the possibility of misalignment by comparing estimates to null distributions. This alternative method indicated that most of the estimated

vector correlations were less than 0.9 (Table 3.2.), consistent with genetic variation, plasticity, and selection gradients being misaligned.

Table 3.2. ψ_θ and proportion of $\psi_\theta > 0$ for the alignment between among – and within-individual variation ($\mathbf{i}_{max}$ and $\mathbf{w}_{max}$ respectively) and selection gradient ($\boldsymbol{\beta}$), and proportion of estimates (out of 1000) that are below three different null expectations (correlation of 0.975. 0.95, and 0.90) for the vector correlation between $\mathbf{i}_{max}$, $\mathbf{w}_{max}$, and $\boldsymbol{\beta}$. Because the proportion of ψ_θ estimates > 0 was less than 0.95 none of the vectors were clearly misaligned. However, the vector correlations where substantially different from the three null expectations (except for $\mathbf{w}_{max}$:$\boldsymbol{\beta}$ compared to the null expectation of a correlation of 0.90), indicating a misalignment between the three vectors.

	ψ_θ	Estimates of $\psi_\theta > 0$	Estimates < 0 Null = 0.975	Estimates < 0 Null = 0.95	Estimates < 0 Null = 0.90
i_{max}:w_{max}	1.24 (-0.25 - 1.80)	0.90	0.99	0.98	0.96
i_{max}:$\boldsymbol{\beta}$	0.14 (-0.92 – 1.42)	0.71	0.99	0.98	0.95
w_{max}:$\boldsymbol{\beta}$	0.22 (-0.41 – 1.58)	0.81	0.97	0.96	0.93

3.5. Discussion

Contrary to our predictions we found little evidence for alignment between among-individual variation, plasticity (within-individual variation), and selection in a wild population of deer mice. Most estimates of alignment were greater than 0 (0.71– 0.90, Table 3.2.), indicating matrix differences. In addition, the results from the Z-transformation analysis indicated that all the three vectors have a vector correlation substantially different from 1, i.e., they are misaligned (Table 3.2.). A misalignment between among-individual variation and the selection gradient is predicted when there has been a change in selection pressure and the genetic architecture has not had enough time to realign with the fitness landscape (Jones et al. 2007). However, because phenotypic plasticity can respond quicker to selection pressure (Gavrilets and Scheiner 1993), plasticity is still predicted to align with selection. This was not the case in our study system, and there are at least three reasons for the apparent misalignment between the selection gradient and both among- and within-individual variation.

First, the estimates of the selection gradient might be biased. Developmental stage had an effect on recapture probability, with adult individuals having a higher chance of being recaptured than juveniles. Juvenile deer mice have a higher rate of dispersal at the end of the breeding season (Fairbairn 1978), which might be why recapture probability was lower for juveniles in our system. One of the assumptions of the mark-recapture model used is that we were monitoring a closed population, meaning that an individual that has not been recaptured for a long time is assumed to be dead. However, this was not necessarily the case for our population, and deer mice were able to migrate out of our trapping location, which means that there might be some biases in the survival estimates. Unfortunately, given the structure of our data, this assumption was necessary to estimate the effect of behaviors on survival.

Second, the behaviors measured here might be under weak or no directional selection. Consistent with this, only mass had a substantive effect on survival probability, with larger individuals having a greater chance of surviving compared to smaller individuals. Including behavior in the mark-recapture model did not improve the model's fit (Table A.6.). This indicates that the behavioral variables had only small (or no) effects on survival probability, and that they are not under strong directional selection. If selection pressure is weak it might not be strong enough to have shaped **G** to align with selection. Thus, the behaviors measured here might not have been appropriate for this population of deer mice. However, the deer mice increased their time spent on the mirror side of the arena in the mirror image stimulation test compared to the open field test (Figure A.3.A.), and likewise increased their distance from the predator cue compared to the same corner in the open field test (Figure A.3.B.). This indicates that the deer mice did respond to these cues in the predicted manner and validates the use of these assays and behaviors in this study. All behaviours were also repeatable, meaning that deer

mice were consistent in their responses to these cues as well as in their activity level. In addition, previous studies have shown that there is additive genetic variation for distance moved (Careau et al. 2011) and underlying genes influencing aggressive behavior (Shorter et al. 2014) for *P. maniculatus*, further supporting the suitability of these methods to record the behaviors used in this study.

While the most parsimonious explanation for our results is that these behaviors are not under selection, such a finding would be particularly surprising. The behaviors we measured are frequently associated with fitness across taxa (Smith and Blumstein 2008). Exploration, response to cues of predator presence, and mirror image stimulation are also frequently argued to be particularly ecologically meaningful (Dingemanse et al. 2004, Shonfield et al. 2012, Haage et al. 2017, Haapakoski et al. 2018). For example, rats (*Rattus* spp.) decrease their foraging time in the presence of a fresh predator cue (Bytheway et al. 2013). Likewise, female red squirrels (*Tamiasciurus hudsonicus*) with a higher activity level were more risk-prone, which led to a lower winter survival for the female herself, but a higher survival of her offspring because the offspring could remain in their natal territory (Boon et al. 2008).

Third, the behaviors might be under stabilizing selection, be at their fitness peak, or have low genetic variation. Only traits under directional selection are predicted to align with the selection gradient (Berdal and Dochtermann 2019). If the behaviors are under stabilizing selection, the variation of the traits will be around the fitness optima and variation in the traits is predicted to be orthogonal to selection. Consistent with this, Blows et al. (2004) found that the genetic correlations among cuticular hydrocarbons in male *Drosophila serrata* were misaligned with directional selection, most likely due to a reduction in genetic variation caused by sexual selection through strong female preference. Thus, both weak directional selection and stabilizing

selection could explain the misalignment of genetic variation (here among-individual variation) and the selection gradient.

However, we could not determine whether the behaviors were under stabilizing selection based on our original *a priori* analysis. We therefore carried out the described post-hoc analysis (see Appendix A) to estimate stabilizing selection by adding quadratic terms to the mark-recapture model (Lande and Arnold 1983). These results were similar to those focusing on linear terms, and the behaviors were not under clear stabilizing selection (Table A.7., Table A.8.). Specifically, the model without any quadratic terms for behaviors had the best fit but was not distinguishable from a model including a quadratic term for aggression. The coefficient for aggression was negative (Table A.7.), indicating that aggression could be under stabilizing selection. Aggression has been found to be under stabilizing selection in both water voles (*Arvicola terrestris*) and house mice (*Mus musculus*) as a result of female choice, where females favoured males with a medium aggression level over overaggressive or docile males (Evsikov et al. 2006). Similarly, female meadow voles (*Microtus pennsylvanicus*) avoid mating with overaggressive males (Storey 1994), thus hindering selection for increased aggressiveness in male voles, i.e., aggression is not under directional selection. Sexual selection on aggression could also be the case for our population of deer mice and could reduce the alignment between among-individual variation and selection. Unfortunately, at this time, stabilizing selection cannot be clearly distinguished from no selection, due to the AIC values not being substantively differnet.

Haller and Hendry (2014) demonstrated that selection in traits that have reached their fitness peak is difficult to detect and we therefore sought to explore the possibility of stabilizing selection further. Another potential indicator of populations being under strong selection is a loss

of variation (Mousseau and Roff 1987, Houle 1992). As a second post hoc analysis we therefore calculated the mean standardized among-individual variation of the behaviors (i.e., evolvability, I_i). These values were low compared to the estimates in Hansen et al. (2011), indicating low evolvability in these behavioral traits (Table 3.1.). This observed low I_i is consistent with behaviors being close to the fitness peak for this population, which would also explain the misalignment of among- and within-individual variation with the selection gradient.

As mentioned, among- and within-individual covariation were also misaligned. The among- and within-individual correlations were low to moderate, and only the within-individual correlation between activity and aggression and activity and predator cue response were found to be substantive (pMCMC > 0.95 or < 0.05, Table 3.1.). The lack of substantial correlation for the among-individual behaviors is surprising, as this has been found in several other rodents (Dochtermann and Jenkins 2007, Taylor et al. 2012), including closely related species (*P. californicus* and *P. leucopus*, Bester-Meredith and Marler 2007). However, the other *Peromyscus* species where from laboratory colonies (Bester-Meredith and Marler, 2007) and artificial breeding for several generations could have change the behavioral correlations in a different way than natural selection would have.

The within-individual correlation between activity and predator cue response was negative, meaning individuals that reduced their activity level also increased their distance from the cue. Activity and aggression, on the other hand, were positively correlated at the within-individual level, which means that individuals that increase their activity level would also increase their aggression towards a conspecific. However, the among-individual correlation between activity and aggression was negative, indicating that individuals with higher activity levels would be less aggressive. The opposite signs for the correlations between activity and

aggression might be the main reason why we did not observe an alignment between the among- and within-individual covariance matrices. Within-individual variation includes other forms of phenotypic plasticity in addition to adaptive plasticity – e.g., passive plasticity, measurement error, and organismal error – (Whitman and Agrawal 2009, Westneat et al. 2015, Berdal and Dochtermann 2019). Only adaptive plasticity is predicted to align with the selection surface and genetic variation, while other sources of plastic variation are not (Berdal and Dochtermann 2019). Thus, the proportion of the within-individual variation that is made up by adaptive plasticity determines how well within-individual variation should align with both genetic variation and the selection surface (see worked example in Berdal and Dochtermann 2019). Here, our finding of misalignment also shows the importance of separating among- and within-individual covariation as they can differ greatly, are produced by different underlying causes, and have different implications (Dingemanse et al. 2010, Dingemanse and Dochtermann 2013, Berdal and Dochtermann 2019).

Several models have examined how genetic and plastic variation is shaped by selection, as well as how they affect each other (Gavrilets and Scheiner 1993, Jones et al. 2003, Jones et al. 2004, Draghi and Whitlock 2012), but few empirical studies have addressed this topic (but see Lind et al. 2015, Johansson et al. 2020). Here, we found no evidence for alignment among these attributes in this wild population of deer mice. The lack of alignment combined with weak or no selection on these behaviors and the low evolvabilities suggests the possibility that the behaviours measured here might be close to their fitness optimum. Furthermore, the low evolvabilities indicate that this population has very low potential of responding to a change in selection pressure. Subjecting a sample of this population to a new selection regime in a lab setting would provide a better idea of the adaptive potential of this population and whether there

is enough phenotypic plasticity to overcome the apparent lack of genetic variation suggested by the low evolvability of each trait.

4. NO COST OR GENETIC VARIANCE FOR BEHAVIORAL PLASTICITY IN CRICKETS

4.1. Abstract

Plasticity has been assumed to have costs because no individual within a population typically displays the entire range of a trait observed within the population. However, the costs of plasticity are often conflated with the costs of a phenotype, an important distinction because the costs of plasticity compares the trade-off of plasticity and fitness while cost of phenotype measures the cost of expressing one phenotype over another within the same environment. These costs have very different impacts on how selection shapes a phenotype. In this study we used 23 isogenic lines of *Gryllodes sigillatus* to investigate the costs of plasticity in response to environmental cues on reproductive success. We also estimated the inter- and intra-genetic variation among the lines in plasticity, mass, and reproduction. The former indicates whether there is variation in plasticity among the genotypes, i.e., if there is potential for plasticity to undergo evolutionary change. The latter shows whether there are differences in variability among lines. We found no evidence for costs of plasticity of the fitness traits measured here. In addition, there was no inter-genotypic variation in plasticity for either males or females. However, we did find differences in intra-genetic variation in plasticity among lines for females, meaning that some genotypes produced more variation in their responses than others. Previous selection pressure might have optimized plasticity, removing both variation in plasticity among the lines as well as reducing cost of plasticity. The differences in intra-genetic variation among lines could be a bet-hedging strategy to increase variability among individuals to increase the probability that some of them will have a phenotype that better match the environment, since the

lack of inter-genotypic variation in plasticity means there is no longer the potential for evolutionary change in plasticity.

4.2. Introduction

Phenotypic plasticity allows organisms to match their phenotype to different environments and is considered to be adaptive if this phenotypic change increases an individual's fitness (Ghalambor et al. 2007). As for any other trait, genetic variance is needed for the trait to be able to respond to selection. For plasticity this would mean that individuals differ in their ability to respond to environmental cues, another way of describing gene-by-environment interactions (GxE). Put another way, genotypes within the same population will have different reaction norm slopes across environments (Figure 4.1., Dingemanse et al. 2010). Several studies have found that individuals within the same population differ in how plastic they are for certain traits (Mathot et al. 2011, Morand-Ferron et al. 2011). For example, plasticity has been shown to respond to selection in several traits, including thorax size in *Drosophila melanogaster* (Scheiner and Lyman 1991*)* and timing of reproduction in great tits, *Parus major*, (Nussey et al. 2005). Phenotypic plasticity is therefore evolvable and is an important factor for a population's evolutionary trajectories, i.e., the evolutionary changes of a trait over time (Stearns and Koella 1986, Draghi and Whitlock 2012, Fischer et al. 2016).

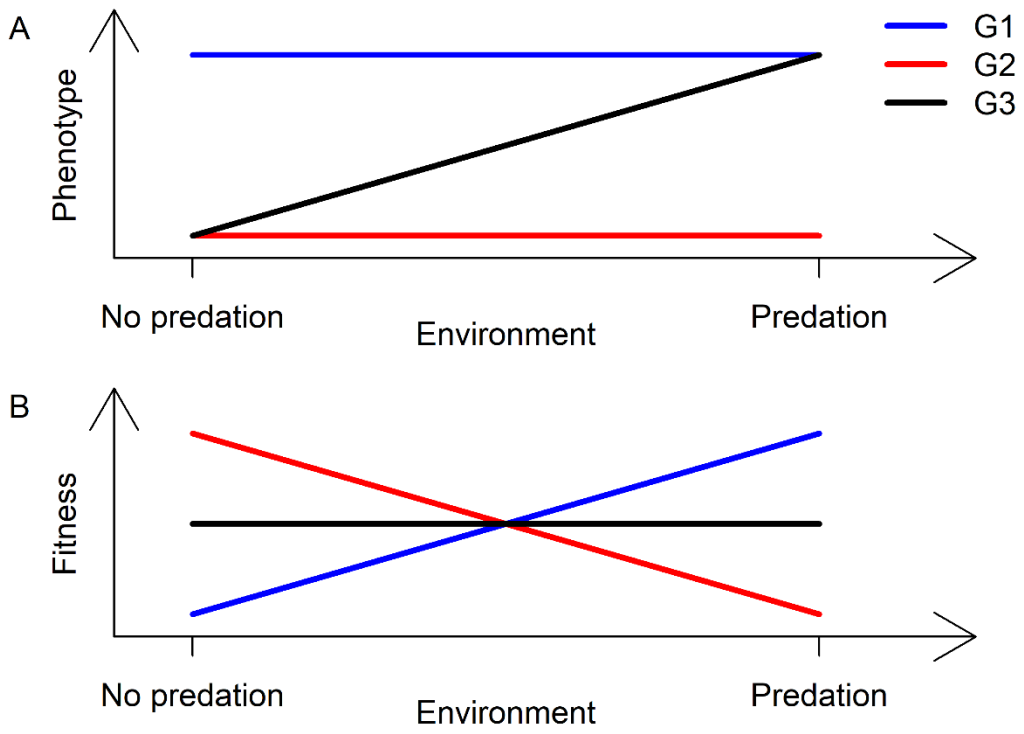

Figure 4.1. Differences in cost of phenotype and cost of plasticity A) G1 (blue line) and G2 (red line) are fixed genotypes and the expressed phenotype is the same regardless of the environment (absence vs. presence of a predator). G3 (black line) is a plastic genotype and the phenotypic expression changes according to the environment the genotype is expressed in. B) The fitness of G1 and G2 depends on which environment they are expressed in, where fitness will be higher in an environment where the phenotype matches. The difference between the red and the blue line measures the cost of phenotype. The plastic phenotype matches both environment and will have the same fitness in both. However, it will have a lower fitness than the optimized fixed phenotype due to the costs of being plastic. The difference between the red and the black line in an environment without predators and the difference between the blue and the black line in an environment with predators measures the fitness cost of plasticity.

The observation that individuals differ in their degree of plasticity (Mathot et al. 2011, Morand-Ferron et al. 2011) means that no individual expresses the full range of a trait observed in a population, suggesting that there might be limits and/or costs to plasticity (DeWitt et al. 1998). Often, costs of plasticity have been measured by comparing a trait in individuals raised in the presence of an environmental cue (e.g., predation risk or population density) with individuals that have developed in the absence of the cue (Scheiner and Berrigan 1998). Measure of fitness are then compared between the groups. For example, in *Dahpnia pulex* individuals develop a

defensive neck spine if they are raised in the presence of predator cues. Individuals that develop these spines produce fewer offspring, develop at a smaller size at maturity, and lower intrinsic growth rate compared to individuals that developed without the presence of the cue (Ketola and Vuorinen 1989, Walls and Ketola 1989, Black and Dodson 1990). However, this comparison measures the cost of phenotype, not the actual cost of plasticity (Murren et al. 2015). That is, costs of plasticity are conflated with the costs of actually developing the defensive spine. This is an important distinction because even if an individual does not need to change the trait because it stays in a constant environment, the trait still has the potential to change, meaning that the cost of the plastic machinery has already been paid. The cost of plasticity should therefore be considered as the cost of having a trait with the *potential* to vary with the environment as opposed to being fixed (Van Buskirk and Steiner 2009).

To properly estimate these costs of plasticity, it is therefore necessary to compare variation in potential plasticity to variation in fitness (Van Tienderen 1991, DeWitt et al. 1998, Van Buskirk and Steiner 2009). Further, to measure variation in plasticity the same genotype needs to be repeatedly measured across different environments. This allows for the estimation of magnitude of plasticity for each genotype, or, put another way, different reaction norm slopes within a population (Figure 4.1.). Repeated measures of the same genotype in the same and in different environment can be achieved using organisms with known pedigrees, like sib-families, or by using iso-genic lines (Van Tienderen 1991, Van Buskirk and Steiner 2009). The use of isogenic lines is especially powerful because individuals within a line will be genetically identical, and the only difference in trait expression should be caused by differences in the environment, i.e., plastic responses. For example, Scheiner and Berrigan (1998) used isogenic lines of *Daphnia spp.* where individuals from the same line were divided into two groups:

developing in the presence or absence of a predatory cue. The absolute difference in trait expression (spine length, mass, and time to maturity) between the two groups within the same line was used as a proxy for plasticity while clutch size and intrinsic rate of increase were used as fitness measurements. Both plasticity and the mean measurements of the traits for each line in each environment were regressed on fitness, where the trait mean refers to the phenotypical cost of the trait. No cost of plasticity was found in this case, a finding that seems to be common when investigating plasticity costs. In a meta-analysis examining only those studies that estimated costs of plasticity by regressing fitness over plasticity, Van Buskirk and Steiner (2009) found that actual costs of plasticity were most often weak. However, they also found that these costs varied broadly across studies and that increased stress had the potential of increasing the chance of detecting a cost. This indicates that a cost of plasticity might only be present under highly stressful situations.

Cost of plasticity is in most cases measured in trait with developmental plasticity, i.e., traits that responds to a cue early in life and stays the same for the rest of that organism's lifespan, i.e., irreversible plasticity. This includes neck spine in daphnia (Krueger and Dodson 1981), birdsong in some species of birds (Marler and Peters 1987), and tail development in tadpoles (Van Buskirk 2000). However, some traits can be changed throughout an organism's lifetime, i.e., reversible plasticity. Irreversible and reversible plasticity, also named developmental and activational plasticity respectively for behavior in a review by Snell-Rood (2013), are expected to have different costs associated with them because the underlying mechanisms for detecting cues and responding to these cues differ. Traits with developmental plasticity only respond to a cue at one time in an organism's lifetime and will not be responsive to any cues after the trait has been developed. These traits can possibly be more integrated as

several traits will develop in response to the cue at once, leading to increased correlations among traits. Activational plasticity, on the other hand, respond to their environment constantly, which means that the neural networks to detect environmental cues and respond appropriate to these cues are always activated. Activational plasticity might therefore have a higher cost than developmental plasticity due to the need to maintain larger neural network over an individual's lifetime rather than just during development (Snell-Rood 2013).

Another example of both types of plasticity is observed in individuals from the same isogenic line raised in the same environment conditions, but differ in their trait expression, with some genotypes being more variable than others (Stamps et al. 2013). This variation, called intra-genotypic variation (IGV), might be the result of bet-hedging or organismal error. For bet-hedging, genotypes producing different phenotypes based on random expression processes rather than as a response to the environment. The different expressed phenotypes might either be fixed within individuals throughout their lifetime or vary across time (Stamps et al. 2013, Ayroles et al. 2015). Both increases the variability of the phenotype, but at different levels, i.e., within an individual or among individuals within a population. Organismal error also increases the variability of the genotype, but where bet-hedging increases the variability to increase the chance that some of the expressed phenotypes will match the environment and benefit from an increase in fitness, organismal error will result in a mismatch between the expressed phenotype and the environment and should be selected against.

In this study we used 23 isogenic populations of *Gryllodes sigillatus* to investigate cost of plasticity on reproductive success. Because this species is not clonal in nature, like for example daphnia, this system lets us use the power of iso-genetic lines in a species where this is usually not possible. We focused on a behavior that expresses reversible plasticity. As discussed above,

costs of reversible plasticity are expected to be greater than as observed for irreversible plasticity, making these costs easier to detect. Plasticity was measured as absolute change in activity in the absence and presence of a predator cue and reproductive success was measured as spermatophore mass in males and egg count in females. If plasticity is costly, we would predict a negative correlation between magnitude of plastic response and reproductive output among lines. We also investigated the potential for evolution in this species by seeing if there was significant among-line variation in plasticity, as well as differences in within-line variance.

4.3. Methods

4.3.1. Breeding

The crickets used for this experiment were from nine lines of crickets that have been inbred through full sibling crossings for 20 generations (Ivy 2005), which results in an inbreeding coefficient of 0.986 (Falconer and Mackay 1996). The lines have been allowed to continue to breed within-lines for another 55-60 generations after the initial crosses, leading to an even higher inbreeding coefficient. Individuals from the same line are therefore close to genetically identical, making it possible to investigate genetic differences in plasticity among lines. Eight of the inbred lines were bred together within lines and among lines creating 23 genetically distinct lines. Two females from one line and two males from the same line or a different line was put in a box (34.6 x 21 x 12.4 cm) where they were provided with water, chick feed, egg carton for shelter, and an oviposition cup filled with peat moss. The boxes were kept at around 26°C. The oviposition cup was checked every other day to make sure it was kept damp and to see whether there were any hatchlings. As soon as hatchlings were observed the parents were removed.

4.3.2. Rearing

Boxes with hatchlings were checked every other day to ensure that the soil was damp and that the hatchlings had access to food and water. As the hatchlings started to mature the boxes were checked every day, and adults were put into individual cups (11.5cm in diameter). This ensured that all the individuals used in the study were virgins. Each individual box was provided with a water vial, a small dish with chick feed, and egg carton for shelter, and was kept at around 26°C. Adult crickets were then run through behavioral trials.

4.3.3. Behavioral tests

Plasticity in activity was measured by introducing crickets to two different environments: environments with and without the presence of a predatory cue. Other species of crickets have shown response to chemical cues from predators fed on crickets (Hoefler et al. 2012) and the predatory cue used here consisted of a filter paper soaked in diluted excreta collected from leopard geckos (*Eublepharis macularius*) fed crickets. The filter paper was put in the bottom of a round arena (15 cm in diameter). The test without the predatory cue was set up in the same way, but with a filter paper soaked in distilled water. Whether a cricket was run through the trial with predatory cues or control cues first was randomized to control for any order effect.

Four crickets were tested at the same time in separate arenas. At the start of the trial the crickets were placed under a plastic cup in a hole cut in the middle of the filter papers for 30 seconds to acclimate. This way the crickets would not interact with the cues until the start of the trial. The cups were then removed, and the crickets had 3 minutes and 40 seconds to explore the arena. After the trial, the crickets were weighted and then put back into their individual housing.

All trials were recorded, and the videos were analyzed using Ethovision version 10. Tracking was started after 40 seconds to make sure that the observer had time to release all the

crickets and place the lid over the arena, and thus avoid any disturbance during video analysis. Activity level in both trials were recorded as distance moved (in cm), and plasticity was calculated as the absolute difference in distance moved between the two trials. Previous studies have shown that *G. sigillatus* increase their activity level in the presence of a predatory cue (unpublished data) and that activity level is higher for individuals that have previously been in contact with predators (Bucklaew and Dochtermann 2021), which could therefore be an escape response. Other species of closely related crickets also increase their activity levels in the presence of these predator cues (Royauté and Dochtermann 2017, Royauté et al. 2019). However, crickets have also been observed to freeze in the presence of predators (Hedrick and Kortet 2012) and using absolute values will therefore capture the plasticity in both directions.

4.3.4. Reproductive measurements

After both behavioral trials were complete, reproductive measurements were recorded for both males and females. Male reproduction was measured as spermatophore mass, which was sampled by holding the cricket between the thumb and index finger, and gently pressing the thumb on the abdomen to release the spermatophore. A probe was used to transport the spermatophore to a weighing paper, and the spermatophore was weighted to the nearest milligram. If a male had no spermatophore it was noted as 0 mg. The mass of males was also recorded.

Female reproduction was measured as number of eggs laid. Females were mated with males from the same line that were not otherwise used in the behavioural trials to avoid any effects of stress due to handling. Because some females lay eggs in the water vials, all water vials were changed before the males were added. The males had their mass recorded and then were added to the female's home container for 24 hours. The males were then removed, and an

oviposition cup with sand was added to the female's box. The females were given three days to lay eggs before the oviposition cup was removed. The cups were covered in cling film and stored in a fridge at ~4 °C. The eggs were counted by mixing the sand with water and pouring the water into a petri dish. Because the eggs are lighter than the sand the eggs would be poured out with the water and could easily be counted.

4.3.5. Statistical analysis

4.3.5.1. Trait correlations within and between sexes

The correlations between plasticity, mass, and fitness were estimated using linear mixed effect models using the nlme packaged in R (Pinheiro et al. 2017), and correlations were estimated piecewise in separate bivariate models. Among-line correlations within the sexes are expected to be negative, as this would indicate a trade-off between fitness, growth, and the plastic machinery. Similarly, if the underlying genetics are similar for the sexes, the among-line correlations between sexes are predicted to be negative, also indicating trade-offs. The within-sex correlation within-line is predicted to be positive, as individuals within a line would be genetically identical and the only difference between them would be due to the environment. Thus, any variance within a line would be due to unmeasurable micro-environmental differences.

For between-sex correlations, only among-line correlations can be estimated. Bivariate response models were fitted by using two of the traits (e.g., fitness and plasticity) as response variables and trait type as a fixed effect in combination with sex, resulting in a factor with four levels (e.g., female fitness, male fitness, female plasticity, and male plasticity). For the random effect, this factor was nested within line allowing for random slopes for each sex and trait type. This allows us to use a univariate model to estimate multivariate parameters, like among-line correlations.

A model without the among-line correlations were run to allow an AIC comparison to see if the correlations were substantial. The correlations between sexes for the same trait were qualitatively compared to one, indicating whether or not the trait should be considered to be the same for both sexes (correlation ≈ 1) or to be two different traits (correlation < 1).

The models estimating the within-sex correlations were run in a similar way but allowed for within-line correlations (full model), since this was estimable here. Additional models allowing for only among-line correlation, only within-line correlation, or no correlation (null model) were run to allow AIC comparison to see whether the among- and/or within-line correlations were substantial. Models different from the null model with an AIC greater than 2 indicate substantively better fit of this model, i.e., the correlation(s) were substantial. If a model including the among-line correlations and the full model had a significantly better fit than the null model, but the model with only the within-line correlation did not, only the among-line correlation was considered substantial, and vice versa.

4.3.5.2. Among- and within-line variation

The inter- and intra-genotypic (among- and within-line) variation for the three traits were estimated for each sex separately and was done by comparing four univariate models of increasing complexity. The null model was a linear model without any random effects, thus not allowing for any variance differences either among or within lines. The second model was a mixed effect model with line as the random effect, which allows for among-line variance. The third model was a linear model without line as random effect, but that did allow for heterogeneous residuals, i.e., differences in within-line variation. Finally, the full model was a mixed effect model with line as a random effect and allowing heterogeneous residuals. Similar to the model comparisons for correlations, models that had substantively better fit than the null

model indicated substantial effect of the variance partitioning. Again, the model with only among-line variance must have a substantively better fit than the null model for the variance to be considered to have a substantial effect. If the full model has a significantly better fit than the null model, but the model only allowing for among-line variance does not, then the among-line variance is not considered substantial, even though the full model includes among-line variance. The same goes for the model only allowing for within-line variance.

4.4. Results

4.4.1. Trait correlations within and between sexes

No cost of plasticity was found for either males or females as none of the among-line correlations between fitness and plasticity were negative (Figure 4.1.). However, a strong positive correlation was found within males, indicating that genotypes with a higher level of plasticity would also produce larger spermatophores (Figure 4.2., Table 4.1.-4.2.). This relationship held even after controlling for an isometric relationship between fitness and mass. In addition, the correlation between fitness and mass was significant among- and within line for males and within-lines for females. This means that males with a genotype for increased size will also produce a larger spermatophore, and that larger individuals within a line will also produce larger spermatophores (or lay more eggs). Only two between-sex correlations were significant, namely between male fitness and either female plasticity or female mass (Figure 4.1., Table 4.1.-4.2.).

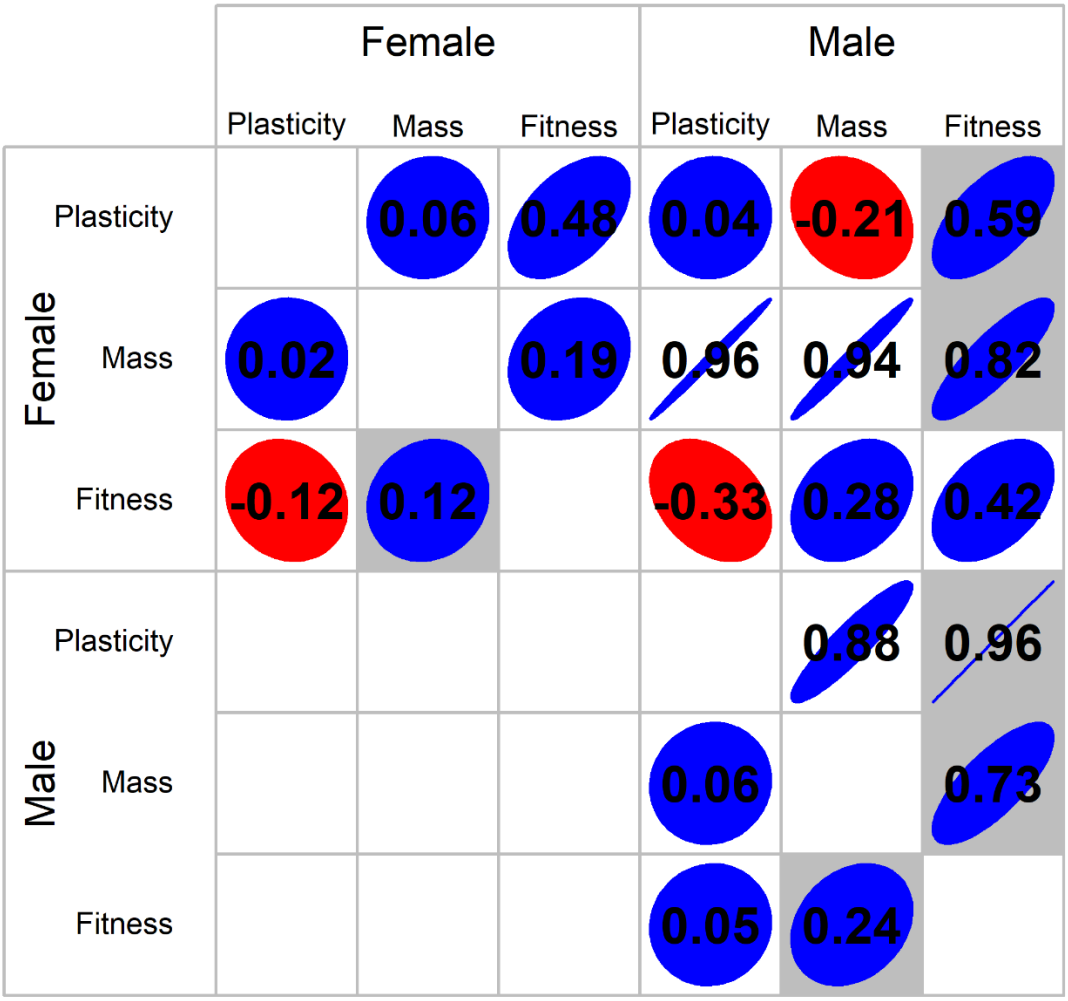

Figure 4.2. Correlations between plasticity, mass, and fitness in G. sigillatus among and within the two sexes. The values above the diagonal are among-line correlations and below the diagonal are within-line correlations. Substantial correlations are shown with a grey background.

Table 4.1. Model comparisons to investigate whether the among-line correlations between plasticity, mass, and fitness between the sexes were substantial. ΔAIC values are calculated by subtracting the AIC values of the model with correlations estimated from the AIC values of the model without correlation estimated. ΔAIC < -2 indicate substantial correlations. Models with substantially better fit are shown in bold. For all models with correlation K = 7 and for all models without correlation K = 6.

	AIC values		
Model	**With correlations**	**Without correlations**	**ΔAIC**
Female fitness x male fitness	6304.13	6302.37	1.76
Female mass x male mass	4955.10	4954.83	0.27
Female plasticity x male plasticity	6265.53	6265.24	0.29
Female fitness x male plasticity	5981.59	5981.27	0.32
Female mass x male fitness	**6013.60**	6015.61	-2.01
Female plasticity x male mass	5539.91	5540.85	-0.94
Female fitness x male mass	5255.83	5254.45	1.38
Female mass x male plasticity	5692.53	5690.58	1.95
Female plasticity x male fitness	**6587.07**	6591.95	-4.88

Table 4.2. Model comparisons to investigate whether the among- and within-line correlations between plasticity, mass, and fitness within the sexes were substantial. Null models are fitted without any correlations, "only among" models include only the among-line correlations, "only within" models include only within-line correlations, and full models are fitted with both correlations. Models with substantially better fit than the null model (ΔAIC > 2) are shown in bold.

Within female							
Model		**Fitness x Mass**		**Fitness x Plasticity**		**Plasticity x Mass**	
	df	**AIC**	**ΔAIC**	**AIC**	**ΔAIC**	**AIC**	**ΔAIC**
Full	8	5552.55	0.59	6121.39	1.99	5839.97	2.92
Only among	7	5553.73	1.77	6122.90	3.50	5838.03	0.98
Only within	7	**5551.96**	**0**	6119.40	0	5839.04	1.99
Null	6	5554.74	2.78	6121.08	1.68	5837.05	0
Within male							
	df	**AIC**	**ΔAIC**	**AIC**	**ΔAIC**	**AIC**	**ΔAIC**
Full	8	**5702.07**	**0**	6443.27	1.16	5396.74	2.62
Only among	7	5712.90	10.83	**6442.11**	**0**	5395.51	1.39
Only within	7	5716.47	14.4	6450.58	8.47	5395.45	1.33
Null	6	5730.32	28.25	6449.09	6.98	5394.12	0

4.4.2. Among- and within-line variation

Female fitness had significant intra- and inter genetic variance, meaning that there is a difference in both mean fitness and within-line variance among lines (Figure 4.3., Table 4.3.).

For male fitness, male mass, and female mass only among-line variance was significant, meaning there are mean differences among lines, but the variance within line is similar. Neither among- or within-line variance was significant for male plasticity, but within-line variance was significant for female variance, indicating that mean plasticity is similar across lines, but that the within-line variances differ among lines for females (Figure 4.3., Table 4.3.).

Table 4.3. Model comparison to investigate whether the among- and within-line variation was substantial. For null models, the variances among and within lines were equal. For "only among" models the variance among lines were allowed to vary, and for "only within" models the within-line variances were allowed to vary, while both among-and within line variance was allowed to vary for full models. Models with substantially better fit than the null model ($\Delta AIC > 2$) are shown in bold.

Within female								
Model			**Fitness**		**Mass**		**Plasticity**	
	Variance	**df**	**AIC**	**ΔAIC**	**AIC**	**ΔAIC**	**AIC**	**ΔAIC**
Null	$V_A =, Vw =$	2	2919.18	3.24	3645.46	11.56	3416.27	7.77
Only among	$V_A \neq, Vw =$	3	2918.67	2.73	**3633.56**	**0**	3417.74	9.24
Only within	$V_A =, Vw \neq$	24	2927.86	11.92	2661.99	28.43	**3408.50**	**0**
Full	$V_A \neq, Vw \neq$	25	**2915.94**	**0**	2654.95	21.39	3408.74	0.27
Within male								
	Variance	**df**	**AIC**	**ΔAIC**	**AIC**	**ΔAIC**	**AIC**	**ΔAIC**
Null	$V_A =, Vw =$	2	3392.59	8.23	2425.61	89.88	**3265.94**	**0**
Only among	$V_A \neq, Vw =$	3	**3384.36**	**0**	**2335.73**	**0**	3267.95	2.01
Only within	$V_A =, Vw \neq$	24	3415.95	31.59	2402.87	67.14	3279.44	13.50
Full	$V_A \neq, Vw \neq$	25	3396.99	12.63	2355.45	19.72	3281.45	15.51

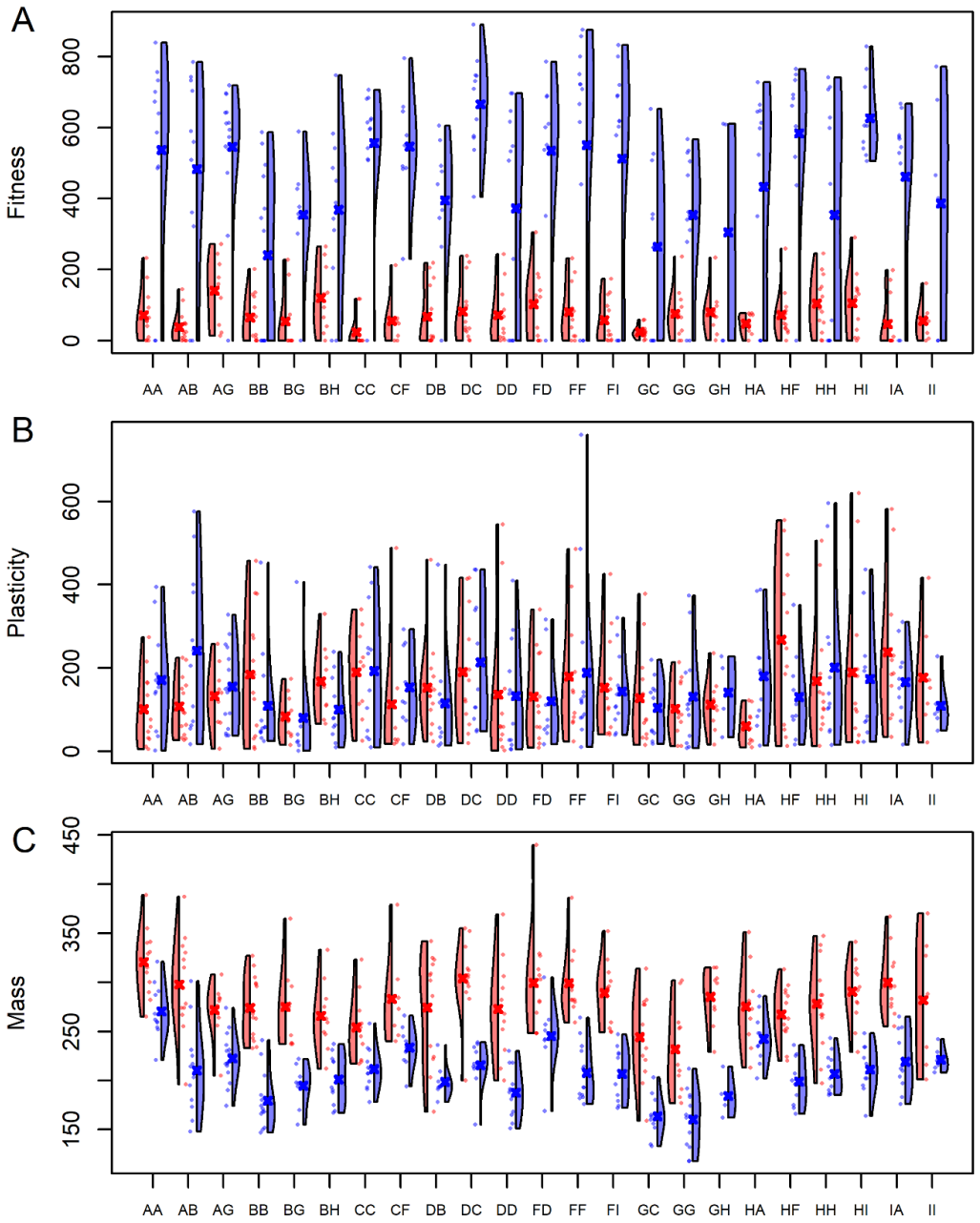

Figure 4.3. Raincloud plots showing the inter- and intra-genetic variation for A) fitness, B) plasticity, and C) mass for males (blue) and females (red) for each line. Because the fitness measures differed for the two sexes (spermatophore mass for males and egg count for females) the mean fitness for male is higher than for females. There is also a significant difference in fitness among the lines for both males and females, and the intra-genetic variance is significant for females. No among-line differences were found for plasticity, though again there was a significant intra-genetic variance for females. Lastly, females were significantly larger than males across all lines and among-line variance was significant for both sexes.

4.5. Discussion

We found no evidence for cost of plasticity in the cricket *G. sigillatus*, i.e., we did not find a negative correlation between fitness and plasticity among lines (Figure 4.1., Table 4.1.-4.2.). We did, however, find a strong significant positive correlation between fitness and plasticity among lines for males, which means that males from lines that produced larger spermatophores also showed higher levels of plasticity. This is contrary to our predictions that individuals investing in higher levels of plasticity will not have as many resources to invest in reproduction. Similarly, we found a positive correlation between fitness and mass among lines for males, again contrary to our predictions of a trade-off. One reason for this lack of trade-off between plasticity, fitness, and mass is that the trade-off is with a trait not measured here, for example longevity. The lack of an observed trade-off could also be a result of variation in acquisition and allocation of resources between genotypes (Van Noordwijk and de Jong 1986). This means that some genotypes have a higher efficiency at acquisition of resources which leads to them being able to put more energy into several traits at the same time, thus masking the trade-off between investment in a specific trait and fitness (Van Noordwijk and de Jong 1986). This has been shown to be the case for several species, for example seed beetles (*Callosobruchus maculatus*), where a positive correlation was observed between fecundity and longevity when seed availability was high (Messina 2003). However, when seeds were scarce, the correlation became negative. The trade-off between fecundity and longevity was therefore only observed during stressful conditions, while during seed abundance some individuals were able to more allocate energy to both fecundity and longevity, masking the trade-off (Messina 2003).

The within-line correlation between fitness and mass was found to be significant and positive for both males and females (Figure 4.1., Table 4.1.-4.2.). Because the lines are

genetically identical and are raised in the same environment, the variation within lines is caused by differences in the unmeasurable micro-environment, which would in this case be slight differences in temperature and possible accessibility to food and water. Individuals within the same line that were, for example, housed in a location with a slightly higher temperature would have a benefit over individuals of the same genotype the were exposed to slightly lower temperatures. The positive correlation observed here is therefore as predicted, as individuals of the same genotype with more resources would be able to invest more in both reproduction and growth at the same time.

Cost of plasticity might be assumed to be low if previous selection has reduced the costs (Murren et al. 2015). If that is the case plasticity might be hard to detect. However, activational plasticity is predicted to remain costly due to the need to maintain a large neural network to constantly detect environmental cues (Snell-Rood 2013). This was the basis of focusing on plasticity in a behavioral response towards a predator cue here. However, since detection of predators is likely highly beneficial, previous selection pressure might have consistently favored plasticity in this trait and reduced associated costs.

The correlations between sexes among lines for the same trait varied among the three traits measured here (Figure 4.1.). The between-sex correlation for mass was close to 1 (0.944), which indicates that mass can be considered the same trait for both sexes, i.e., it has the same underlying genetic architecture in both sexes. The between-sex correlation for plasticity, on the other hand, was close to 0 (0.038). There is therefore no apparent relationship between the genes for plasticity for the two sexes, and plasticity should therefore be considered different traits for males and females. Lastly, the between-sex correlation for fitness was intermediate between 0 and 1 (0.417), which means that there are some shared genetics between the sexes for fitness, but

that there are also sex-specific genes controlling this trait. This is not surprising as fitness was measured differently for the two sexes (egg count for females and spermatophore mass for males). However, both traits demand high investment and would therefore have the potential for a trade-off with other traits that demand investments, like growth and plasticity, which was why they were used as fitness measures in this study. If the selection pressure for a trait differs between the two sexes, a high between-sex correlation will lead to sexual discordant selection and could impact evolutionary trajectories (Westneat and Sih 2009). For example, if males in our population were selected to have smaller mass, while females were selected to be larger, only the sex with the strongest selection pressure would reach its optimum, and the time to reach the optimum would be slowed down due to the opposite selection pressure in the other sex (Westneat and Sih 2009). If the selection pressure differs between the sexes for plasticity, on the other hand, the trait could evolve freely in both sexes due to the low between-sex correlation.

The only cross-trait correlation between sexes that were significant were for male fitness with both female plasticity and female mass, which were both positively correlated (Figure 4.1., Table 4.1.-4.2.). This means in lines where males produce larger spermatophores, the females would be larger and exhibit a higher level of plasticity. This is, again, contrary to our prediction of a trade-off between these traits, though this prediction depended on the traits being controlled by the same underlying genetics for the two sexes, which in this case was only true for mass. If the traits were controlled by the same underlying genetics for both sexes, a trade-off between fitness, mass, and plasticity is predicted to be the same between sexes as within sexes among lines. However, due to the low cross-sex correlation fitness can be considered a semi-independent trait and plasticity a completely independent trait for the two sexes the trade-off

between fitness and plasticity between the two sexes is no longer predicted as these traits are controlled by different genes.

Both fitness and mass were found to have significant inter-genotypic variation for both males and females, meaning that there were genetic differences among lines in mean level for these two traits (Figure 4.2., Table 4.3.). For female fitness there was also significant intra-genotypic variation, which means that some genotypes were more variable in fitness output than others. Significant intra-genotypic variation was also found for plasticity in females, though inter-genotypic variation was not significant. Thus, there was no difference in mean level of plasticity among the lines, but some lines showed more variation in their responses than others. The intra-genotypic variation could be a result of bet-hedging, where some genotypes produce a wider range of phenotypes based on gene expression process, like epigenetic markers and maternal effects, rather than as a response to an environmental cue (Stamps et al. 2013, Ayroles et al. 2015). This increases the probability of producing a phenotype that will better match the environment, especially in a habitat where environmental cues are less reliable. Neither inter- nor intra-genotypic variation were found for plasticity in males.

The lack of evidence for cost of plasticity and inter-genetic variation in plasticity among lines could be caused by prior selection pressure that has optimized plasticity in this population, reducing the variance and any associated cost of plasticity (Murren et al. 2015). Another possibility for the lack of cost of plasticity is that the olfactory machinery is used in other contexts, like foraging, and is therefore already being maintained. Being able to detect predators might not have any additional cost to the perseverance of the olfactory machinery, or only a small additional cost, which would be hard to detect.

Because there was no genetic variation in plasticity there can be no evolutionary changes in this trait. The intra-genetic variation for plasticity in females might therefore be a bet-hedging strategy to increase variability among individuals increasing the chance that some of them might develop a phenotypic response that better matches the environment they live in. Females might have a greater benefit from this variability than males, as females must move through different environments in their search for a mate, while males stay in one place, often close to a burrow, where they call to attract females (Sakaluk 1987). The male strategy will therefore be the same every time a predatory cue is detected, namely run and hide in the burrow, while female's strategy depends on which habitat she is in. A futures study on the benefit of intra-genotypic variation in plasticity, and whether this benefit is greater for female than male crickets could increase our understanding for the evolutionary consequences for intra-genetic variation.

REFERENCES

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

Animal Behaviour Society. 2020. Guidelines for the treatment of animals in behavioural research and teaching. Animal Behaviour **159**:I - XI.

Armbruster, W. S. and K. E. Schwaegerle. 1996. Causes of covariation of phenotypic traits among populations. Journal of Evolutionary Biology **9**:261-276.

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

Ayroles, J. F., S. M. Buchanan, C. O'Leary, K. Skutt-Kakaria, J. K. Grenier, A. G. Clark, D. L. Hartl, and B. L. De Bivort. 2015. Behavioral idiosyncrasy reveals genetic control of phenotypic variability. Proceedings of the National Academy of Sciences **112**:6706-6711.

Berdal, M. A. and N. A. Dochtermann. 2019. Adaptive alignment of plasticity with genetic variation and selection. Journal of Heredity **110**:514-521.

Bester-Meredith, J. K. and C. A. Marler. 2007. Social experience during development and female offspring aggression in Peromyscus mice. Ethology **113**:889-900.

Black, A. R. and S. I. Dodson. 1990. Demographic costs of chaoborus-induced phenotypic plasticity in daphnia-pulex. Oecologia **83**:117-122.

Blows, M. W., S. F. Chenoweth, and E. Hine. 2004. Orientation of the genetic variance-covariance matrix and the fitness surface for multiple male sexually selected traits. The American Naturalist **163**:329-340.

Boon, A. K., D. Réale, and S. Boutin. 2008. Personality, habitat use, and their consequences for survival in North American red squirrels Tamiasciurus hudsonicus. Oikos **117**:1321-1328.

Brehm, A. M. and A. Mortelliti. 2018. Mind the trap: large-scale field experiment shows that trappability is not a proxy for personality. Animal Behaviour **142**:101-112.

Brehm, A. M., S. Tironi, and A. Mortelliti. 2020. Effects of trap confinement on personality measurements in two terrestrial rodents. Plos One **15**:e0221136.

Brinkerhoff, R. J., N. M. Haddad, and J. L. Orrock. 2005. Corridors and olfactory predator cues affect small mammal behavior. Journal of Mammalogy **86**:662-669.

Brommer, J. E. and B. Class. 2017. Phenotypic correlations capture between-individual correlations underlying behavioral syndromes. Behavioral Ecology and Sociobiology **71**:50.

Bucklaew, A. and N. A. Dochtermann. 2021. The effects of exposure to predators on personality and plasticity. Ethology **127**:158-165.

Bytheway, J. P., A. J. Carthey, and P. B. Banks. 2013. Risk vs. reward: how predators and prey respond to aging olfactory cues. Behavioral Ecology and Sociobiology **67**:715-725.

Careau, V., D. Thomas, F. Pelletier, L. Turki, F. Landry, D. Garant, and D. Réale. 2011. Genetic correlation between resting metabolic rate and exploratory behaviour in deer mice (Peromyscus maniculatus). Journal of Evolutionary Biology **24**:2153-2163.

Cheverud, J. M. 1988. A comparison of genetic and phenotypic correlations. Evolution **42**:958-968.

Conner, J. K. 2002. Genetic mechanisms of floral trait correlations in a natural population. Nature **420**:407-410.

Conner, J. K., K. Karoly, C. Stewart, V. A. Koelling, H. F. Sahli, and F. H. Shaw. 2011. Rapid independent trait evolution despite a strong pleiotropic genetic correlation. The American Naturalist **178**:429-441.

DeWitt, T. J., A. Sih, and D. S. Wilson. 1998. Costs and limits of phenotypic plasticity. Trends in Ecology & Evolution **13**:77-81.

Dingemanse, N. J., I. Barber, J. Wright, and J. E. Brommer. 2012. Quantitative genetics of behavioural reaction norms: genetic correlations between personality and behavioural plasticity vary across stickleback populations. Journal of Evolutionary Biology **25**:485-496.

Dingemanse, N. J., C. Both, P. J. Drent, and J. M. Tinbergen. 2004. Fitness consequences of avian personalities in a fluctuating environment. Proceedings of the Royal Society B-Biological Sciences **271**:847-852.

Dingemanse, N. J. and N. A. Dochtermann. 2013. Quantifying individual variation in behaviour: mixed-effect modelling approaches. Journal of Animal Ecology **82**:39-54.

Dingemanse, N. J. and N. A. Dochtermann. 2014. Individual behaviour: behavioural ecology meets quantitative genetics. Oxford University Press.

Dingemanse, N. J., A. J. Kazem, D. Réale, and J. Wright. 2010. Behavioural reaction norms: animal personality meets individual plasticity. Trends in Ecology & Evolution **25**:81-89.

Dochtermann, N. A. 2011. Testing Cheverud's conjecture for behavioral correlations and behavioral syndromes. Evolution **65**:1814-1820.

Dochtermann, N. A. and S. H. Jenkins. 2007. Behavioural syndromes in Merriam's kangaroo rats (Dipodomys merriami): a test of competing hypotheses. Proceedings of the Royal Society B-Biological Sciences **274**:2343-2349.

Dochtermann, N. A., S. H. Jenkins, M. J. Swartz, and A. C. Hargett. 2012. The roles of competition and environmental heterogeneity in the maintenance of behavioral variation and covariation. Ecology **93**:1330-1339.

Dochtermann, N. A., T. Schwab, M. Anderson Berdal, J. Dalos, and R. Royauté. 2019. The heritability of behavior: A meta-analysis. Journal of Heredity **110**:403-410.

Dochtermann, N. A., T. Schwab, and A. Sih. 2015. The contribution of additive genetic variation to personality variation: heritability of personality. Proceedings of the Royal Society B **282**:20142201.

Draghi, J. A. and M. C. Whitlock. 2012. Phenotypic plasticity facilitates mutational variance, genetic variance, and evolvability along the major axis of environmental variation. Evolution **66**:2891-2902.

Endler, J. A. 1986. Natural selection in the wild. Princeton University Press.

Evsikov, V., M. Potapov, G. Nazarova, and O. Potapova. 2006. Stabilizing function of sexual selection in relation to male aggressiveness in rodents. Page 510 *in* Doklady Biological Sciences. Springer Nature BV.

Fairbairn, D. J. 1978. Dispersal of deer mice, Peromyscus maniculatus. Oecologia **32**:171-193.

Falconer, D. and T. Mackay. 1996. Introduction into quantitative genetics. Essex: Prentice Hall.

Fischer, E. K., C. K. Ghalambor, and K. L. Hoke. 2016. Can a network approach resolve how adaptive vs nonadaptive plasticity impacts evolutionary trajectories? Oxford University Press.

Fisher, R. A. 1930. The genetical theory of natural selection. Clarendon, Oxford.

Forsman, A. 2015. Rethinking phenotypic plasticity and its consequences for individuals, populations and species. Heredity **115**:276-284.

Gavrilets, S. and S. M. Scheiner. 1993. The genetics of phenotypic plasticity .5. evolution of reaction norm shape. Journal of Evolutionary Biology **6**:31-48.
Ghalambor, C. K., J. K. McKay, S. P. Carroll, and D. N. Reznick. 2007. Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. Functional Ecology **21**:394-407.
Gibson, G. and I. Dworkin. 2004. Uncovering cryptic genetic variation. Nature Reviews Genetics **5**:681-690.
Gotthard, K. and S. Nylin. 1995. Adaptive plasticity and plasticity as an adaptation: a selective review of plasticity in animal morphology and life history. Oikos:3-17.
Haage, M., T. Maran, U. A. Bergvall, B. Elmhagen, and A. Angerbjörn. 2017. The influence of spatiotemporal conditions and personality on survival in reintroductions–evolutionary implications. Oecologia **183**:45-56.
Haapakoski, M., A. Hardenbol, and K. D. Matson. 2018. Exposure to chemical cues from predator-exposed conspecifics increases reproduction in a wild rodent. Scientific Reports **8**:1-9.
Hadfield, J. D. 2010. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. Journal of statistical software **33**:1-22.
Haller, B. C. and A. P. Hendry. 2014. Solving the paradox of stasis: squashed stabilizing selection and the limits of detection. Evolution **68**:483-500.
Hansen, T. F., C. Pélabon, and D. Houle. 2011. Heritability is not evolvability. Evolutionary Biology **38**:258-277.
Hedrick, A. V. and R. Kortet. 2012. Sex differences in the repeatability of boldness over metamorphosis. Behavioral Ecology and Sociobiology **66**:407-412.
Hegab, I. M., G. Shang, M. Ye, Y. Jin, A. Wang, B. Yin, S. Yang, and W. Wei. 2014. Defensive responses of Brandt's voles (Lasiopodomys brandtii) to chronic predatory stress. Physiology & Behavior **126**:1-7.
Herde, A. and J. A. Eccard. 2013. Consistency in boldness, activity and exploration at different stages of life. Bmc Ecology **13**:49.
Hewes, M. E., K. Delventhal, and J. Chaves-Campos. 2017. Behavioral plasticity and consistency in the naked-footed mouse (Peromyscus nudipes) with habitat disturbance. Journal of Ethology **35**:279-292.
Hoefler, C. D., L. C. Durso, and K. D. McIntyre. 2012. Chemical-mediated predator avoidance in the European house cricket (Acheta domesticus) is modulated by predator diet. Ethology **118**:431-437.
Houle, D. 1992. Comparing evolvability and variability of quantitative traits. Genetics **130**:195-204.
Ivy, T. M. 2005. The evolution of polyandry in the decorated cricket, Gryllodes sigillatus. Illinois State University.
Johansson, F., P. Watts, S. Sniegula, and D. Berger. 2020. Natural selection mediated by seasonal time constraints increases the alignment between evolvability and developmental plasticity. Evolution.
Jones, A. G., S. J. Arnold, and R. Burger. 2004. Evolution and stability of the G-matrix on a landscape with a moving optimum. Evolution **58**:1639-1654.
Jones, A. G., S. J. Arnold, and R. Bürger. 2003. Stability of the G-matrix in a population experiencing pleiotropic mutation, stabilizing selection, and genetic drift. Evolution **57**:1747-1760.

Jones, A. G., S. J. Arnold, and R. Bürger. 2007. The mutation matrix and the evolution of evolvability. Evolution: International Journal of Organic Evolution **61**:727-745.
Ketola, M. and I. Vuorinen. 1989. Modification of life-history parameters of Daphnia pulex Leydig and D. magna Straus by the presence of Chaoborus sp. Hydrobiologia **179**:149-155.
Krueger, D. A. and S. I. Dodson. 1981. Embryological induction and predation ecology in Daphnia pulex. Limnology and Oceanography **26**:219-223.
Kruuk, L. E. 2004. Estimating genetic parameters in natural populations using the 'animal model'. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences **359**:873-890.
Laake, J. L. 2013. RMark: an R interface for analysis of capture-recapture data with MARK.
Lande, R. 1979. Quantitative genetic-analysis of multivariate evolution, applied to brain - body size allometry. Evolution **33**:402-416.
Lande, R. 1980. The genetic covariance between characters maintained by pleiotropic mutations. Genetics **94**:203-215.
Lande, R. 2009. Adaptation to an extraordinary environment by evolution of phenotypic plasticity and genetic assimilation. Journal of Evolutionary Biology **22**:1435-1446.
Lande, R. and S. J. Arnold. 1983. The measurement of selection on correlated characters. Evolution **37**:1210-1226.
Lind, M. I., K. Yarlett, J. Reger, M. J. Carter, and A. P. Beckerman. 2015. The alignment between phenotypic plasticity, the major axis of genetic variation and the response to selection. Proceedings of the Royal Society B-Biological Sciences **282**.
Lynch, M. and B. Walsh. 1998. Genetics and analysis of quantitative traits. Sinauer Sunderland, MA.
Marler, P. and S. Peters. 1987. A sensitive period for song acquisition in the song sparrow, Melospiza melodia: a case of age-limited learning. Ethology **76**:89-100.
Mathot, K. J., P. J. van den Hout, T. Piersma, B. Kempenaers, D. Réale, and N. J. Dingemanse. 2011. Disentangling the roles of frequency-vs. state-dependence in generating individual differences in behavioural plasticity. Ecology Letters **14**:1254-1262.
Messina, F. J. F., J. D. 2003. Environment-dependent reversal of a life history trade-off in the seed beetle Callosobruchus maculatus. Journal of Evolutionary Biology **16**:501-509.
Morand-Ferron, J., E. Varennes, and L.-A. Giraldeau. 2011. Individual differences in plasticity and sampling when playing behavioural games. Proceedings of the Royal Society B: Biological Sciences **278**:1223-1230.
Mousseau, T. A. and D. A. Roff. 1987. Natural selection and the heritability of fitness components. Heredity **59**:181-197.
Murren, C. J., J. R. Auld, H. Callahan, C. K. Ghalambor, C. A. Handelsman, M. A. Heskel, J. G. Kingsolver, H. J. Maclean, J. Masel, H. Maughan, D. W. Pfennig, R. A. Relyea, S. Seiter, E. Snell-Rood, U. K. Steiner, and C. D. Schlichting. 2015. Constraints on the evolution of phenotypic plasticity: limits and costs of phenotype and plasticity. Heredity **115**:293-301.
Noble, D. W., M. Lagisz, R. E. O'dea, and S. Nakagawa. 2017. Nonindependence and sensitivity analyses in ecological and evolutionary meta-analyses. Wiley Online Library.
Noble, D. W., R. Radersma, and T. Uller. 2019. Plastic responses to novel environments are biased towards phenotype dimensions with high additive genetic variation. Proceedings of the National Academy of Sciences **116**:13452-13461.

Nolte, D. L., J. R. Mason, G. Epple, E. Aronov, and D. L. Campbell. 1994. Why are predator urines aversive to prey? Journal of Chemical Ecology **20**:1505-1516.
Nussey, D. H., E. Postma, P. Gienapp, and M. E. Visser. 2005. Selection on heritable phenotypic plasticity in a wild bird population. Science **310**:304-306.
Ovaskainen, O., J. M. Cano, and J. Merilä. 2008. A Bayesian framework for comparative quantitative genetics. Proceedings of the Royal Society B: Biological Sciences **275**:669-678.
Phillips, P. C. and S. J. Arnold. 1989. Visualizing multivariate selection. Evolution **43**:1209-1222.
Piersma, T. and J. Drent. 2003. Phenotypic flexibility and the evolution of organismal design. Trends in Ecology & Evolution **18**:228-233.
Piersma, T. and J. A. Van Gils. 2011. The flexible phenotype: a body-centred integration of ecology, physiology, and behaviour. Oxford University Press.
Pinheiro, J., D. Bates, S. DebRoy, D. Sarkar, S. Heisterkamp, B. Van Willigen, and R. Maintainer. 2017. Package 'nlme'. Linear and nonlinear mixed effects models, version **3**.
Price, T. D., A. Qvarnström, and D. E. Irwin. 2003. The role of phenotypic plasticity in driving genetic evolution. Proceedings of the Royal Society of London. Series B: Biological Sciences **270**:1433-1440.
R Development Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Räsänen, K. and L. Kruuk. 2007. Maternal effects and evolution at ecological time-scales. Functional Ecology **21**:408-421.
Reale, D., S. M. Reader, D. Sol, P. T. McDougall, and N. J. Dingemanse. 2007. Integrating animal temperament within ecology and evolution. Biological Reviews **82**:291-318.
Roff, D. A. 1997. Evolutionary quantitative genetics. Springer Science & Business Media.
Roff, D. A., J. M. Prokkola, I. Krams, and M. J. Rantala. 2012. There is more than one way to skin a G matrix. Journal of Evolutionary Biology **25**:1113-1126.
Royauté, R. and N. A. Dochtermann. 2017. When the mean no longer matters: developmental diet affects behavioral variation but not population averages in the house cricket (Acheta domesticus). Behavioral Ecology **28**:337-345.
Royauté, R., C. Garrison, J. Dalos, M. A. Berdal, and N. A. Dochtermann. 2019. Current energy state interacts with the developmental environment to influence behavioural plasticity. Animal Behaviour **148**:39-51.
Sakaluk, S. K. 1987. Reproductive behaviour of the decorated cricket, Gryllodes supplicans (Orthoptera: Gryllidae): calling schedules, spatial distribution, and mating. Behaviour **100**:202-225.
Scheiner, S. M. 1998. The genetics of phenotypic plasticity. VII. Evolution in a spatially-structured environment. Journal of Evolutionary Biology **11**:303-320.
Scheiner, S. M. and D. Berrigan. 1998. The genetics of phenotypic plasticity. VIII. The cost of plasticity in Daphnia pulex. Evolution **52**:368-378.
Scheiner, S. M. and R. F. Lyman. 1991. The genetics of phenotypic plasticity. II. Response to selection. Journal of Evolutionary Biology **4**:23-50.
Schluter, D. 1996. Adaptive radiation along genetic lines of least resistance. Evolution:1766-1774.
Schneider, R. F. and A. Meyer. 2017. How plasticity, genetic assimilation and cryptic genetic variation may contribute to adaptive radiations. Molecular Ecology **26**:330-350.

Shonfield, J., R. W. Taylor, S. Boutin, M. M. Humphries, and A. G. McAdam. 2012. Territorial defence behaviour in red squirrels is influenced by local density. Behaviour:369-390.
Shorter, K. R., A. Owen, V. Anderson, A. C. Hall-South, S. Hayford, P. Cakora, J. P. Crossland, V. R. Georgi, A. Perkins, and S. J. Kelly. 2014. Natural genetic variation underlying differences in Peromyscus repetitive and social/aggressive behaviors. Behavior Genetics **44**:126-135.
Sikes, R. S., A. Care, and U. C. o. t. A. S. o. Mammalogists. 2016. 2016 Guidelines of the American Society of Mammalogists for the use of wild mammals in research and education. Journal of Mammalogy **97**:663-688.
Smith, B. R. and D. T. Blumstein. 2008. Fitness consequences of personality: a meta-analysis. Behavioral Ecology **19**:448-455.
Snell-Rood, E. C. 2013. An overview of the evolutionary causes and consequences of behavioural plasticity. Animal Behaviour **85**:1004-1011.
Stamps, J. A., J. B. Saltz, and V. V. Krishnan. 2013. Genotypic differences in behavioural entropy: unpredictable genotypes are composed of unpredictable individuals. Animal Behaviour **86**:641-649.
Stearns, S. C. and J. C. Koella. 1986. The evolution of phenotypic plasticity in life-history traits: predictions of reaction norms for age and size at maturity. Evolution **40**:893-913.
Stirling, D. G., D. Reale, and D. A. Roff. 2002. Selection, structure and the heritability of behaviour. Journal of Evolutionary Biology **15**:277-289.
Storey, A. E. 1994. Pre-implantation pregnancy disruption in female meadow voles Microtus pennsylvanicus (Rodentia: Muridae): male competition or female mate choice? Ethology **98**:89-100.
Sultan, S. E. and H. G. Spencer. 2002. Metapopulation structure favors plasticity over local adaptation. The American Naturalist **160**:271-283.
Svendsen, G. E. and K. B. Armitage. 1973. Mirror-Image Stimulation Applied to Field Behavioral Studies. Ecology **54**:623-627.
Taylor, R. W., A. K. Boon, B. Dantzer, D. Réale, M. M. Humphries, S. Boutin, J. C. Gorrell, D. W. Coltman, and A. G. McAdam. 2012. Low heritabilities, but genetic and maternal correlations between red squirrel behaviours. Journal of Evolutionary Biology **25**:614-624.
Thomson, C. E., I. S. Winney, O. C. Salles, and B. Pujol. 2018. A guide to using a multiple-matrix animal model to disentangle genetic and nongenetic causes of phenotypic variance. Plos One **13**:e0197720.
Uchida, K., T. Shimamoto, H. Yanagawa, and I. Koizumi. 2020. Comparison of multiple behavioral traits between urban and rural squirrels. Urban Ecosystems:1-10.
Van Buskirk, J. 2000. The costs of an inducible defense in anuran larvae. Ecology **81**:2813-2821.
Van Buskirk, J. and U. K. Steiner. 2009. The fitness costs of developmental canalization and plasticity. Journal of Evolutionary Biology **22**:852-860.
Van de Pol, M. and J. Wright. 2009. A simple method for distinguishing within-versus between-subject effects using mixed models. Animal Behaviour **77**:753.
Van Noordwijk, A. J. and G. de Jong. 1986. Acquisition and allocation of resources: their influence on variation in life history tactics. The American Naturalist **128**:137-142.
Van Tienderen, P. H. 1991. Evolution of generalists and specialist in spatially heterogeneous environments. Evolution:1317-1331.

Waller, J. and E. I. Svensson. 2016. The measurement of selection when detection is imperfect: How good are naïve methods? Methods in Ecology and Evolution **7**:538-548.
Walls, M. and M. Ketola. 1989. Effects of predator-induced spines on individual fitness in Daphnia pulex. Limnology and Oceanography **34**:390-396.
West-Eberhard, M. J. 2003. Developmental plasticity and evolution. Oxford University Press.
Westneat, D. F. and A. Sih. 2009. Sexual conflict as a partitioning of selection. Biology Letters **5**:675-677.
Westneat, D. F., J. Wright, and N. J. Dingemanse. 2015. The biology hidden inside residual within-individual phenotypic variation. Biological Reviews **90**:729-743.
Whitman, D. W. and A. A. Agrawal. 2009. What is phenotypic plasticity and why is it important. Phenotypic plasticity of insects: Mechanisms and consequences:1-63.
Wright, J., G. H. Bolstad, Y. G. Araya-Ajoy, and N. J. Dingemanse. 2019. Life-history evolution under fluctuating density-dependent selection and the adaptive alignment of pace-of-life syndromes. Biological Reviews **94**:230-247.
Zar, J. H. 1999. Biostatistical analysis. Pearson Education

APPENDIX

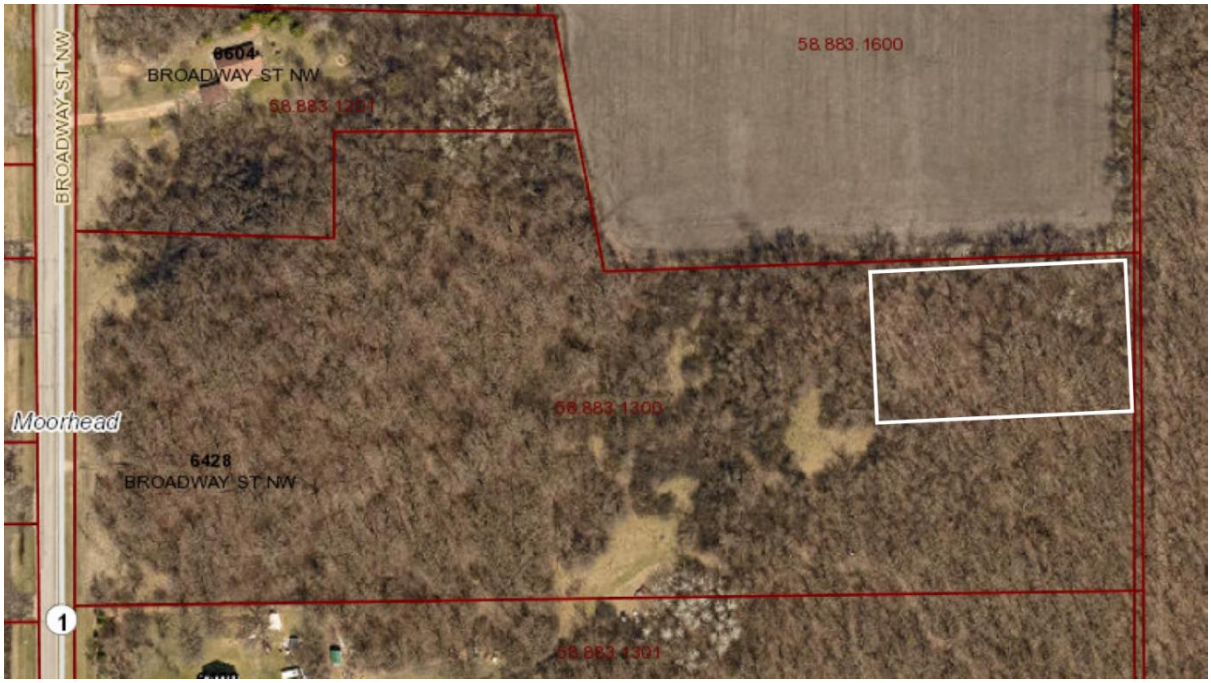

Figure A.1. Map of the Cassel wood, Minnesota (46°56'34.6"N, 96°47'03.2"W). The trapping grid's location is indicated by the white square.

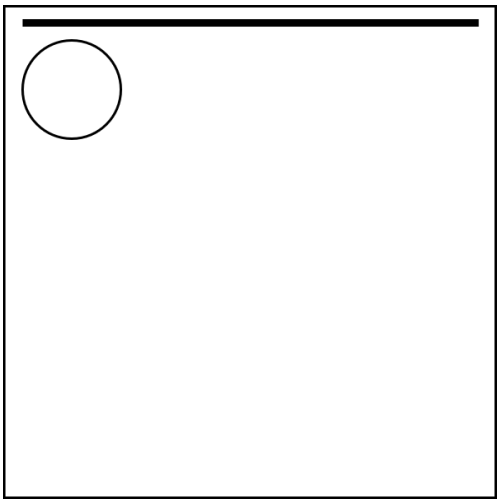

Figure A.2. Arena set up. The black line indicates the position of the mirror in the mirror arena, while the circle indicates the position of the predator cue in the anti-predator arena.

Table A.1. The eight mark-recapture models fit to investigate the effect of behavior on survival.

Model		Behavioral covariates
1	Effect of all behaviors on survival	Activity, aggression, and anti-predator response
2	Effect of activity and aggression on survival	Activity and aggression
3	Effect of activity and anti-predator response on survival	Activity and anti-predator response
4	Effect of aggression and anti-predator response on survival	Aggression and anti-predator response
5	Effect of Activity on survival	Activity
6	Effect of aggression on survival	Aggression
7	Effect of anti-predator response on survival	Anti-predator response
8	No effect of behaviors on survival	No behaviors

A.1. Testing for historical and stabilizing selection

Because we found that the behavioral traits were under weak (or no) directional selection, two post-hoc analyses were carried out to investigate if the behaviors had low evolvability and/or were under stabilizing selection. The first post-hoc analysis was the calculation of mean standardized variances for the three behaviors, i.e. evolvability, I_A, (Hansen et al. 2011).

$$I_A = \frac{V_A}{\mu^2} \quad \text{(Equation A.1)}$$

where V_A is the additive genetic variance of a trait, and μ is the mean of the same trait. This gives an estimate of the evolutionary potential of a population. Mean-scaling the variances allows us to compare these estimates across traits and species. Low values of evolvability suggest that a trait has been under strong historical selection, leading to a depletion of variation. However, in our case evolvability is estimated based on among-individual variance instead of just the additive genetic variance, meaning our estimate also included dominance- and epistatic variance as well as permanent environmental effects (Dochtermann and Royauté 2019). The estimated intercepts from the multi-response generalized mixed effect models were used for the behavioral means in equation A.1.

Second, we added quadratic terms for behaviors to the mark-recapture model (Arnold and Lande 1983) where a negative coefficient on the quadratic term would indicate stabilizing selection (Schluter 1988). As before, eight models were fit (Table A.1.) and AICc values were used to compare the model fit. All models included sex, developmental stage, mass, and the linear terms (i.e. the BLUPS) for the behaviors (Table A.1., Model 1), and only differed in the number of quadratic terms for the behaviors that were added (none, one, two, or three quadratic terms). If the behavioral traits were under stabilizing selection, models including these terms and with negative coefficients should have a better fit.

Table A.2. The eight mark-recapture models fit to investigate stabilizing selection on the three different behaviors. The directional selection terms for the three behaviors are included in all models.

Model		Behavior quadratic covariates
1	Quadratic term for all three behaviors	Activity, aggression, and anti-predator response
2	Quadratic term for activity and aggression	Activity and aggression
3	Quadratic term for activity and anti-predator response	Activity and anti-predator response
4	Quadratic term for aggression and anti-predator response	Aggression and anti-predator response
5	Quadratic term for activity	Activity
6	Quadratic term for aggression	Aggression
7	Quadratic term for anti-predator response	Anti-predator response
8	No quadratic terms	No quadratic terms for behaviors

Table A.3. Number of trials for the different sexes and developmental stages.

	Activity	Aggression	Anti-predator response
Adult female	98	84	93
Adult male	84	67	71
Adult unknown	3	3	3
Juvenile female	13	12	11
Juvenile male	21	19	17
Juvenile unknown	6	6	6
Total	225	191	201

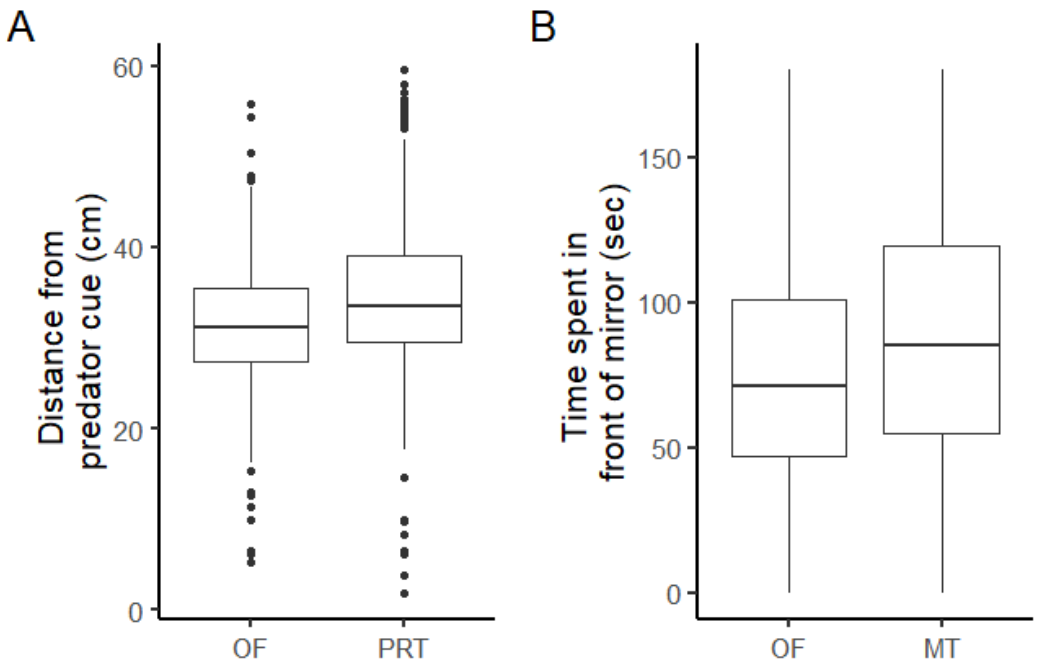

Figure A.3. Mean differences in A) distance from predator cue in the predator response test (PRT) and the open field (OF), and B) time spent on the mirror side of the arena in the mirror test (MT) and open field. The distance from the predator cue was substantially larger in the predator response test (pMCMC = 0.00), and the time spent in front of the mirror was substantially longer in the mirror test compared to the open field (pMCMC = 0.008).

Table A.4. Coefficients as standard deviation for the fixed effect from the multi-response generalized mixed effect model. Values in parenthesis shows 95% HDP-intervals.

	Activity	**Aggression**	**AP**
Shelter	**-1.07 (-1.46 – -0.69) pMCMC = 1.00**	-0.45 (-1.01 – 0.30) pMCM = 0.90	**-0.41 (-0.87 – 0.03) pMCM = 0.96)**
SexM	**0.51 (0.09 – 0.91) pMCM = 0.007**	0.13 (-0.29 – 0.57) pMCM = 0.29	-0.09 (-0.60 – 0.31) pMCM = 0.65
SexU	0.67 (-0.31 – 1.58) pMCM = 0.07	0.25 (-0.83 – 1.16) pMCM = 0.32	0.18 (-0.81 – 1.17) pMCM = 0.35
Developmental stage	-0.05 (-0.46 – 0.33) pMCM = 0.59	0.27 (-0.21 – 0.76) pMCM = 0.14	0.01 (-0.52 – 0.48) pMCM = 0.47
Mass (among-individual centred)	0.01 (-0.03 – 0.05) pMCM = 0.29	0.01 (-0.03 – 0.05) pMCM = 0.27	0.02 (-0.06 – 0.03) pMCM = 0.79
Mass (within-individual centred)	-0.04 (-0.10 – 0.01) pMCM = 0.94	0.01 (-0.06 – 0.07) pMCM = 0.45	0.05 (-0.02 – 0.10) pMCM = 0.06

A.2. Mark-recapture model

Only mass had a substantive effect on survival and only developmental stage had a substantive effect on recapture probability (Table A.5.). Larger individuals were found to have a higher survival probability than smaller individuals and adult individuals had a higher chance of being recaptured compared to juveniles. The selection coefficient for the behaviors were relatively low, especially for aggression. This low effect of behaviours on survival is further emphasised by the model comparison results, where the model with no behavioral terms had the lowest AICc score (Table A.6.). The same results were found when investigating the presence of stabilizing selection on the behaviors, where the model with no quadratic terms had the best fit (Table A.7, Table A.8.).

Table A.5. Survival and recapture coefficients estimated for sex, developmental stage, mass, and the three behavioral measurements based on the full mark-recapture model (Table A.1., Model 1).

Trait	**Survival coefficients Estimate (95 % CIs)**	**Recapture coefficients Estimate (95 % CIs)**
Intercept	0.46 (-2.24 – 3.15)	-0.66 (-3.24 – 1.92)
Sex1	0.03 (-1.54 – 1.62)	0.56 (-1.26 – 2.39)
Sex2	-0.66 (-2.39 – 1.07)	0.68 (-1.17 – 2.52)
Developmental stage	-0.88 (-2.32 – 0.57)	**1.18 (0.15 – 2.21)**
Mass	**0.19 (0.08 – 0.30)**	-0.03 (-0.09 – 0.03)
Activity	-0.79 (-1.61 – 0.03)	0.15 (-0.31 – 0.61)
Aggression	-0.23 (-1.21 – 0.76)	-0.22 (-0.66 – 0.23)
Anti-predator response	-0.70 (-1.85 – 0.45)	0.06 (-0.49 – 0.61)

Table A.6. AICc values for the 8 mark-recapture models. The reduced model has the best fit, indicating that the behavioral traits had very little impact on survival.

Model	AICc	ΔAICc	K
No effect of behaviour on survival ~ Sex + Developmental stage + Mass	739.28	0.00	3
Effect of only activity on survival ~ Sex + Developmental stage + Mass + Activity	741.33	2.05	4
Effect of only aggression on survival ~ Sex + Developmental stage + Mass + Aggression	742.48	3.20	4
Effect of only anti-predator response on survival ~ Sex + Developmental stage + Mass + Anti-predator response	743.51	4.23	4
Effect of activity and anti-predator response on survival ~ Sex + Developmental stage + Mass + Activity + Anti-predator response	744.52	5.24	5
Effect of activity and aggression response on survival ~ Sex + Developmental stage + Mass + Activity + Aggression	744.57	5.29	5
Effect of aggression and anti-predator response on survival ~ Sex + Developmental stage + Mass + Aggression + Anti-predator response	746.78	7.50	5
Effect of all three behaviors on survival ~ Sex + Developmental stage + Mass + Activity + Aggression + Anti-predator response	747.67	8.39	6

Table A.7. Effect of sex, developmental stage, mass, and the linear and quadratic terms for the three behavioral measurements on survival and recapture probability based on the post-hoc mark-recapture model.

Trait	**Survival probability**	**Recapture probability**
Intercept	0.20 (-2.69 – 3.09)	-0.71 (-3.25 – 1.83)
Sex 1	-0.41 (-1.99 – 1.16)	0.75 (-0.97 – 2.46)
Sex 2	-0.74 (-2.50 – 1.02)	0.79 (-0.95 – 2.54)
Developmental stage	-1.30 (-2.77 -0.17)	**1.43 (0.47 -2.40)**
Mass	**0.21 (0.09 – 0.34)**	-0.04 (-0.10 – 0.03)
Activity	-0.65 (-1.83 -0.53)	0.11 (-0.37 – 0.59)
Aggression	0.03 (-1.41 – 1.48)	-0.29 (-0.72 – 0.14)
Anti-predator response	-0.15 (-1.64 – 1.34)	-0.02 (-0.68 – 0.65)
Activity^2	0.55 (-0.65 – 1.74)	-0.15 (-0.77 – 0.48)
Aggression^2	2.66 (-0.76 – 6.08)	-0.17 (-1.00 – 0.67)
Anti-predator response^2	-0.76 (-2.82 – 1.30)	0.05 (-1.12 – 1.21)

Table A.8. AICc comparison for the mark-recapture models with quadratic terms. The reduced model has the best fit, but not significantly different from the model with the quadratic term for aggression.

Model	AICc	ΔAICc	K
No quadratic terms for behavior ~ Sex + Developmental stage + Mass + Activity + Aggression + Anti-predator response	747.67	0.00	6
Quadratic term for aggression ~ Sex + Developmental stage + Mass + Activity + Aggression + Anti-predator response + Aggression^2	749.44	1.77	7
Quadratic term for activity ~ Sex + Developmental stage + Mass + Activity + Aggression + Anti-predator response + Activity^2	751.69	4.02	7
Quadratic term for anti-predator response ~ Sex + Developmental stage + Mass + Activity + Aggression + Anti-predator response + anti-predator response^2	752.15	4.48	7
Quadratic term for activity and aggression ~ Sex + Developmental stage + Mass + Activity + Aggression + Anti-predator response + Activity^2 + Aggression^2	753.21	5.54	8
Quadratic term for aggression and anti-predator response ~ Sex + Developmental stage + Mass + Activity + Aggression + Anti-predator response + Aggression^2 + anti-predator response^2	753.71	6.04	8
Quadratic term for activity and anti-predator response ~ Sex + Developmental stage + Mass + Activity + Aggression + Anti-predator response + Activity^2 + anti-predator response^2	756.11	8.44	8
Quadratic term for all three behaviors ~ Sex + Developmental stage + Mass + Activity + Aggression + Anti-predator response + Activity^2 + Aggression^2 + anti-predator response^2	757.33	9.66	9

Table A.9. Vector correlations (above the diagonal) and angles (below the diagonal) between the dominant eigen vector of the among- and within-individual covariance matrices and the selection gradient (**β**).

	$\mathbf{i}_{max}$	$\mathbf{w}_{max}$	**β**
$\mathbf{i}_{max}$		0.36(0.00004 – 0.87)	0.01(0.0004 – 0.90)
$\mathbf{w}_{max}$	68.63˚(29.47˚ – 90.00˚)		0.18(0.009 – 0.94)
β	89.23˚(26.50˚ – 89.97˚)	79.42˚(20.32˚ - 89.75˚)	

A.3. Appendix references

Dochtermann, N. A. and R. Royauté. 2019. The mean matters: going beyond repeatability to interpret behavioural variation. Animal Behaviour **153**.

Hansen, T. F., C. Pélabon, and D. Houle. 2011. Heritability is not evolvability. Evolutionary Biology **38**:258-277.

Lande, R. and S. J. Arnold. 1983. The measurement of selection on correlated characters. Evolution **37**:1210-1226.

Schluter, D. 1988. Estimating the form of natural selection on a quantitative trait. Evolution **42**:849-861.