Environmental Stress Responses and Biological Interactions Investigated in the Drosophila Model System

Ørsted, Michael

DOI (link to publication from Publisher):
10.5278/vbn.phd.eng.00030

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

ENVIRONMENTAL STRESS RESPONSES AND BIOLOGICAL INTERACTIONS INVESTIGATED IN THE DROSOPHILA MODEL SYSTEM

BY
MICHAEL ØRSTED

DISSERTATION SUBMITTED 2017

AALBORG UNIVERSITY
DENMARK

ENVIRONMENTAL STRESS RESPONSES AND BIOLOGICAL INTERACTIONS INVESTIGATED IN THE *DROSOPHILA* MODEL SYSTEM

PHD THESIS

BY

MICHAEL ØRSTED

DEPARTMENT OF CHEMISTRY AND BIOSCIENCE

FACULTY OF ENGINEERING AND SCIENCE

AALBORG UNIVERSITY

Dissertation submitted 14 October 2017

AALBORG UNIVERSITY
DENMARK

Dissertation submitted: 14-10-2017

PhD supervisor: Professor Torsten Nygaard Kristensen,
Department of Chemistry and Bioscience
Aalborg University
Denmark

PhD committee: Associate Professor Majken Pagter (chairman)
Aalborg University

Professor Yvonne Willi
University of Basel

Professor Juan L. Bouzat
Bowling Green State University

PhD Series: Faculty of Engineering and Science, Aalborg University

Department: Department of Chemistry and Bioscience

ISSN (online): 2446-1636
ISBN (online): 978-87-7210-088-3

Published by:
Aalborg University Press
Skjernvej 4A, 2nd floor
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

Printed in Denmark by Rosendahls, 2017

Credits:
Front/back page pictures: iStock.com/janeff
Layout inspiration: Mads F. Schou, Palle D. Rohde & Kristian Trøjelsgaard

PREFACE

This thesis represent the culmination of three years of PhD study at the Department of Chemistry and Bioscience in the Section of Biology and Environmental Science, Aalborg University. The thesis is divided into two main parts. The first part is a general introduction providing a broad overview of the fields of environmental stress responses, biological interactions and the genetics of inbreeding, and serves to introduce the main ideas behind the included papers and projects, and puts them into perspective. The second part consists of four papers and as well as a presentation of the results of ongoing work, which is not yet formulated into a full manuscript. These papers and projects are the principal products of my PhD.

Michael Ørsted
Aalborg, October 2017

ACKNOWLEDGEMENTS

First and foremost, I want to express my sincerest gratitude to my supervisor, Torsten Nygaard Kristensen, for supporting and guiding me throughout my PhD. You have always trusted me and showed an immensely positive attitude towards my work, and given me the freedom to pursue whichever path my studies took me. I owe thanks for introducing me to a vast network of like-minded researchers - encounters that have driven my desire to pursue a scientific career. Beyond your role as my academic advisor, I have enjoyed your company outside the lab and office, and I hope our collaboration will continue in many future projects.

I would like to thank fellow young scientists, Mads F. Schou and Palle D. Rohde from Aarhus University for inspirational and invaluable scientific discussions and for being great collaborators. Furthermore, you showed me the ins and outs of being a PhD student, and you both have my deepest gratitude for your aid with data analysis, programming, and for educating me in the realm of bioinformatics. My thanks go to all colleagues, past and present, at the Department of Chemistry of Bioscience, in particular, Simon, Kristian, Niels I. and Peter for always being helpful and smiling and for sharing a Friday afternoon beer, and to Neda for joyful talks in the office - academic or otherwise - and for always bringing me coffee. I would also like to thank our lab technicians, especially Helle, Susan and Henriette. Without your help throughout all three years, this would not have been possible. I am grateful to you for your ever-positive attitude. I enjoyed coming to work every day, for which all my colleagues have the credit. Thanks to all the master students who helped me in the lab.

To Ary A. Hoffmann I owe a lot. You generously hosted me at the infamous Bio21 Institute at the University of Melbourne for six months, and I value your kindness and our great scientific discussions. Your immense knowledge and your ability to remember practically every person and paper in our field, never cease to amaze me. I truly admire your inspirational story telling approach to science both on paper and in person, something I will always remember in my research communication. Also at Bio21, I thank Kelly and Nancy for lab assistance, and Jason and Perran for introducing me to the work with mosquitoes, a venue of research I hope to pursue in the future. Thanks to my housemate in Melbourne, Josh, for showing me around town, for various hiking trips and for your open-minded approach to people.

I would like to thank friends and family for your support and for your patience when I had to go to the lab at the most inconvenient times. Above all, I want to express my special appreciation to Iben for her unconditional love, support and encouragement, and for being a great discussion partner and travel buddy at conferences. I owe the greatest of thanks for your help throughout my entire PhD, but especially in the last eight weeks since our son, Sigurd, was born, conveniently coinciding with the finalization of this thesis. Talk about multiple stresses.

TABLE OF CONTENTS

LIST OF PAPERS

Included papers are referred to by Roman numerals (I-IV)

PAPER I **Ørsted, M.**, Schou, M.F., & Kristensen, T.N. (2017). Biotic and abiotic factors investigated in two *Drosophila* species – evidence of both negative and positive effects of interactions on performance. *Scientific Reports*, 7, 40132.

PAPER II **Ørsted, M.**, Malmendal, A., Muñoz, J. & Kristensen, T.N. (2017). Metabolic and functional phenotypic profiling of *Drosophila melanogaster* reveal reduced sex differentiation under stressful environmental conditions. *Biological Journal of the Linnean Society* (In print).

PAPER III **Ørsted, M.**, Rohde, P.D., Hoffmann, A.A., Sørensen, P. & Kristensen, T.N. (2017). Environmental variation partitioned into separate heritable components. *Evolution* (In review after first revision).

PAPER IV Jensen, C.*, **Ørsted, M.*** & Kristensen, T.N. (2017). Temporal dynamics and effects of genetic distance in genetic rescue investigated in a *Drosophila melanogaster* model system. *Genetica* (In review).

* Shared first authorship

In addition, I will present and briefly discuss results from ongoing work entitled: "Consequences of population bottlenecks on adaptive genetic variation revealed in a highly replicated experimental evolution study".

SUMMARY

The reoccurring theme of this thesis is the use of fruit flies as model organisms for studying how natural populations respond and adapt when faced with a multitude of environmental stresses and the consequences of reduced populations size and loss of genetic variation on the ability to evolve. The common denominator in the papers presented here, is the investigation of responses to many different ecologically relevant environmental stresses. Biological interactions are likewise a major constituent of many of the papers, both interactions between multiple environmental conditions, interactions between sex and the environment, or how the environmental factors interacted with the genetic constitution of individuals, across both environments and time. For this purpose, I used different species of the genus *Drosophila* originating from wild-caught populations from Denmark or Australia or from a panel of sequenced isogenic lines of *D. melanogaster*.

In **PAPER I**, two naturally co-occurring species of *Drosophila* was tested to investigate the responses to combinations of both biotic and abiotic environmental conditions, on a range of fitness related traits. The study found that, although interactions between stresses do sometimes occur and can have highly adverse effects on performance, additive effects of combinations of environmental stress were most common. Furthermore, the responses were highly species-, trait-, and sex dependent. This highlighted the importance of considering the combined effect of environmental stresses in prediction models of species responses to e.g. climate change, and in ecological risk assessments. The study also revealed the need for a re-conceptualized terminology for describing the complexity of interactions between environmental conditions. Building on the differential phenotypic responses to environmental stressors and the sex dependency of such responses in **PAPER I**, it was investigated in **PAPER II**, whether a general metabolic stress response (using NMR metabolomics) could be identified in males and females across a range of different environmental stresses that fruit flies are likely to encounter in the wild. I found a difference between *D. melanogaster* males and females in the way they plastically responded across a range of different types of stress. At both the metabolite level and at the functional phenotypic level, this resulted in a decrease of the sexual dimorphism with the severity of the stress, with possible implications for the effects of environment on sexual selection. No evidence of a generic stress response was found in the metabolome.

In **PAPER III**, I investigated how environmental stress can interact with the genotype of individuals, and what genetic architecture governs why some individuals are more variable and plastic in their ability to adapt to a range of different environments, while others are more canalized. For this purpose, the *Drosophila* Genetic Reference Panel (DGRP) was used. DGRP is a set of ~200 fully inbred and

sequenced lines, suitable for studying the genetic basis of complex traits. I found that genetic variation (V_G) and environmental variation (V_E) are not independent, as a genetic control of V_E was confirmed. In this study it is proposed that environmental variation can be partitioned into four different conceptual components. Genetic control of all four V_E components encompassing variation across and within environments was identified. I found little overlap in the genetic background between some of these V_E measures, while others were genetically correlated.

In **PAPER IV**, the focus shifted towards genetic stress in the form of inbreeding, and how to alleviate some of the consequences of inbreeding and loss of genetic variation. **PAPER IV** has a conservation-oriented perspective, and focuses on how to save small, fragmented, extinction-threatened populations with little genetic variation, by translocating individuals from other populations to re-establish gene flow, a technique known as 'genetic rescue'. For this purpose, the DGRP system was also used, in this context to simulate genetically deteriorated populations expressing high levels of inbreeding depression. It was investigated whether the success of a translocation depended on the genetic distance between the receiver and donor population. The results provided clear evidence of high fitness enhancements in hybrid offspring (heterosis), but also a temporal decline of such benefits. Genetic distance between donor and recipient population did not have strong impact on the level of heterosis.

Small populations might suffer from inbreeding depression as illustrated in **PAPER IV**. They may also suffer from lack of genetic variation due to genetic drift, which can reduce the evolutionary potential. While this is often highlighted as one of the major concerns for small extinction prone populations, large-scale empirical evidence of this hypothesis is surprisingly scarce and some recent evidence suggest that associations between the effective population size and the amount of genetic variation is more complex than hitherto assumed. To investigate this in more detail, I set up a highly replicated evolution experiment with lines of a wild caught population of *D. melanogaster* inbred to different degrees, from which I will present some analyses and result, and briefly discuss possible implications. The first data suggest high line specificity, but generally support the expectations, that increasing levels of inbreeding leads to reduced evolutionary response to selection.

In summary, this thesis investigates how, and to what extent, insect model species respond to a multitude of different environmental stresses, how the environment interacts with the genetic composition of individuals, and lastly the consequences of inbreeding on the adaptive ability, and how to possible alleviate some of the negative fitness effects of inbreeding.

RESUMÉ

Det gennemgående tema for denne afhandling er brugen af bananfluen som modelorganisme for at undersøge af hvordan naturlige populations reagerer på og tilpasser sig miljøstress, samt hvilke konsekvenser en reduceret populationsstørrelse og deraf resulterende tab af genetisk variation har for arters evne til at tilpasse sig evolutionært. De præsenterede artikler har det tilfælles, at de undersøger responsen på en række økologisk relevante miljøstresser. Biologiske interaktioner udgør et centralt element i flere af artiklerne; både fitness konsekvenser og adaptive responser på flere samtidige miljøfaktorer, effekter af interaktioner mellem køn og miljø, eller hvordan miljøfaktorer interagerer med den genetiske sammensætning af individer, dels på tværs af miljøgradienter og på tværs af tid. Til dette formål har jeg benyttet forskellige bananfluearter af slægten *Drosophila*, som stammer enten fra vildtfangede populationer fra Danmark eller Australien, eller fra et panel af sekvenserede isogenetiske linjer af *Drosophila melanogaster*.

I **ARTIKEL I**, undersøges det hvordan to naturligt sameksisterende *Drosophila* arter reagerer på kombinationer af både biotiske og abiotiske miljøfaktorer på en række fitness relaterede træk. Dette studie fandt, at på trods af, at stressfaktorer kan interagere i deres effekt på fitness, og at disse kan have meget negative konsekvenser, så var de additive effekter af kombinationerne af miljøstress hyppigst. Derudover var de observerede responser meget arts- og kønsafhængige, samt afhængige af hvilket træk, der blev undersøgt. Dette understreger vigtigheden af at inkludere de kombinerede effekter af miljøstress i prædiktionsmodeller over arters respons på eksempelvis klimaforandringer samt i risikovurderinger af fx kemikalier og forurening. Studiet afslørede desuden et behov for at udvide begreberne, som bruges til at beskrive komplekse interaktioner mellem miljøfaktorer. For yderligere at undersøge baggrunden for de kønsafhængige fænotypiske responser på miljøstres i **ARTIKEL I**, blev det i **ARTIKEL II** undersøgt, om der kunne findes kønsspecifikke eller generelle metaboliske stress responser (ved brug af NMR metabolomics) i hanner og hunner, på tværs af en række vidt forskellige miljøstresser. Disse typer stress var alle nogle bananfluer vil kunne opleve i naturen. Jeg fandt en væsentlig forskel i den plastiske respons på stress i hanner og hunner af *D. melanogaster* på tværs af de forskellige typer af stress. For både funktionelle fænotyper og på metabolit niveau resulterede dette i en reduktion af kønsforskellen med stress, og denne var proportionel med intensiteten af de forskellige typer stress,. Dette kan have betydning for, hvilken indflydelse miljøet har på graden af seksuel selektion. Jeg fandt ingen antydninger af en universel stress respons i metabolomet.

I **ARTIKEL III** undersøgte jeg hvordan miljøstress kan interagere med individers genotype, samt den genetiske baggrund, der styrer hvorfor nogle individer er variable og plastiske i deres evne til at tilpasses en række forskellige miljøer, mens andre er

mere ensartede og ude af stand til at reagere plastisk. Til dette formål brugte jeg en ressource kaldet *Drosophila* Genetic Reference Panel (DGRP). DGRP er et sæt af ca. 200 komplet indavlede sekvenserede linjer, som er egnede til at studere den genetiske baggrund for komplekse træk. Jeg fandt, at genetisk varians (V_G) og miljøvarians (V_E) ikke er uafhængige - graden af miljøvarians er genetisk bestemt. I studiet foreslås det, at miljøvarians kan opdeles i fire konceptuelle delelementer, der inkluderer variation indenfor og på tværs af miljøer. Alle fire delkomponenter af V_E var genetisk kontrollerede, og der blev fundet meget lidt overlap i den genetiske baggrund for disse V_E mål.

I **ARTIKEL IV**, ændres fokus til at omhandle genetisk stress i form af indavl, samt hvordan konsekvenserne af indavl og tab af genetisk variation kan modvirkes. **ARTIKEL IV** har et bevaringsorienteret perspektiv, og fokuserer på hvordan man kan redde små fragmenterede udryddelsestruede populationer med lav genetisk variation ved at flytte individer fra andre populationer for at sikre genudveksling, en teknik der kaldes 'genetic rescue'. Til dette formål blev DGRP igen benyttet, denne gang til at simulere populationer som lider under indavlsdepression. Det blev undersøgt hvorvidt successen af en translokation af individer afhang af den genetiske afstand mellem modtager- og donorpopulationen. Resultater viste tydelig evidens for store fitness forbedringer i hybridafkommet (kaldet heterosis), men også en nedgang i sådanne fordele med tiden. Genetisk afstand viste sig ikke at have en stor effekt på mængden af heterosis.

Små populationer kan lide under indavlsdepression, som det blev belyst i **ARTIKEL IV**. De kan desuden lide under manglen på genetisk variation på grund af genetisk drift, hvilket muligvis reducerer det evolutionære potentiale. Selvom dette ofte fremhæves, som en af de største bekymringer for små udryddelsestruede populationer, er empiriske beviser for denne hypotese overraskende sjældne. Desuden tyder nylige studier på, at sammenhængen mellem populationsstørrelse og mængden af genetisk variation er mere kompleks end først antaget. For at undersøge dette opsatte jeg et eksperimentelt evolutionsforsøg med et højt antal linjer fra en vildtfanget *D. melanogaster* population, som blev indavlet til forskellige niveauer. Jeg præsenterer analyser og resultater fra dette forsøg og diskuterer kort mulige konsekvenser. Ind- og udavlede linjer blev holdt i 10 generationer på et stressende medie, and evolution i fitness relaterede træk blev undersøgt. De første data afslører en høj linjespecificitet, men generelt understøtter at øget indavl medfører en reduceret evne til at tilpasses sig gennem evolutionære ændringer.

Samlet set undersøger denne afhandling hvordan, og i hvilken grad, insekt model arter responderer på en række forskellige miljøstressorer, hvordan miljøet vekselvirker med individers genetiske komposition, og slutteligt konsekvenserne af indavl på evnen til at tilpasse sig, og hvordan de negative fitness effekter af indavl muligvis kan lettes.

ABBREVIATIONS

Commonly used abbreviations:

A(+/-)	Antagonistic interaction (positive/negative)
BPH	Best-parent heterosis
CT_{max}	Critical thermal maximum
CT_{min}	Critical thermal minimum
DGRP	*Drosophila* Genetic Reference Panel
F	Coefficient of inbreeding
F_{ST}	Fixation index
G x E	Genotype-by-environment interaction
GBLUP	Genomic best linear unbiased prediction
GD	Genetic distance
GFBLUP	Genomic feature best linear unbiased prediction
GLM	General linear model
GLMM	General linear mixed model
GO	Gene ontology
GWAS	Genome-wide association study
HCA	Hierarchical cluster analysis
I x E	Inbreeding-by-environment interaction
MPH	Mid-parent heterosis
n	Sample size
N	Census population size
N_e	Effective population size
NMR	Nuclear Magnetic Resonance
PCA	Principal component analysis
PR	Potence ratio
REML	Restricted maximum likelihood
RING	Rapid Iterative Negative Geotaxis
RNAi	RNA mediated gene interference
RO	Reproductive output
S(+/-)	Synergistic interaction (positive/negative)
SD	Standard deviation
SE	Standard error
SNP	Single nucleotide polymorphism
SR	Starvation resistance

INTRODUCTION

AN EVER-CHANGING ENVIRONMENT

When organisms are faced with change in their immediate surroundings, they are forced to respond, if they are to maintain optimal function. Especially, ectotherms must deal with environmental changes on a regular basis, thus their survival and reproductive success depend on their ability to adjust according to the environmental cues. In the short term, e.g. daily or seasonal temperature fluctuations, organisms respond to variable environmental conditions through behavioral, physiological and/or morphological adjustments (Hoffmann & Parsons 1991; Angilletta 2009). While some environmental changes have so little impact that a response is hardly observed, other changes may occur with a magnitude or rate of change that exceeds the capabilities of the organism. If unable to respond sufficiently, such environmental changes will harm the normal functioning of the organism, and potentially decrease survival and reproductive fitness. Such environmental changes are defined as 'environmental stress' (Hoffmann & Parsons 1991), and will be employed as a working definition throughout the current thesis.

Many short term fluctuations such as daily or seasonal variation occur within the same generation, however some environmental disturbances span many generations and might require long-term evolutionary responses in order to maintain a normal functionality in a changing environment (Hoffmann & Willi 2008; Willi & Hoffmann 2009; Chown *et al.* 2010). One example of such environmental disturbance is the steadily increasing human impact on natural ecosystems, e.g. anthropogenic climate change. In the last few centuries, many species have experienced unprecedented rates of climate change (Smith *et al.* 2015). Despite an average temperature increase of only ~1 °C since before industrial times, the global footprint of a growing human population is well documented across all ecosystems on the planet (Parmesan & Yohe 2003), and effects are present on all biological levels from genes to biomes (Scheffers *et al.* 2016). In addition to an increase in mean temperature, it is also predicted that both temperature and precipitation patterns become more variable (IPCC 2013). Since the fitness of individuals depends on their ability to accurately predict the environmental change (Manenti *et al.* 2014), an increase in variability could mean that species will struggle to anticipate future climate conditions (Ketola *et al.* 2013). It has been suggested that evolutionary responses might be either too slow or constrained to allow species to adapt to the rapidly deteriorating state of their environments (Kelly *et al.* 2012; Kellermann *et al.* 2012; Araújo *et al.* 2013; Hoffmann *et al.* 2013; Schou *et al.* 2014; Kristensen *et al.* 2015). This means that some organisms will have to depend in part on adjusting their phenotype according to environmental cues. This is

termed ‘phenotypic plasticity’ and is a re-occurring theme throughout this thesis that will be discussed in details (**PAPER III**).

STUDYING ENVIRONMENTAL STRESS RESPONSES

The reasons for studying environmental stress are many. There is a fundamental curiosity, which drives research, but also there is an increasing need to elucidate the effects of a myriad of different environmental factors on a wide range of biological organizational levels, from DNA to entire ecosystems. Recently, the list of chemical, biological, and physical stressors that are considered to be potentially harmful to the environment has grown rapidly (Novacek & Cleland 2001; Folke *et al.* 2004; Halpern *et al.* 2008). Scientists, conservation managers, and policy makers are urged to consider the ecological consequences of stressors for appropriate regulation and management of natural resources (De Lange *et al.* 2010).

Traditionally, assessment of the effects of environmental stressors has predominately been based on the results of laboratory experiments where a test organism has been exposed to an individual stressor. This is especially pertinent to the assessments of potentially harmful chemicals, where a single compound is tested often across a range of concentrations to obtain a dose-response relationship, and establish toxicity data, e.g. the concentration resulting in 50 % mortality (LC_{50}). Such measures enable easy comparisons across compounds, and used by policy makers for appropriate management of chemicals. In such tests, the test organisms are usually maintained at optimal and constant temperature, humidity, pH, etc. and are given food in abundance. Examples include many of the standardized toxicity tests still employed by governmental and international institutions (e.g. US-EPA 2002; ISO 2012).

In nature, however, species rarely experience optimal environmental conditions, but are forced to cope with sub-optimal and often stressful conditions for the majority of their life, with large fluctuations in e.g. food availability or climatic conditions as discussed above. Beyond an increase in temperature mean and variability linked to climate change, an increase in the intensity and diversity of other anthropogenic environmental stressors has also been observed as a result of a growing human population in the last decades (Halpern *et al.* 2007). These include e.g. habitat loss, urbanization, pollution, increase in invasive species and diseases, and many derived effects of climate change like increasing sea levels, and ocean acidification (Novacek & Cleland 2001; Allison & Bassett 2015). For a realistic and ecologically relevant assessment of stress responses, they must be viewed in the context of a plethora of environmental conditions, and their potential interactions acting simultaneously.

ENVIRONMENTAL INTERACTIONS

Ecological research have been elucidating the effects of the abovementioned effects individually, empirical studies on the cumulative effects and potential interactions between individual stressors are far less frequent (Crain *et al.* 2008; Darling & Côté 2008), despite natural systems being exposed to several human-derived stressors simultaneously for most of the time (Halpern *et al.* 2007; Laskowski *et al.* 2010). The fitness impact of an environmental factor may be minute when considered in isolation. However, multiple environmental factors may interact and yield effects that are widely different from the sum of the individual stressors on the fitness of individual organisms as well as on the community structure in an ecosystem. Understanding the ecological effects of environmental stressors and the effects of their potential interactions on fitness is of great importance for global climate change prediction models (Kaunisto *et al.* 2016), where multiple stressors may interact in a manner, that is not predictable from individual stressors. Some studies predict that multiple stresses will interact and accelerate biodiversity loss (Sala *et al.* 2000) and/or amplify the effects of already existing anthropogenic stresses (Halpern *et al.* 2008). In any case, when interactions either mitigate or exacerbate the effects of individual stresses in natural environments (Didham *et al.* 2007; Mora *et al.* 2007), this has sometimes been termed 'ecological surprises' (Paine *et al.* 1998), and exemplify a key uncertainty in projections of biodiversity (Pereira *et al.* 2010) and ecosystem resilience (Folke *et al.* 2004). Consequently, neglecting interactions of environmental factors can make predictions of individual performance and community structure inaccurate (Relyea & Hoverman 2006; Schuwirth *et al.* 2015; Kéfi *et al.* 2016). There is a potential risk for underestimating the severity of the effect of multiple environmental stresses on species distributions and extinction risks e.g. thermal extremes in combination with draught or chemical stress (Visser 2008; Bellard et al. 2012). In ecotoxicology and ecological risk assessments, not incorporating knowledge on multiple stressors can lead to underestimating risk (Bednarska *et al.* 2013), which of course is problematic, but also overestimating the risk which can have substantial undesirable economic consequences (Holmstrup *et al.* 2010).

Amongst the studies that have been conducted on multiple stressors, the majority investigates the potential interactions between only two environmental conditions, the far most common combination being between a chemical compound and some other abiotic stressor, e.g. another chemical or temperature stress (Holmstrup *et al.* 2010; Laskowski *et al.* 2010). In the context of ecological relevance, this can be problematic, because such studies ignore biotic interactions, which play an important role in the evolution (Thorpe *et al.* 2011) and distribution of many, if not all, species, through e.g. predation, competition, mutual dependencies etc. (Wisz *et al.* 2013). Some even

argue that biotic interactions are more important than abiotic habitat requirements for determining distribution ranges and community compositions (Schuwirth *et al.* 2015).

In **PAPER I**, I investigated the consequences of exposing two naturally co-occurring species of fruit flies (*Drosophila hydei* and *Drosophila melanogaster*) to both biotic and abiotic environmental factors in a full factorial manner, i.e. both in isolation and in all combinations, to examine effects of potential interactions on fitness components. In this paper, and in all papers presented in this thesis, I have put much emphasis on the ecological relevance of the environmental conditions, i.e. both the types of stressors and the levels of intensity, are likely to be encountered by insects in a natural setting. Effects of environmental interactions should ideally be included in all studies to provide the ecological context all stressors should be evaluated in. However, these types of experiments (full factorial) are very cumbersome, as the number of interactions increases exponentially with the number of environmental variables considered. In **PAPER I**, though, the purpose was to specifically elucidate the nature of individual stressors and the strength and frequency of their two- and three-way interactions. Recent reviews on fitness effects on interactions give the impression that interactions are more the rule than the exception, and that most interactions are of the synergistic type, i.e. when combined effects are greater than the expected additive sum, and stressors exacerbate their mutual effects (Crain *et al.* 2008; Darling & Côté 2008). Contrary to this notion, the results from **PAPER I** suggested that although interactions did occur, additive effects of stressors were more common. This discrepancy could be explained by researchers tending to be biased towards publishing "positive" results, i.e. findings of interactions rather than simply the additive effects (Holmstrup *et al.* 2010), which could cause the frequency of interactions in nature to be incorrectly reflected.

Interestingly, I also found a high proportion of positive effects of interactions, e.g. *D. hydei* benefitted greatly in many traits from co-occurring alongside *D. melanogaster*. This result might seem counterintuitive in a study of stressful environmental conditions, however, the findings are congruent with other studies showing that the number of positive interaction increase with stress (Callaway *et al.* 2002; Brooker *et al.* 2008). Positive interactions, e.g. the development of intrinsic mutual dependencies might be a mechanism that will be increasingly adopted by species communities to counteract the increase in environmental stress with global climate change (He *et al.* 2013). I initially viewed these environmental interactions in the context of the classically defined terms of synergism and antagonism (when combined effects are smaller than expected) (Folt *et al.* 1999). However, due to the complexity of the results, especially in situations where individual stresses were of opposite effect directions (some with positive effects, others with negative), it became quickly clear that it was necessary to update the terminology of interactions to offer more informative descriptions. Such re-conceptualized terms has recently been

suggested by others for two-way interactions (Piggott *et al.* 2015), however, I expanded them to include three-way interactions as well.

BIOLOGICAL INTERACTIONS

The stress responses in **PAPER I** were in some cases sex specific, where males and females responded differently to the environmental conditions, congruent with other studies (Hoffmann *et al.* 2005; Sørensen *et al.* 2007). Similarly, I found that many of the responses were highly dependent on specific stressors, and on which trait was investigated. These observations were explored in further detail in **PAPER II**, where I investigated to what extent a general stress response could be recognized both across environments and sexes. The initial idea was partly to try to identify generic responses to a wide range of different ecologically relevant stressors on a sub-organismal level and compare these to responses on the functional phenotypic level. In ecological risk assessments many studies rely on rather dichotomous and insensitive endpoints at the organismal level such as mortality (Darling & Côté 2008) or mobility (ISO 2012), which are 'either-or', and leaves little room for quantifying gradual stress responses. As a result, assays that examines responses on the sub-organismal level, e.g. using molecular, physiological, or biochemical parameters, so-called biomarkers, have received increased attention (Forbes *et al.* 2006). Biomarkers may characterize initial responses to stressors and toxicants that can be detected before survival is affected (Ørsted & Roslev 2015), and can represent efficient ways to quantify sub-lethal effects on e.g. growth and reproduction at the organismal or population level (Forbes *et al.* 2006). For this purpose, I employed nuclear magnetic resonance (NMR) metabolomics in **PAPER II**, to study the effects of environmental stressors on metabolite composition. Metabolomics is a characterization of endo- and exogenous low molecular mass metabolites within a biological sample, e.g. cells, tissues or whole-organism homogenates. For this purpose NMR technology is particularly helpful, as it allows for a non-targeted and comprehensive analysis of all or most of all the metabolites in a sample, that is possibly closer to the organismal phenotype than the other 'omics' techniques, e.g. gene expression (transcriptomics) and protein changes (proteomics), which are both subject to rather complex feedback and homeostatic control mechanisms (Nicholson *et al.* 1999; Ankley *et al.* 2006; van Ravenzwaay *et al.* 2007).

In **PAPER II**, I exposed *D. melanogaster* to different 'natural' stressful treatments by varying media contents and thermal environments, and investigated the metabolite composition as well as functional phenotypes (size and survival) of both males and females. I found that the difference in metabolite compositions between sexes were greatest in benign environments, and decreased linearly with the severity of the stress. Similarly, in terms of body mass I found that females responded more under

environmental stress, i.e. they were more plastic than males, resulting in a similar decrease in the sexual dimorphism of body size with increased stress, concurrent with the metabolomic results. Some of the metabolites found in highest concentrations in control females as compared to both males and stressed females were seemingly related to reproduction, and suggested that the reduced sexual dimorphism in stressful environments was associated with a trade-off between reproduction and stress resistance, which is a commonly observed trade-off (Partridge *et al.* 2005). This could have some really interesting evolutionary implications, and I speculate that environmental factors can play an important role in shaping sexual selection, an idea that goes all the way back to Alfred Russel Wallace (Wallace 1889). Ketola *et al.* (2012) found that sexual dimorphism in heritability for heat tolerance in *D. melanogaster* was affected by developmental temperature, and that genetic variation for the trait was genetically uncorrelated in the two sexes, suggesting potential for independent evolution between sexes.

PHENOTYPIC AND ENVIRONMENTAL VARIATION

I have now introduced interactions between environment factors themselves, and between the environment and sex. In discussing and introducing **PAPERS III-IV**, I will focus more specifically on the impact of genetics on the phenotype, and how genetic factors can interact with environmental conditions. Genetic and environmental factors have for very long been viewed as independent and have founded the alliterative expression '*nature versus nurture*'. However, this is often too simplistic, as genotypes may respond differently to changes in the environment. Therefore, phenotypic variation is determined by the sum of genetic variation and environmental variation as well as genotype-by-environment (G x E) interactions in modern quantitative genetic theory (Falconer & Mackay 1996; Lynch & Walsh 1998):

$$V_P = V_G + V_E + V_{G\,x\,E}$$

The G x E interaction is sometimes described as genotypic differences in what is referred to as 'environmental sensitivity'. This can be defined in two ways. The first definition of environmental sensitivity is the mean phenotypic changes of a given genotype in different environments (Jinks & Pooni 1988). This has been extensively studied in quantitative genetics (Falconer & Mackay 1996), evolutionary biology (Via & Lande 1985), breeding of livestock (Huquet *et al.* 2012), and plants (El-Soda *et al.* 2014), and in human medical genetics (Hutter *et al.* 2013). The second definition of environmental sensitivity is differences in the environmental variance of different genotypes in the same environment (Jinks & Pooni 1988). This second definition implies that there is a genetic component to environmental variance. While

heterogeneity of variance among genotypes have been known for a while, this venue of research have received little attention as compared to the effects of genetic variation on trait means, and even less devotion has been given to elucidate the genetic architecture of environmental variation. Only within the last decade or so have researchers started to realize and investigate this genetic control of the expression of V_E itself (**PAPER III**; Ros *et al.* 2004; Willmore *et al.* 2007; Ibáñez-Escriche *et al.* 2008; Ayroles *et al.* 2015; Morgante *et al.* 2015; Sørensen *et al.* 2015; Blasco *et al.* 2017). The two definitions of environmental sensitivity are not necessarily mutually exclusive, as they both describe mechanisms by which variable phenotypes arise from a uniform genetic background within and across environments. This combination of different aspects of the genetic control of V_E was the main research aim of **PAPER III**. Based on suggestions that multiple forms of V_E exists, and that such sources may be genetically independent (Hill & Mulder 2010), I suggested four different components of environmental variation encompassing variation both across environments, and conceptualized their computations. The four components were found to be heritable, and largely genetically decoupled, however, there were some exceptions of genetic correlations between different components. Our results suggest that the some of the components of V_E might represent separate selection targets with different constraints acting upon them, and some might in practice be indistinguishable by selection.

For **PAPER III**, I used the *Drosophila* Genetic Reference Panel (DGRP) (Mackay *et al.* 2012), which is a set of >200 lines, originating from a single wild-caught *D. melanogaster* population from North Carolina, USA. This population was initially sub-divided into lines, which were then extensively inbred through full-sibling matings, until essentially no genetic variation was left within each line, while maintaining the full extent of natural genetic variation between lines. The panel can be purchased from a stock center (http://flystocks.bio.indiana.edu), and full genome sequence data is available for each line (http://dgrp2.gnets.ncsu.edu). This unique resource allows researchers to investigate the correlation between phenotypic variation and genetic variation. I reared each line at five thermal environments and subsequently measured their cold tolerance. I exploited the fact that any variation in the phenotypic measures of multiple individuals from each line within and across environment is due to environmental variation, as there is practically no genetic variation within these lines. By measuring phenotypes of many individuals from the same line across the whole panel of lines, one can obtain a precise estimation of a lines performance. This can then be related to the sequence data in order to identify single genetic variants or genes associated with the phenotypic variation through genome-wide-association studies (GWAS) (Mackay *et al.* 2012; Huang *et al.* 2014), or to identify biological features (gene ontologies; GO) predictive of the trait value given the genotypic variation (Sarup *et al.* 2016; Edwards *et al.* 2016). Besides this

unique DGRP resource, there are numerous advantages to using *D. melanogaster* as a model organism to study this, beyond the well-known ease of maintenance, short generation times, and the immense knowledge base as reviewed by Jennings (2011). One such advantage is various ways to genetically modify this organism. For instance, a whole array of techniques has been developed to functionally validate the genes or genetic features identified in the association analyses described above. An example is by disrupting gene function by RNA-mediated gene interference (RNAi), where gene expression is supressed (Dietzl *et al.* 2007), which I used in **PAPER III**, to validate candidate genes for the different V_E components.

PLASTICITY AND ADAPTATION

Long-term phenotypic responses to environmental change is likely constituted by a mix of phenotypic plasticity and adaptive evolution. Phenotypic plasticity is one of the sources of environmental variation investigated in **PAPER III**, and perhaps the most studied form of V_E (DeWitt & Scheiner 2004; Valladares *et al.* 2006). Phenotypic plasticity is defined here as the ability of a given genotype to express different phenotypes depending on the environment. Phenotypic plasticity allows organisms to respond to rapid environmental changes to maintain overall fitness, and is believed to be an important determinant for the success of species under the environmental stress of anthropogenic climate change (Teplitsky *et al.* 2008; Hoffmann & Sgrò 2011; Anderson *et al.* 2012). In some cases phenotypic plasticity increase the fitness of an organism, a term referred to as adaptive phenotypic plasticity (Ghalambor *et al.* 2007). Under continuous environmental change, e.g. increasing temperature, a common way of characterizing phenotypic plasticity is as the norm-of-reaction of the phenotypic trait across the environmental gradient, and there are a myriad of different indices for phenotypic plasticity, each with different pros and cons (Valladares *et al.* 2006). For a linear reaction norm, as in the change in cold tolerance as a result of developmental temperature in **PAPER III**, slope of the linear regression is the most commonly used measure of phenotypic plasticity (Valladares *et al.* 2006). If traits are not displaying phenotypic plasticity, the reaction norm is horizontal. It can be costly for an organism to maintain a high phenotypic plasticity, as they must be flexible on a number of biological levels. Some argue that there is a trade-off between trait mean value and trait plasticity (Murren *et al.* 2015), and for stress resistance, hardening or acclimation which both can be considered plasticity, might constrain the organism's basal stress resistance (Stillman 2003; Calosi *et al.* 2008; Chown *et al.* 2010; Gerken *et al.* 2015). Congruent with a recent cross-taxa review (Gunderson *et al.* 2015), I found no evidence of a trade-off between basal cold tolerance and plasticity in **PAPER III**.

In some scenarios, e.g. with recent climate change, phenotypic plasticity might not be sufficient to maintain high fitness, and might be complemented or substituted by adaptive evolution instead. Adaptive evolution is characterized by a change in the genetic constitution of a population as a result of natural selection, thus in order to demonstrate the occurrence of adaptation, proof of genetic change and natural selection as the driving force is needed. This can prove challenging partly because precise estimates of natural selection can be hard to obtain (Kingsolver *et al.* 2012), and because the genetic architecture of many traits is still unknown (Anderson *et al.* 2014). Because of the perception of natural selection as a strong force, it was previously assumed by default that phenotypic changes were due to adaptive evolution. However, phenotypic plasticity is increasingly becoming the parsimonious (null) model (Merilä & Hendry 2014), which can be rejected with direct evidence of genetic change; in fact, some observations of phenotypic differences, that were initially assumed to be a result of genetic changes have subsequently been recognised as phenotypic plasticity (Charmantier *et al.* 2008; Teplitsky *et al.* 2008). Further complicating things is the fact that plasticity is heritable (**PAPER III**), and thus adaptive phenotypic plasticity itself can evolve as also suggested by earlier studies (Schlichting 1986; Stearns 1989; Scheiner & Lyman 1991), and it might not be easy to disentangle the two in natural or domestic populations (Gienapp *et al.* 2008). In any case, adaptive evolution, be it in trait means or traits plasticity, is dependent on available genetic variation, which in turn in dependent on populations sizes and inbreeding, as discussed below.

POPULATION SIZE AND INBREEDING

I have introduced genetic variation and environmental factors, and how they mutually interact, and for the remaining project of this thesis, I looked more into what determines genetic variation in a population, how the effects of inbreeding and low genetic variation can be alleviated, and lastly how loss of genetic variation affects adaptive evolution. A number of factors determine the amount of available genetic variation, one being the population size. In **PAPERS I-II**, I maintained fly cultures at a high number of individuals in a population, typically >500 individuals. In many natural and especially domestic populations, (effective) population sizes are smaller than this (Palstra & Ruzzante 2008). Because of finite population sizes, the sampling of genes, passed on to the next generation result in drifting allele frequencies, which is termed random genetic drift. This change in frequency is dependent on the starting frequencies of genetic variants and the number of samples (individuals) (Wright 1931). It follows, that it is not the number of individuals present in a population (called census population size, N), but rather the number of contributing individuals, and how well these individuals represent the gene variant frequencies in the original

populations, which determines the effective population size (N_e). As N_e is not easy to quantify because it is affected by reproduction and breeding strategies (inbreeding, outcrossing, asexual reproduction, sex ratio etc.) (Frankham 1995; Allendorf *et al.* 2013), N can be used to approximate N_e as a proxy for available genetic variation and selection efficiency (Wright 1931; Falconer & Mackay 1996; Frankham 2012). In our laboratory we normally keep N>500 to minimize genetic drift. Genetic drift will over generations lead to changes in allele frequencies (loss and fixation of alleles) and increased homozygosity at a rate that depends on N_e (Garner *et al.* 2005). If fixed loci are associated with phenotypic variation, genetic drift will result in drifting trait values (Falconer & Mackay 1996), which in combination with loss of genetic variation resulting from a small population size can lead to a decreased ability to adapt to a stressful environment (Charlesworth & Charlesworth 1987; Frankham *et al.* 1999; Willi *et al.* 2006; Hoffmann & Willi 2008). This is one of major concerns with the generally increasing stress levels experienced by many populations, e.g. under recent climate change and is one the primary reasons for investigating the effects of small populations sizes. Some evidence suggests that associations between the N_e and the amount of genetic variation is more complex than previously assumed (Bouzat 2010; Wood *et al.* 2016; Hoffmann *et al.* 2017), necessitating further studies in this highly relevant field of research (see additional results presented at the end of this thesis).

In studies of small and fragmented population, inbreeding is also a large concern. Inbreeding is most commonly defined as the non-random mating among related individuals, and the coefficient of inbreeding (F), designates the probability that the two alleles at a given locus in the offspring are both inherited from a common ancestor, so-called identical by descent (Falconer & Mackay 1996; Frankham *et al.* 2013). One effect of inbreeding is an increase in homozygosity within the population (Hartl & Clark 2007). A direct consequence of the increase in homozygosity following inbreeding is the increased expression of rare recessive deleterious alleles (Falconer & Mackay 1996). This often leads to a decrease in the fitness of inbred populations relative to outbred populations, a phenomenon termed inbreeding depression, and is often reported in natural populations (Crnokrak & Roff 1999; Frankham *et al.* 2013; Hoffman *et al.* 2014). A reduction in fitness caused by inbreeding depression, genetic load or reproductive incompatibility is sometimes referred to as ‘genetic stress’ (Pertoldi *et al.* 2006; Willi *et al.* 2006). Genetic stress as a result of inbreeding depression is an important theme of this thesis as such intrinsic genetic stress can interact with external environmental stress. One such interaction is the well known inbreeding-by-environment interaction (I x E), where the effects of inbreeding is dependent on environmental conditions, and is often reported in scenarios where the deleterious effects of inbreeding depression is exacerbated under environmental stress (Armbruster & Reed 2005; Schou *et al.* 2015), with proposed large impacts for small

inbred populations under existing environmental stresses (Fox & Reed 2011; Reed *et al.* 2012).

Understanding the causes and consequences of inbreeding depression is central in population biology (Armbruster & Reed 2005), including the evolution of mating systems (Charlesworth & Charlesworth 1987; Uyenoyama *et al.* 1993), animal- and plant breeding programs (Falconer & Mackay 1996), and the conservation of rare and extinction prone populations (Crnokrak & Roff 1999; Hedrick & Kalinowski 2000; Reed & Frankham 2003). At the end of this thesis I present the results of an ongoing study (described below), which provide additional insights into the relationship between population bottlenecks, inbreeding and adaptive capacity.

THE ADAPTIVE POTENTIAL OF SMALL POPULATIONS

Faced with the plethora of environmental stresses as described above, the long-term persistency of natural population will ultimately depend on their ability to respond either through plasticity and/or evolutionary changes. It has for long been theorized that populations with low genetic variation will have a lower evolutionary potential (Fisher 1958), and this has since been a central topic of debate in evolutionary biology and conservation genetics. Recent studies provide evidence that some populations are evolutionary constrained in ecologically important stress resistance traits (Kelly *et al.* 2012; Kellermann *et al.* 2012; Araújo *et al.* 2013; Hoffmann *et al.* 2013; Schou *et al.* 2014; Kristensen *et al.* 2015). The availability of genetic variation relevant for adaptation in populations is frequently measured by the additive genetic variance V_A of the trait in question, typically expressed as the heritability (h^2) or evolvability of the trait. While theory predicts a relationship between N_e and V_A (Falconer & Mackay 1996; Willi *et al.* 2006), there is considerable ambiguity in the empirical evidence of the relationship between population size, genetic variation, and evolutionary potential (Bouzat 2010; Wood *et al.* 2016; Hoffmann *et al.* 2017). While some studies find that larger populations respond faster to selection in morphological traits (Jones *et al.* 1969; Weber 1990) and stress tolerance (Weber & Diggins 1990), the meta-analysis by Wood *et al.* (2016) suggested a poor association between population size and adaptive potential. Some evidence from laboratory experiments with insects suggest that inbreeding due to low Ne, which also increases genetic drift, reduces V_A and heritability estimates (Saccheri *et al.* 2001; Kristensen *et al.* 2005; Dierks *et al.* 2012). However a meta-analysis of experimental studies investigating the association between inbreeding levels and V_A conclude that V_A are not reduced with increasing inbreeding to the extent predicted from theory (Taft & Roff 2012). Contrary to theoretical expectations, some studies suggest that that population bottlenecks can in fact increase V_A (Taft & Roff 2012), but not necessarily increase response to selection (Van Heerwaarden *et al.* 2008). The connection between heritability and evolutionary

response is also unresolved. Many of the studies reviewed by Wood *et al.* (2016) investigate morphological traits, which tend to have high heritability estimates and uncertain connections to fitness. Thus, such studies might not correctly reflect the evolvability of important fitness components in natural populations, where low heritabilities are common (Carlson & Seamons 2008; Hansen *et al.* 2011). Also, the fact that heritability estimates are inherently noisy (Hansen *et al.* 2011) especially for traits that are highly responsive to environmental variability (Hoffmann *et al.* 2017), highlights the necessity for large sample sizes or highly replicated inbreeding designs to yield reliable estimates for low heritability traits (Hoffmann *et al.* 2016). This is problematic because traits with low variances is arguably the most interesting in terms of conservation, because low heritabilities can suggest a constraint on further adaptation that would have otherwise allowed populations to evolve to overcome e.g. current fast climate change (Hoffmann *et al.* 2017). In addition, the levels of inbreeding is perhaps unrealistically high inmany experimental studies (Pemberton *et al.* 2017), complicating comparisons with natural populations, and contributing to the complexity of the relationship between inbreeding, fitness, genetic diversity, adaptive capability, and extinction risk.

In response to some of the abovementioned ambiguities and the current lack of large-scale empirical evidence on the connection between adaptive potential and inbreeding and loss of genetic variation, I set up a laboratory evolution experiment. As the experimental work was finalized only a few weeks prior to the completion of this thesis, I will present the preliminary analyses and results of this ongoing work (under 'Additional results'). This work is not yet formulated into a full manuscript. The experiment was set up with ~120 lines of *D. melanogaster* inbred to three different F levels (40 at each level), by undergoing a varying number of population bottlenecks. These lines and outbred control lines were reared on in novel stressful environment for 10 generations while productivity and body size was assessed every generation. The initial analyses suggest that the results generally supported the expectations, that increasing levels of inbreeding lead to reduced evolutionary response to selection, however there was a large degree of line specificity, emphasizing the need for a large number of replicated lines in such studies. I also found highly trait specific responses among the lines of the different inbreeding levels. Across all inbreeding levels, there was a significant positive correlation between nucleotide diversity and selection response measured as the slope of the respective traits across generations. Assessment of viability before and after selection indicated that inbred lines performed better in the stressful environment, while they performed slightly worse in a benign environment, suggestive of an evolutionary trade-off.

MANAGING POPULATIONS WITH LOW GENETIC VARIATION

Detrimental effects of inbreeding and low genetic variation are commonly reported in natural populations (Charlesworth & Charlesworth 1987; Crnokrak & Roff 1999; Charpentier *et al.* 2005; Da Silva *et al.* 2006; Hanski & Saccheri 2006; Fox *et al.* 2008; Grueber *et al.* 2008; Frankham *et al.* 2013; Hoffman *et al.* 2014). In genetically deteriorated populations suffering from inbreeding depression and potentially low evolutionary potential (as described in the experiment above), human intervention in the form of conservation management may be necessary to prevent extinction. Different management strategies are being employed in the conservation of endangered populations including artificial feeding, fencing, fostering of offspring, vaccination, culling, and/or management of environments such as preserving or restoring habitats and establishing corridors between suitable habitats. Different management strategies are reviewed in e.g. Bodini *et al.* (2008). In small and fragmented populations genetic management might be necessary to re-establish gene flow e.g. by translocating individuals or genetic material from a donor population to the genetically deteriorated recipient population (Edmands *et al.* 2003; Frankham 2010). The beneficial outcome of such outcrossing have been reported as increased long-term survival, reproduction, population growth and reduced extinction risk (Madsen *et al.* 1999; Marr *et al.* 2002; Pimm *et al.* 2006; Bijlsma *et al.* 2010; Miller *et al.* 2012; Hufbauer *et al.* 2015). An increase in population fitness due to immigration of new alleles, is sometimes referred to as 'genetic rescue' (Whiteley et al., 2015), and is the central theme of **PAPER IV**.

The positive effects of genetic rescue are primarily caused by heterosis (or hybrid vigour), which is the outperformance of the hybrid offspring compared to the mean of parents. The result is an increase in population fitness compared to the original population prior to outcrossing, and as with inbreeding depression, the greatest effects are typically seen in traits closely related to fitness (Tallmon et al., 2004; Whiteley et al., 2015). This is due to the optimization of fitness related traits, and resulting higher degree of non-additive genetic variation essential for the expression of heterosis. The positive effect of heterosis is particularly utilized in animal and plant breeding to enhance the performance in specific production traits such as yield, disease resistance, and the production of uniform phenotypes (Kawamura et al., 2016; Pandey et al., 2015; Solieman et al., 2013; Sørensen et al., 2008). Genetic rescue is however not always preferable and some concerns have been raised in its use for managing wild populations (Tallmon *et al.* 2004; Frankham *et al.* 2011; Hedrick & Garcia-Dorado 2016). One concern is the risk of outbreeding depression, caused by either local adaptive differences between immigrants and the resident population (Allendorf *et al.* 2001; Edmands *et al.* 2003) or by the disruption of beneficial interactions of co-

adapted gene complexes between loci in linkage-disequilibrium (Templeton *et al.* 1986; Allendorf *et al.* 2001; Montalvo & Ellstrand 2001).

To gain most heterosis, and thus the largest fitness enhancements and due to the risk of outbreeding depression, knowledge on the genetic relatedness of donor and recipient population is imperative. For instance, genetic distance (GD) between the two populations can be predictive of the magnitude of heterosis and fitness benefits (Mohamed & Pirchner 1998; Pandey *et al.* 2015). This was the central research aim of **PAPER IV**, where I again used the DGRP resource to simulate populations in need of genetic rescue, to investigate the effects of GD on the temporal effects of heterosis. The benefits of the DGRP system in this context was that I could calculate precise genetic distances based on many-fold more genetic markers, than have previously been used in studies of the correlation between GD and heterosis (Goddard & Ahmed 1982; Graml & Pirchner 1984; Ehiobu *et al.* 1990; Mohamed & Pirchner 1998; Geleta *et al.* 2004; Singh & Singh 2004; Teklewold & Becker 2006; Pandey *et al.* 2015; Kawamura *et al.* 2016). In addition, I had information on each DGRP line's performance in a range of traits, allowing me to identify lines likely to suffer the most from inbreeding depression to represent weak lines in need of genetic rescue. The results of **PAPER IV** clearly demonstrated genetic rescue as a viable conservation management strategy with large fitness benefits in the hybrid offspring, however this study also revealed potential caveats with genetic translocation as the magnitude of heterosis decreased from the first to the third generation. Overall, GD had little effect on the amount of expressed heterosis, while other measures turned out to be better predictors of heterosis, e.g. the phenotypic difference between parents used in the outcrossing, as also suggested by others (Teklewold & Becker 2006).

CONCLUSIONS AND PERSPECTIVE

Many studies on the effects of environmental stresses, especially studies investigating the effects of climate change as a potential stress, tend to focus on changes in mean and variability of a single parameter, e.g. temperature. However, several papers in this thesis highlight the fact that environmental stresses should not be considered in isolation, on the contrary most species are exposed to multiple environmental conditions simultaneously, and environmental factors are likely to interact, either with other environmental stresses, and/or with the sex or the genetic constitution of individuals. Such biological interactions, whether they are interactions between multiple environmental conditions (**PAPER I**), interactions between sex and environment (**PAPER II**), or genotype-by-environment interactions (**PAPER III**), is important for our understanding of how multiple stresses impact ecosystem resilience (Folke *et al.* 2004) and for projections of biodiversity (Sala *et al.* 2000; Pereira *et al.* 2010). This knowledge is also vital for our ability to predict and eliminate ecological

surprises (Paine *et al.* 1998; Didham *et al.* 2007; Mora *et al.* 2007). Although environmental interactions do occur, it should not be assumed by default that exposure to multiple stresses is worse (or better) than the individual environmental conditions alone. An important conclusion of **PAPER I** was that the additive sum of individual environmental conditions was more common than interactions, and that the seemingly prevalent notion that synergistic interactions are omnipresent is perhaps a result of bias towards publishing 'positive' results.

In **PAPER II**, I concluded that males and females displayed different degrees of phenotypic plasticity, resulting in the degree of sexual dimorphism of the metabolite composition being environment dependent. Although only suggestive, I postulated that the direction and magnitude of sexual selection may be environment dependent as well, an aspect that deserves much more attention, especially in the light of the maintenance of sexual dimorphism in traits relevant to stress resistance and population viability. Further studying how environmental stresses affect biosynthesis and metabolism may provide new insight regarding 'mode of action' and can possibly clarify responses observed on a higher biological level. Such studies also may help in the development of alternative endpoints for toxicity testing (Pablos *et al.* 2015; Ørsted & Roslev 2015), and techniques for rapid screening of environments for extracellular compounds indicative of a general stress response of the ecosystem.

In **PAPER III**, I dove more into what determines phenotypic variation within and across environments, and conceptualized four different components. The results delineate selection targets associated with environmental variation and the constraints acting upon them, offering a backdrop for applied evolutionary studies on environmental sensitivity. This decomposition of the genetic control of environmental variation extends well beyond what has been attempted before. I partly consider this work as a proof-of-concept, and hope that this study can work as a hypothesis-generating platform motivating future studies elucidating the nature of the evolutionary forces maintaining segregating variation for each V_E component and how they are interrelated. I think that animal breeders having access to extremely large dataset have a lot to offer in this context (Hill & Mulder 2010; Sanchez-Garcia *et al.* 2012; Rönnegård *et al.* 2013). I applied the deconstructed V_E terms to cold resistance as a function of thermal rearing conditions, but it is my intention and hope, that the concepts are broadly applicable, and it will be interesting to see how they adopt to other traits in other environments.

In **PAPER IV**, I show that the effects of inbreeding and low genetic variation can be somewhat alleviated through genetic rescue, which is viable management strategy for the conservation of small and fragmented populations to increase fitness. However, I also found that the level of heterosis declined strongly over time, and thus the results suggest that in genetic rescue projects, continuous translocations may be necessary to maintain fitness benefits of the outcrossing in hope of preventing

extinction in the future. Despite the results suggesting that genetic distance did not have a large effect on the amount of expressed heterosis, the study proposed that other measures, e.g. parental phenotypic distance may be a better predictor of heterosis (Teklewold & Becker 2006), i.e. in our study lower fitness in the crossed population led to higher heterosis. This means that populations suffering from inbreeding depression have the potential to gain the most from a genetic rescue operation. Lastly, I will add that although translocations can result in fitness increases of populations on the verge of extinction, the long-term persistence of populations in the wild will depend on the availability of suitable habitat, so unless the conditions that led to the species decline in the first place are reversed, e.g. through environmental restoration, the efforts of genetic rescue will be futile (Bouzat *et al.* 2009).

The additional results presented at the end suggested a connection between population bottlenecks and resulting inbreeding levels, and adaptive potential. Further analyses will elucidate the full extent of the adaptive responses, e.g. whether a linear regression is the best model to describe the selection response, since the response for productivity seem non-linear with rapid early evolution followed by plateauing responses, perhaps indicating a fast initial depletion of V_A and reaching a selection limit. Furthermore, the results can perhaps help disentangle the effects of inbreeding from the effects of the effective population size, although for now this is only on a conceptual level, and needs to be further explored.

REFERENCES

Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001) The problems with hybrids: setting conservation guidelines. *Trends in Ecology and Evolution*, **16**, 613–622.

Allendorf FW, Luikart G, Aitken SN (2013) *Conservation and the genetics of population*. Wiley-Blackwell, Hoboken, NJ, USA.

Allison EH, Bassett HR (2015) Climate change in the oceans: Human impacts and responses. *Science*, **350**, 778–782.

Anderson JT, Inouye DW, McKinney AM, Colautti RI, Mitchell-Olds T (2012) Phenotypic plasticity and adaptive evolution contribute to advancing flowering phenology in response to climate change. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 3843–3852.

Anderson JT, Wagner MR, Rushworth CA, Prasad KVSK, Mitchell-Olds T (2014) The evolution of quantitative traits in complex environments. *Heredity*, **112**, 4–12.

Angilletta MJ (2009) *Thermal Adaptation: a Theoretical and Empirical Synthesis*. Oxford University Press, Oxford, UK.

Ankley GT, Daston GP, Degitz SJ *et al.* (2006) Toxicogenomics in regulatory ecotoxicology. *Environmental Science and Technology*, **40**, 4055–4065.

Araújo MB, Ferri-Yáñez F, Bozinovic F *et al.* (2013) Heat freezes niche evolution. *Ecology Letters*, **16**, 1206–1219.

Armbruster P, Reed DH (2005) Inbreeding depression in benign and stressful environments. *Heredity*, **95**, 235–242.

Ayroles JF, Buchanan SM, O'Leary C *et al.* (2015) Behavioral idiosyncrasy reveals genetic control of phenotypic variability. *Proceedings of the National Academy of Sciences*, **112**, 6706–6711.

Bednarska AJ, Jevtić DM, Laskowski R (2013) More ecological ERA: incorporating natural environmental factors and animal behavior. *Integrated environmental assessment and management*, **9**, 39-46.

Bijlsma R, Westerhof MDD, Roekx LP, Pen I (2010) Dynamics of genetic rescue in inbred *Drosophila melanogaster* populations. *Conservation Genetics*, **11**, 449–462.

Blasco A, Martínez-Álvaro M, García M-L, Ibáñez-Escriche N, Argente M-J (2017) Selection for environmental variance of litter size in rabbits. *Genetics Selection Evolution*, **49**, 48.

Bodini A, Baumgärtner J, Gilioli G (2008) Conservation strategies evaluation in an adaptive management framework. *Animal Conservation*, **11**, 472–475.

Bouzat JL (2010) Conservation genetics of population bottlenecks: The role of chance, selection, and history. *Conservation Genetics*, **11**, 463–478.

Bouzat JL, Johnson JA, Toepfer JE *et al.* (2009) Beyond the beneficial effects of translocations as an effective tool for the genetic restoration of isolated populations. *Conservation Genetics*, **10**, 191–201.

Brooker RW, Maestre FT, Callaway RM *et al.* (2008) Facilitation in plant communities: The past, the present, and the future. *Journal of Ecology*, **96**, 18–34.

Callaway RM, Brooker RW, Choler P *et al.* (2002) Positive interactions among alpine plants increase with stress. *Nature*, **417**, 844–848.

Calosi P, Bilton DT, Spicer JI (2008) Thermal tolerance, acclimatory capacity and vulnerability to global climate change. *Biology Letters*, **4**, 99–102.

Carlson SM, Seamons TR (2008) A review of quantitative genetic components of fitness in salmonids: implications for adaptation to future change. *Evolutionary Applications*, **1**, 222–238.

Charlesworth D, Charlesworth B (1987) Inbreeding depression and its evolutionary consequences. *Annual Review of Ecology and Systematics*, **18**, 237–268.

Charmantier A, McCleery RH, Cole LR *et al.* (2008) Adaptive phenotypic plasticity in response to climate change in a wild bird population. *Science*, **320**, 800–803.

Charpentier MJE, Setchell JM, Prugnolle F *et al.* (2005) Genetic diversity and reproductive success in mandrills (*Mandrillus sphinx*). *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 16723–16728.

Chown SL, Hoffmann AA, Kristensen TN *et al.* (2010) Adapting to climate change: a perspective from evolutionary physiology. *Climate Research*, **43**, 3–15.

Crain CM, Kroeker K, Halpern BS (2008) Interactive and cumulative effects of multiple human stressors in marine systems. *Ecology Letters*, **11**, 1304–1315.

Crnokrak P, Roff DA (1999) Inbreeding depression in the wild. *Heredity*, **83**, 260–270.

Darling ES, Côté IM (2008) Quantifying the evidence for ecological synergies. *Ecology Letters*, **11**, 1278–1286.

DeWitt T, Scheiner S (2004) *Phenotypic plasticity, functional and conceptual approaches*. Oxford University Press, Oxford, UK.

Didham RK, Tylianakis JM, Gemmell NJ, Rand TA, Ewers RM (2007) Interactive effects of habitat modification and species invasion on native species decline. *Trends in Ecology and Evolution*, **22**, 489–496.

Dierks A, Baumann B, Fischer K (2012) Response to selection on cold tolerance is constrained by inbreeding. *Evolution*, **66**, 2384–2398.

Dietzl G, Chen D, Schnorrer F *et al.* (2007) A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. *Nature*, **448**, 151–157.

Edmands S, Timmerman CC, Timmermant CC (2003) Modeling factors affecting the severity of outbreeding depression. *Conservation Biology*, **17**, 883–892.

Edwards SM, Sørensen IF, Sarup P, Mackay TFC, Sørensen P (2016) Genomic prediction for quantitative traits is improved by mapping variants to gene ontology categories in *Drosophila melanogaster*. *Genetics*, **203**, 1871–1883.

Ehiobu NG, Goddard ME, Taylor JF (1990) Prediction of heterosis in crosses between inbred lines of *Drosophila melanogaster*. *TAG: Theoretical and Applied Genetics*, **80**, 321–325.

El-Soda M, Malosetti M, Zwaan BJ, Koornneef M, Aarts MGM (2014) Genotype x environment interaction QTL mapping in plants: lessons from *Arabidopsis*. *Trends in Plant Science*, **19**, 390–398.

Falconer DS, Mackay TFC (1996) *Introduction to Quantitative Genetics*. Longman Publishing Group, Harlow, UK.

Fisher RA (1958) *The Genetical Theory of Natural Selection*. Dover Publication, Mineola, NY, USA.

Folke C, Carpenter S, Walker B *et al.* (2004) Regime shifts, resilience, and biodiversity in ecosystem management. *Annual Review of Ecology, Evolution, and Systematics*, **35**, 557–581.

Folt C, Chen C, Moore M, Burnaford J (1999) Synergism and antagonism among multiple stressors. *Limnology and Oceanography*, **44**, 864–877.

Forbes VE, Palmqvist A, Bach L (2006) The use and misuse of biomarkers in ecotoxicology. *Environmental Toxicology and Chemistry*, **25**, 272–280.

Fox CW, Reed DH (2011) Inbreeding depression increases with environmental stress: an experimental study and meta-analysis. *Evolution*, **65**, 246–258.

Fox CW, Scheibly KL, Reed DH (2008) Experimental evolution of the genetic load and its implications for the genetic basis of inbreeding depression. *Evolution*, **62**, 2236–2249.

Frankham R (1995) Effective population size/adult population size ratios in wildlife: a review. *Genetical Research*, **66**, 95–107.

Frankham R (2010) Challenges and opportunities of genetic approaches to biological conservation. *Biological Conservation*, **143**, 1919–1927.

Frankham R (2012) How closely does genetic diversity in finite populations conform to predictions of neutral theory? Large deficits in regions of low recombination. *Heredity*, **108**, 167–178.

Frankham R, Ballou JD, Briscoe DA (2013) *Introduction to Conservation Genetics*. Cambridge University Press, Cambridge, UK.

Frankham R, Ballou JD, Eldridge MDB *et al.* (2011) Predicting the probability of outbreeding depression. *Conservation Biology*, **25**, 465–475.

Frankham R, Lees K, Montgomery ME *et al.* (1999) Do population size bottlenecks reduce evolutionary potential? *Animal Conservation*, **2**, 255–260.

Garner A, Rachlow JL, Hicks JF (2005) Patterns of genetic diversity and its loss in mammalian populations. *Conservation Biology*, **19**, 1215–1221.

Geleta LF, Labuschagne MT, Viljoen CD (2004) Relationship between heterosis and genetic distance based on morphological traits and AFLP markers in pepper. *Plant Breeding*, **123**, 467–473.

Gerken AR, Eller OC, Hahn DA, Morgan TJ (2015) Constraints, independence, and evolution of thermal plasticity: probing genetic architecture of long- and short-term thermal acclimation. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 4399–4404.

Ghalambor CK, McKay JK, Carroll SP, Reznick DN (2007) Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments. *Functional Ecology*, **21**, 394–407.

Gienapp P, Teplitsky C, Alho JS, Mills JA, Merilä J (2008) Climate change and evolution: disentangling environmental and genetic responses. *Molecular Ecology*, **17**, 167–178.

Goddard ME, Ahmed AM (1982) The use of the genetic distance between cattle breeds to predict the heterosis in crosses. In: *Proceedings of the 2nd World Congress on Genetics applied to Livestock Production, Madrid (Vol. 8)*, pp. 377–382.

Graml R, Pirchner F (1984) Relation of genetic distance between cattle breeds and heterosis of resulting crosses. *Animal Blood Groups and Biochemical Genetics*, **15**, 173–180.

Grueber CE, Wallis GP, Jamieson IG (2008) Heterozygosity-fitness correlations and their relevance to studies on inbreeding depression in threatened species. *Molecular Ecology*, **17**, 3978–3984.

Gunderson AR, Stillman JH, Gunderson AR (2015) Plasticity in thermal tolerance has limited potential to buffer ectotherms from global warming. *Proceedings of the Royal Society B: Biological Sciences*, **282**, 1–8.

Halpern BS, Selkoe KA, Micheli F, Kappel C V. (2007) Evaluating and ranking the vulnerability of global marine ecosystems to anthropogenic threats. *Conservation Biology*, **21**, 1301–1315.

Halpern BS, Walbridge S, Selkoe KA *et al.* (2008) A global map of human impact on marine ecosystems. *Science*, **319**, 948–52.

Hansen TF, Pélabon C, Houle D (2011) Heritability is not evolvability. *Evolutionary Biology*, **38**, 258–277.

Hanski I, Saccheri I (2006) Molecular-level variation affects population growth in a butterfly metapopulation. *PLOS Biology*, **4**, 719–726.

Hartl DL, Clark AG (2007) *Principles of Population Genetics*. Sinauer Associates, Sunderland, UK.

He Q, Bertness MD, Altieri AH (2013) Global shifts towards positive species interactions with increasing environmental stress. *Ecology Letters*, **16**, 695–706.

Hedrick PW, Garcia-Dorado A (2016) Understanding inbreeding depression, purging, and genetic rescue. *Trends in Ecology and Evolution*, **31**, 940–952.

Hedrick PW, Kalinowski ST (2000) Inbreeding depression in conservation biology. *Annual Review of Ecology and Systematics*, **31**, 139–162.

Van Heerwaarden B, Willi Y, Kristensen TN, Hoffmann AA (2008) Population bottlenecks increase additive genetic variance but do not break a selection limit in rain forest *Drosophila*. *Genetics*, **179**, 2135–2146.

Hill WG, Mulder HA (2010) Genetic analysis of environmental variation. *Genetics Research*, **92**, 381–395.

Hoffman JI, Simpson F, David P *et al.* (2014) High-throughput sequencing reveals inbreeding depression in a natural population. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 3775–3780.

Hoffmann A a., Chown SL, Clusella-Trullas S (2013) Upper thermal limits in terrestrial ectotherms: how constrained are they?. *Functional Ecology*, **27**, 934–949.

Hoffmann AA, Hallas R, Anderson AR, Telonis-Scott M (2005) Evidence for a robust sex-specific trade-off between cold resistance and starvation resistance in *Drosophila melanogaster*. *Journal of Evolutionary Biology*, **18**, 804–810.

Hoffmann AA, Merilä J, Kristensen TN (2016) Heritability and evolvability of fitness and nonfitness traits: lessons from livestock. *Evolution*, **70**, 1770–1779.

Hoffmann AA, Parsons PA (1991) *Evolutionary Genetics and Environmental Stress*. Oxford University Press, Oxford, UK.

Hoffmann A, Sgrò C (2011) Climate change and evolutionary adaptation. *Nature*, **470**, 479–485.

Hoffmann AA, Sgrò CM, Kristensen TN (2017) Revisiting adaptive potential, population size, and conservation. *Trends in Ecology and Evolution*, **32**, 506–517.

Hoffmann AA, Willi Y (2008) Detecting genetic responses to environmental change. *Nature Reviews Genetics*, **9**, 421–432.

Holmstrup M, Bindesbøl AM, Oostingh GJ *et al.* (2010) Interactions between effects of environmental chemicals and natural stressors: a review. *Science of the Total Environment*, **408**, 3746–3762.

Huang W, Massouras A, Inoue Y *et al.* (2014) Natural variation in genome architecture among 205 *Drosophila melanogaster* Genetic Reference Panel lines. *Genome Research*, **24**, 1193–1208.

Hufbauer RA, Szűcs M, Kasyon E *et al.* (2015) Three types of rescue can avert extinction in a changing environment. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 10557–10562.

Huquet B, Leclerc H, Ducrocq V (2012) Modelling and estimation of genotype by environment interactions for production traits in French dairy cattle. *Genetics Selection Evolution*, **44**, 35.

Hutter CM, Mechanic LE, Chatterjee N *et al.* (2013) Gene-environment interactions in cancer epidemiology: a national cancer institute think tank report. *Genetic Epidemiology*, **37**, 643–657.

Ibáñez-Escriche N, Varona L, Sorensen D, Noguera JL (2008) A study of heterogeneity of environmental variance for slaughter weight in pigs. *Animal*, **2**, 19–26.

IPCC (2013) *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (TF Stocker, D Qin, G-K Plattner, et al., Eds,). Cambridge University Press, Cambridge, UK.

ISO (2012) ISO 6341. Water quality – Determination of the inhibition of the mobility of *Daphnia magna* Straus (Cladocera, Crustacea) – Acute toxicity test. International Organization of Standardization, Geneve, Switzerland.

Jennings BH (2011) *Drosophila* - a versatile model in biology & medicine. *Materials Today*, **14**, 190–195.

Jinks JL, Pooni HS (1988) The genetic basis of environmental sensitivity. In: *Proceedings of the 2nd International Conference on Quantitative Genetics* (eds Weir BS, Eisen EJ, Goodman MM, Namkoong G), pp. 505–522. Sinauer

Associates, Sunderland, UK.

Jones LP, Frankham R, Sheridan AK (1969) Correlation between bristle systems in *Drosophila melanogaster*. *Australian Journal of Biological Sciences*, **22**, 1473–1484.

Kaunisto S, Ferguson L V., Sinclair BJ (2016) Can we predict the effects of multiple stressors on insects in a changing climate? *Current Opinion in Insect Science*, **17**, 55–61.

Kawamura K, Kawanabe T, Shimizu M *et al.* (2016) Genetic distance of inbred lines of Chinese cabbage and its relationship to heterosis. *Plant Gene*, **5**, 1–7.

Kéfi S, Holmgren M, Scheffer M (2016) When can positive interactions cause alternative stable states in ecosystems? *Functional Ecology*, **30**, 88–97.

Kellermann V, Overgaard J, Hoffmann AA *et al.* (2012) Upper thermal limits of *Drosophila* are linked to species distributions and strongly constrained phylogenetically. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 16228–33.

Kelly MW, Sanford E, Grosberg RK (2012) Limited potential for adaptation to climate change in a broadly distributed marine crustacean. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 349–56.

Ketola T, Kristensen TN, Kellermann VM, Loeschcke V (2012) Can evolution of sexual dimorphism be triggered by developmental temperatures? *Journal of Evolutionary Biology*, **25**, 847–855.

Ketola T, Mikonranta L, Zhang J *et al.* (2013) Fluctuating temperature leads to evolution of thermal generalism and preadaptation to novel environments. *Evolution*, **67**, 2936–2944.

Kingsolver JG, Diamond SE, Siepielski AM, Carlson SM (2012) Synthetic analyses of phenotypic selection in natural populations: lessons, limitations and future directions. *Evolutionary Ecology*, **26**, 1101–1118.

Kristensen TN, Overgaard J, Lassen J, Hoffmann AA, Sgrò C (2015) Low evolutionary potential for egg-to-adult viability in *Drosophila melanogaster* at high temperatures. *Evolution*, **69**, 803–814.

Kristensen TN, Sørensen AC, Sorensen D *et al.* (2005) A test of quantitative genetic theory using *Drosophila* - effects of inbreeding and rate of inbreeding on heritabilities and variance components. *Journal of Evolutionary Biology*, **18**, 763–770.

De Lange HJ, Sala S, Vighi M, Faber JH (2010) Ecological vulnerability in risk assessment - a review and perspectives. *Science of The Total Environment*, **408**, 3871–3879.

Laskowski R, Bednarska AJ, Kramarz PE *et al.* (2010) Interactions between toxic chemicals and natural environmental factors - a meta-analysis and case studies. *Science of the Total Environment*, **408**, 3763–3774.

Lynch M, Walsh B (1998) *Genetics and analysis of quantitative traits*. Sinauer Associates, Sunderland, UK.

Mackay TFC, Richards S, Stone EA *et al.* (2012) The *Drosophila melanogaster* genetic reference panel. *Nature*, **482**, 173–178.

Madsen T, Shine R, Olsson M, Wittzell H (1999) Restoration of an inbred adder population. *Nature*, **402**, 34–35.

Manenti T, Sørensen JG, Moghadam NN, Loeschcke V (2014) Predictability rather than amplitude of temperature fluctuations determines stress resistance in a natural population of *Drosophila simulans*. *Journal of Evolutionary Biology*, **27**, 2113–2122.

Marr AB, Keller LF, Arcese P (2002) Heterosis and outbreeding depression in descendants of natural immigrants to an inbred population of song sparrows (*Melospiza melodia*). *Evolution*, **56**, 131–142.

Merilä J, Hendry AP (2014) Climate change, adaptation, and phenotypic plasticity: The problem and the evidence. *Evolutionary Applications*, **7**, 1–14.

Miller JM, Poissant J, Hogg JT, Coltman DW (2012) Genomic consequences of genetic rescue in an insular population of bighorn sheep (*Ovis canadensis*). *Molecular Ecology*, **21**, 1583–1596.

Mohamed SA, Pirchner F (1998) Relationships between genetic distance and dominance and epistatic effects in line-crosses of mice. *Journal of Animal Breeding and Genetics*, **115**, 313–321.

Montalvo AM, Ellstrand NC (2001) Nonlocal transplantation and outbreeding depression in the subshrub *Lotus scoparius* (Fabaceae). *American Journal of Botany*, **88**, 258–269.

Mora C, Metzger R, Rollo A, Myers RA (2007) Experimental simulations about the effects of overexploitation and habitat fragmentation on populations facing environmental warming. *Proceedings of the Royal Society B: Biological Sciences*, **274**, 1023–1028.

Morgante F, Sørensen P, Sorensen D, Maltecca C, Mackay TFC (2015) Genetic architecture of micro-environmental plasticity in *Drosophila melanogaster*. *Scientific Reports*, **5**, 9785.

Murren CJ, Auld JR, Callahan H *et al.* (2015) Constraints on the evolution of phenotypic plasticity: limits and costs of phenotype and plasticity. *Heredity*, **115**, 293–301.

Nicholson JK, Lindon JC, Holmes E (1999) "Metabonomics": understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica*, **29**, 1181–1189.

Novacek MJ, Cleland EE (2001) The current biodiversity extinction event: scenarios for mitigation and recovery. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 5466–5470.

Pablos M V, Garcia-Hortiguela P, Fernandez C (2015) Acute and chronic toxicity of emerging contaminants, alone or in combination, in *Chlorella vulgaris* and *Daphnia magna*. *Environmental Science and Pollution Research*, **22**, 5417–5424.

Paine RT, Tegner MJ, Johnson EA(1998) Compounded perturbations yield ecological surprises. *Ecosystems*, **1**, 535–545.

Palstra FP, Ruzzante DE (2008) Genetic estimates of contemporary effective

population size: what can they tell us about the importance of genetic stochasticity for wild population persistence? *Molecular Ecology*, **17**, 3428–3447.

Pandey P, Pandey VR, Kumar A *et al.* (2015) Relationship between heterosis and genetic diversity in Indian pigeonpea [*Cajanus cajan* (L.) Millspaugh] accessions using multivariate cluster analysis and heterotic grouping. *Australian Journal of Crop Science*, **9**, 494–503.

Parmesan C, Yohe G (2003) A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, **421**, 37–42.

Partridge L, Gems D, Withers DJ (2005) Sex and death: what is the connection? *Cell*, **120**, 461–472.

Pemberton JM, Ellis PE, Pilkington JG, Bérénos C (2017) Inbreeding depression by environment interactions in a free-living mammal population. *Heredity*, **118**, 64–77.

Pereira HM, Leadley PW, Proença V *et al.* (2010) Scenarios for global biodiversity in the 21st century. *Science*, **330**, 1496–1501.

Pertoldi C, Kristensen TN, Andersen DH, Loeschcke V (2006) Developmental instability as an estimator of genetic stress. *Heredity*, **96**, 122–127.

Piggott JJ, Townsend CR, Matthaei CD (2015) Reconceptualizing synergism and antagonism among multiple stressors. *Ecology and Evolution*, **5**, 1538–1547.

Pimm SL, Dollar L, Bass OL (2006) The genetic rescue of the Florida panther. *Animal Conservation*, **9**, 115–122.

van Ravenzwaay B, Cunha GCP, Leibold E *et al.* (2007) The use of metabolomics for the discovery of new biomarkers of effect. *Toxicology Letters*, **172**, 21–28.

Reed DH, Fox CW, Enders LS, Kristensen TN (2012) Inbreeding-stress interactions: evolutionary and conservation consequences. *Annals of the New York Academy of Sciences*, **1256**, 33–48.

Reed DH, Frankham R (2003) Correlation between fitness and genetic diversity. *Conservation Biology*, **17**, 230–237.

Relyea R, Hoverman J (2006) Assessing the ecology in ecotoxicology: a review and synthesis in freshwater systems. *Ecology Letters*, **9**, 1157–1171.

Ros M, Sorensen D, Waagepetersen R *et al.* (2004) Evidence for genetic control of adult weight plasticity in the snail *Helix aspersa*. *Genetics*, **168**, 2089–2097.

Rönnegård L, Felleki M, Fikse WF, Mulder HA, Stransberg E (2013) Variance component and breeding value estimation for genetic heterogeneity of residual variance in Swedish Holstein dairy cattle. *Journal of Dairy Science*, **96**, 2627–2636.

Saccheri IJ, Nichols R a, Brakefield PM (2001) Effects of bottlenecks on quantitative genetic variation in the butterfly *Bicyclus anynana*. *Genetical research*, **77**, 167–81.

Sala OE, Chapin FS, Iii *et al.* (2000) Global biodiversity scenarios for the year 2100. *Science*, **287**, 1770–1774.

Sanchez-Garcia M, Álvaro F, Martín-Sánchez J a. *et al.* (2012) Breeding effects on the genotype×environment interaction for yield of bread wheat grown in Spain

during the 20th century. *Field Crops Research*, **126**, 79–86.
Sarup P, Jensen J, Ostersen T, Henryon M, Sørensen P (2016) Increased prediction accuracy using a genomic feature model including prior information on quantitative trait locus regions in purebred Danish Duroc pigs. *BMC Genetics*, **17**, 11.
Scheffers BR, De Meester L, Bridge TCL *et al.* (2016) The broad footprint of climate change from genes to biomes to people. *Science*, **354**, 719.
Scheiner SM, Lyman RF (1991) The genetics of phenotypic plasticity II. Response to selection. *Journal of Evolutionary Biology*, **4**, 23–50.
Schlichting CD (1986) The evolution of phenotypic plasticity in plants. *Annual Review of Ecology and Systematics*, **17**, 667–693.
Schou MF, Kristensen TN, Kellermann V, Schlötterer C, Loeschcke V (2014) A *Drosophila* laboratory evolution experiment points to low evolutionary potential under increased temperatures likely to be experienced in the future. *Journal of Evolutionary Biology*, **27**, 1859–1868.
Schou MF, Loeschcke V, Kristensen TN (2015) Inbreeding depression across a nutritional stress continuum. *Heredity*, **115**, 1–7.
Schuwirth N, Dietzel A, Reichert P (2015) The importance of biotic interactions for the prediction of macroinvertebrate communities under multiple stressors. *Functional Ecology*, **30**, 974–984.
Da Silva A, Luikart G, Yoccoz NG, Cohas A, Allainé D (2006) Genetic diversity-fitness correlation revealed by microsatellite analyses in European alpine marmots (*Marmota marmota*). *Conservation Genetics*, **7**, 371–382.
Singh S., Singh M (2004) Multivariate analysis in relation to genetic improvement in *Cuphea procumbens*. *Journal of Genetics and Breeding*, **58**, 105–112.
Smith SJ, Edmonds J, Hartin CA, Mundra A, Calvin K (2015) Near-term acceleration in the rate of temperature change. *Nature Climate Change*, **5**, 333–336.
Solieman THI, El-Gabry MAH, Abido AI (2013) Heterosis, potence ratio and correlation of some important characters in tomato (*Solanum lycopersicum* L.). *Scientia Horticulturae*, **150**, 25–30.
Stearns SC (1989) The evolutionary significance of phenotypic plasticity. *BioScience*, **39**, 436–445.
Stillman JH (2003) Acclimation capacity underlies susceptibility to climate change. *Science*, **301**, 65.
Sørensen JG, Kristensen TN, Kristensen K V., Loeschcke V (2007) Sex specific effects of heat induced hormesis in Hsf-deficient *Drosophila melanogaster*. *Experimental Gerontology*, **42**, 1123–1129.
Sørensen P, de los Campos G, Morgante F, Mackay TFC, Sorensen D (2015) Genetic control of environmental variation of two quantitative traits of *Drosophila melanogaster* revealed by whole-genome sequencing. *Genetics*, **201**, 487–497.
Sørensen MK, Norberg E, Pedersen J, Christensen LG (2008) Invited review: crossbreeding in dairy cattle: a Danish perspective. *Journal of dairy science*, **91**, 4116–4128.

Taft HR, Roff DA (2012) Do bottlenecks increase additive genetic variance? *Conservation Genetics*, **13**, 333–342.

Tallmon DA, Luikart G, Waples R (2004) The alluring simplicity and complex reality of genetic rescue. *Trends in Ecology & Evolution*, **19**, 489–496.

Teklewold A, Becker HC (2006) Comparison of phenotypic and molecular distances to predict heterosis and F1 performance in Ethiopian mustard (*Brassica carinata* A. Braun). *Theoretical and Applied Genetics*, **112**, 752–759.

Templeton AR, Hemmer H, Mace G *et al*. (1986) Local adaptation, coadaptation, and population boundaries. *Zoo Biology*, **5**, 115–125.

Teplitsky C, Mills JA, Alho JS, Yarrall JW, Merilä J (2008) Bergmann's rule and climate change revisited: disentangling environmental and genetic responses in a wild bird population. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 13492–13496.

Thorpe AS, Aschehoug ET, Atwater DZ, Callaway RM (2011) Interactions among plants and evolution. *Journal of Ecology*, **99**, 729–740.

US-EPA (2002) *Health Effects Test Guidelines OPPTS 870.1000 Acute Toxicity Testing*. United States Environmental Protection Agency, December 2002.

Uyenoyama MK, Holsinger KE, Waller DM (1993) Ecological and genetic factors directing the evolution of self-fertilization. *Oxford Surveys in Evolutionary Biology*, **9**, 327–381.

Valladares F, Sanchez-Gomez D, Zavala MA (2006) Quantitative estimation of phenotypic plasticity: bridging the gap between the evolutionary concept and its ecological applications. *Journal of Ecology*, **94**, 1103–1116.

Via S, Lande R (1985) Genotype-environment interaction and the evolution of phenotypic plasticity. *Evolution*, **39**, 505–522.

Wallace AR (1889) *Darwinism: an Exposition of the Theory of Natural Selection with Some of its Applications*. MacMillan, London, UK.

Weber KE (1990) Increased selection response in larger populations. I. Selection for wing-tip height in *Drosophila melanogaster* at three population sizes. *Genetics*, **125**, 579–584.

Weber KE, Diggins LT (1990) Increased selection response in larger populations. II. Selection for ethanol vapor resistance in *Drosophila melanogaster* at two population sizes. *Genetics*, **125**, 585–597.

Whiteley AR, Fitzpatrick SW, Funk WC, Tallmon DA (2015) Genetic rescue to the rescue. *Trends in Ecology & Evolution*, **30**, 42–49.

Willi Y, Van Buskirk J, Hoffmann AA (2006) Limits to the adaptive potential of small populations. *Annual Review of Ecology, Evolution, and Systematics*, **37**, 433–458.

Willi Y, Hoffmann AA (2009) Demographic factors and genetic variation influence population persistence under environmental change. *Journal of Evolutionary Biology*, **22**, 124–133.

Willmore KE, Young NM, Richtsmeier JT (2007) Phenotypic variability: its components, measurement and underlying developmental processes. *Evolutionary Biology*, **34**, 99–120.

Wisz MS, Pottier J, Kissling WD *et al.* (2013) The role of biotic interactions in shaping distributions and realised assemblages of species: implications for species distribution modelling. *Biological Reviews*, **88**, 15–30.

Wood JLA, Yates MC, Fraser DJ (2016) Are heritability and selection related to population size in nature? Meta-analysis and conservation implications. *Evolutionary Applications*, **9**, 640–657.

Wright S (1931) Evolution in mendelian populations. *Genetics*, **16**, 97–159.

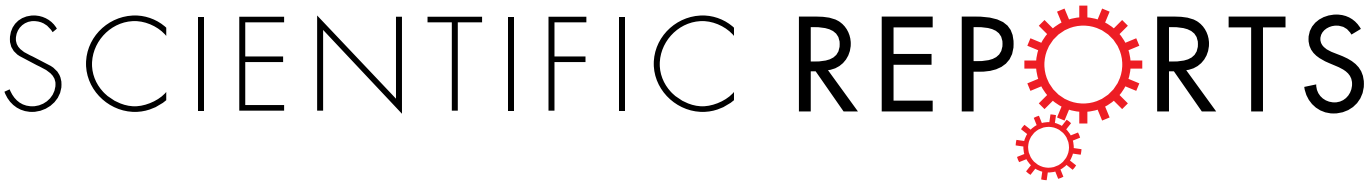
Ørsted M, Roslev P (2015) A fluorescence based hydrolytic enzyme activity assay for quantifying toxic effects of Roundup® to *Daphnia magna*. *Environmental Toxicology and Chemistry*, **34**, 1841–1850.

PAPER I

BIOTIC AND ABIOTIC FACTORS INVESTIGATED IN TWO *DROSOPHILA* SPECIES – EVIDENCE OF BOTH NEGATIVE AND POSITIVE EFFECTS OF INTERACTIONS ON PERFORMANCE

Published in *Scientific Reports*

MICHAEL ØRSTED, MADS FRISTRUP SCHOU & TORSTEN NYGAARD KRISTENSEN



OPEN

Received: 01 August 2016
Accepted: 30 November 2016
Published: 06 January 2017

Biotic and abiotic factors investigated in two *Drosophila* species – evidence of both negative and positive effects of interactions on performance

Michael Ørsted[1], Mads Fristrup Schou[2] & Torsten Nygaard Kristensen[1,2]

Multiple environmental factors acting in concert can interact and strongly influence population fitness and ecosystem composition. Studies investigating interactions usually involve only two environmental factors; most frequently a chemical and another abiotic factor such as a stressful temperature. Here we investigate the effects of three environmental factors: temperature, an insecticide (dimethoate) and interspecific co-occurrence. We expose two naturally co-occurring species of *Drosophila* (*D. hydei* and *D. melanogaster*) to the different environments during development and examine the consequences on several performance measures. Results are highly species and trait specific with evidence of two- and three-way interactions in approximately 30% of all cases, suggesting that additive effects of combined environmental factors are most common, and that interactions are not universal. To provide more informative descriptions of complex interactions we implemented re-conceptualised definitions of synergism and antagonism. We found approximately equal proportions of synergistic and antagonistic interactions in both species, however the effects of interactions on performance differed between the two. Furthermore, we found negative impacts on performance in only 60% of interactions, thus our study also reveals a high proportion of cases with positive effects of interactions.

Natural populations are exposed to multiple environmental stimuli simultaneously[1,2]. The impact of environmental factors may vary, and this is especially pronounced in seasonally fluctuating environments, e.g. during winter[3]. Environmental factors may interact in their impact on organisms resulting in fitness consequences that are different from what would be expected when considering each factor individually[1].

Interactions between environmental factors, both within and between biotic and abiotic factors, play an important role in determining species composition of communities and ecosystems[4,5]. Indeed, such interactions can be more important than abiotic habitat requirements when predicting community assemblies[6], highlighting the importance of integrating interactions in ecological prediction models. Environmental factors may also interact with the genotype of individuals and the genetic constitution of populations. For instance, fitness consequences of inbreeding is typically exacerbated under stressful environmental conditions[7], with proposed large implications for small and fragmented populations suffering from inbreeding and genetic drift[7–9]. Neglecting fitness consequences of interactions within and between biotic and abiotic interactions can have considerable undesirable consequences. This may result in underestimating the effect of multiple environmental factors on population persistency and the stability of communities[4], a risk exemplified by the combination of thermal extremes and drought stress resulting from climate change[10]. The predictability and generalizability of responses to multiple environmental factors should be incorporated in general global climate change models[11], and in ecological risk assessments[2,12] for increased accuracy and prediction power of community assembly modelling[6].

[1]Department of Chemistry and Bioscience, Section of Biology and Environmental Science, Aalborg University, Fredrik Bajers Vej 7H, DK-9220 Aalborg E, Denmark. [2]Department of Bioscience, Section of Genetics, Ecology and Evolution, Aarhus University, Ny Munkegade 114, DK-8000 Aarhus C, Denmark. Correspondence and requests for materials should be addressed to M.Ø. (email: moer@bio.aau.dk)

When the combination of deleterious environmental factors is more harmful to the organism than the sum of individual factors, this has traditionally been referred to as a "synergistic" interaction, e.g. higher temperature exacerbating the harmful effect of a pesticide[13,14]. If the combination of two or more factors is less harmful than the sum of individual factors, the interaction is traditionally termed "antagonistic", e.g. insect herbivory reducing harmful effects of plant competition[15]. However, the validity of these relatively simple terms has been debated[1,16,17], and several authors have proposed a reconceptualization of the typical classifications, enabling an inclusion of, e.g., interactions between individual environmental factors with opposing directions[18,19].

Studies of interactions typically investigate only two environmental factors, and among these the most frequently investigated factors are different temperatures and presence or absence of a chemical[20]. Among more than 150 studies investigating two-way interactions between a chemical and another environmental factor, 74% find an interaction, with 93% of these interactions being synergistic and 7% antagonistic[2]. Interactions have also been identified when assessing fitness consequences of chemicals in combination with biotic stressors, e.g. interspecific competition[21], starvation[22], pathogens/parasites[23] and predation[24]. In general, studies performed so far are highly biased towards assessing pesticides or other chemical compounds and typically involve only one species or several species investigated separately[2,25,26].

In this study, we investigate consequences of exposure to biotic and abiotic factors, that are potentially stressful in isolation or in combination, in two cosmopolitan *Drosophila* species; *Drosophila hydei* and *Drosophila melanogaster*. The effects of three environmental factors were investigated; low, intermediate and high developmental temperatures, presence or absence of the organophosphate insecticide dimethoate, and presence or absence of a co-occurring species. We combined all environmental factors in a full factorial manner, and analysed potential synergistic and antagonistic interactions (Fig. 1). The effects of the three environmental factors were investigated by assessing egg-to-adult viability, developmental time, upper and lower thermal limits and a behavioural trait. A composite measure of performance across traits was also computed. Responses to unfavourable environmental conditions can be highly sex-specific[27,28] and therefore both females and males were assayed. Based on recent reviews suggesting a high frequency of interactions, as well as the seemingly prevalent notion, that synergistic interactions typically have negative impacts on performance, we hypothesised to find 1) multiple environmental factors primarily interact in their effect on performance, 2) interactions are primarily synergistic, 3) the effects of interactions are mostly negative, and 4) the frequency and direction of interactions are both trait and species specific.

Results

Given the large amount of data and high number of potential interactions, we aimed at quantifying general patterns of responses to individual and combined environmental factors. We will refer to the environment experienced by the flies as the treatment, which is thus composed of up to three manipulated environmental factors. The consequences of the different treatments on individual traits as well as on the composite performance are summarized in Table 1. In order to achieve an overall picture of interactions we constructed linear models and extracted the standardised model coefficients of all treatments on the different traits, which are summarized for *D. hydei* in Fig. 2 and for *D. melanogaster* in Fig. 3.

Across all traits we observed both benefits and costs of exposure to potentially stressful environments (Table 1). We observed large variation in the composite performance measure both within and across the treatments (Table 1). In *D. hydei*, co-occurrence with *D. melanogaster* caused a significantly improved performance in all traits except for Critical Thermal minimum (CT_{min}). Conversely, *D. melanogaster* was largely unaffected by the presence of *D. hydei*. Dimethoate affected negative geotaxis behaviour as the sole trait in *D. melanogaster*, while in *D. hydei*, the presence of dimethoate significantly affected egg-to-adult viability, developmental time, and negative geotaxis behaviour. Egg-to-adult viability of *D. melanogaster* was largely unaffected by developmental temperature, whereas in *D. hydei* this trait was greatly impacted by heat alone, and also by cold when combined with other environmental factors (Table 1). In terms of the effect of developmental temperature on thermal tolerance, we also found different results in the two species. While a low developmental temperature resulted in a high cold tolerance and low heat tolerance in *D. melanogaster*, exposure of *D. hydei* to low developmental temperature resulted in significantly higher cold tolerance and unaltered heat tolerance compared to flies developed at an intermediate temperature, confirming results from other studies providing evidence for thermal acclimation[29]. The two species responded similarly in cold and heat tolerance to development at a high temperature, i.e. in both species high developmental temperature resulted in decreased CT_{min} but increased CT_{max}. Overall, the consequences of the different factors were both species and trait specific, and in a few cases sex specific (Table 1).

In *D. hydei* 37% of all tests resulted in significant two- or three-way interactions (Table 2). In *D. melanogaster* we found significant interactions in only 19% of the cases. These interactions were almost equally distributed between synergistic and antagonistic interactions in both species. The proportion of positive and negative synergistic interactions was approximately equally frequent in both species, whereas the pattern of antagonistic interactions differed more between the two species. The majority of antagonistic interactions for *D. hydei* were positive antagonistic (80%), i.e. less positive than expected additively, whereas negative antagonistic interactions were the most frequent type of antagonism in *D. melanogaster* (80%). This resulted in differences in the effects of interactions on performance. For *D. hydei*, most interactions had a negative effect (72%), while in *D. melanogaster* the majority of interactions had a positive effect (67%). Overall, we found more interactions that affected performance negatively. When we observed a significant interaction in both sexes it was always of the same interaction type, however we also found some interactions that only affected one sex (Figs 2 and 3). In *D. melanogaster* in particular, the effects of interactions on developmental time and negative geotaxis seemed to differ between sexes (Fig. 3).

For all traits where the response to co-occurrence was significantly different from that of the control (25 °C, no co-occurrence and no dimethoate), co-occurrence was beneficial to *D. hydei,* whereas *D. melanogaster* was

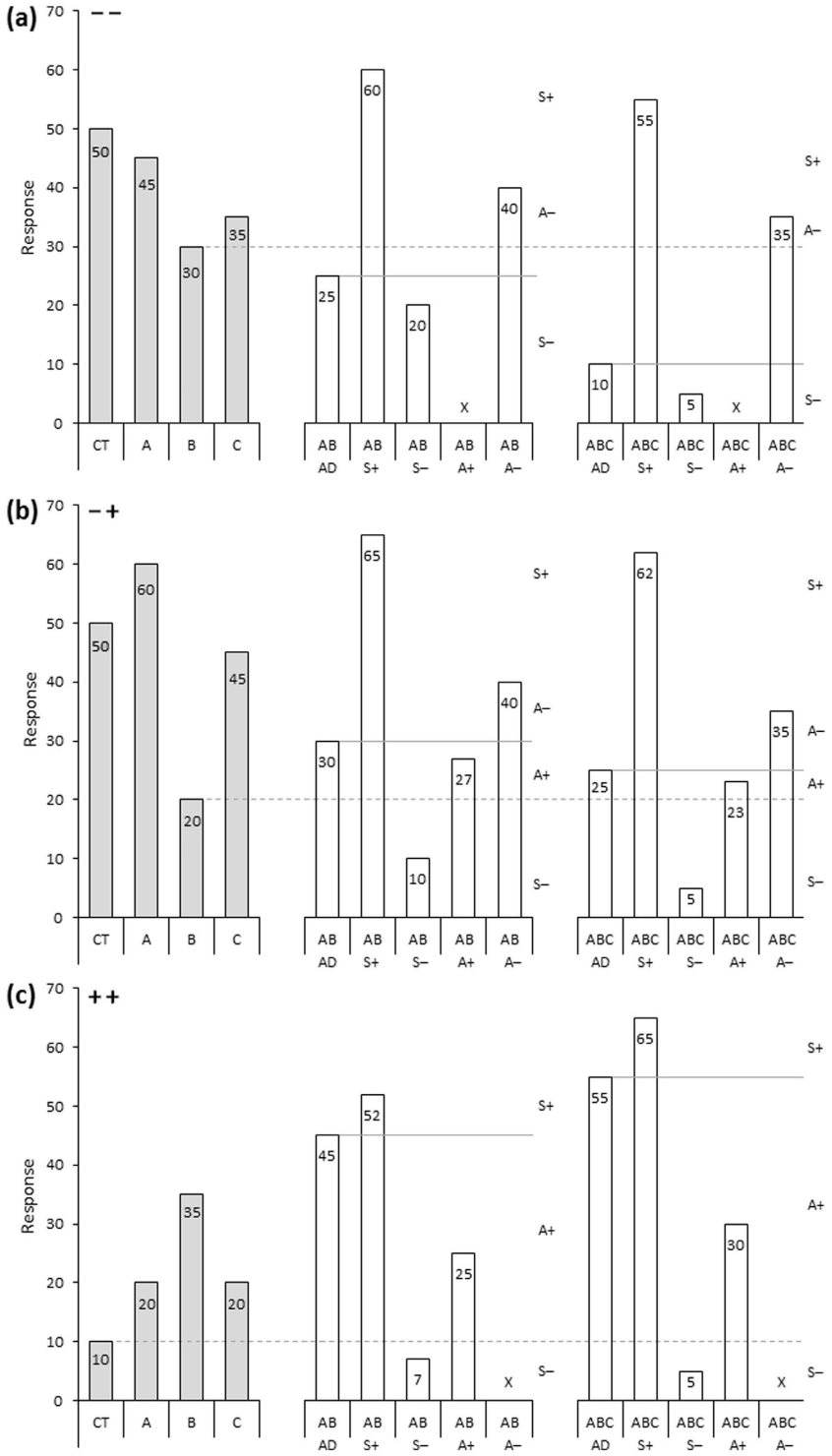


Figure 1. Illustration of our conceptual definitions of interaction types. Our definitions combine the magnitude and response direction of the interaction effect. Treatments in the factorial design include control (CT), with individual factors A, B, and C and with two factors (AB) or three factors (ABC). Directional interaction classes are +Synergistic (S+), −Synergistic (S−), +Antagonistic (A+), and −Antagonistic (A−) which depend on the effect of multiple factors (AB or ABC) compared to the additive sum (AD) of the individual effects of A and B (and C) relative to the control (CT). Height of the bars represents the absolute value of the response to each treatment. Grey shaded bars represent control treatment and the individual factors A, B, and C. Solid horizontal lines illustrate the additive sum for reference. The dashed horizontal lines represent the individual factor with the lowest response. The three plots illustrate interactions types in situations where the effects of individual environmental factors are all negative (**a**), opposing (**b**), and all positive (**c**) on the trait in question. An X indicates that the interaction class is not applicable in a given situation. Redrawn from refs 18, 19.

	Treatment	Egg-to-adult viability (%)	Developmental time (days)		CT_{min} (°C)		CT_{max} (°C)		Negative geotaxis (cm)		Composite performance	
		M/F	M	F	M	F	M	F	M	F	M	F
D. hydei	Control (25 °C)	34.8 ± 2.5	23.4 ± 0.19	22.4 ± 0.23	4.47 ± 0.08	4.85 ± 0.10	39.0 ± 0.13	39.0 ± 0.09	2.35 ± 0.09	2.4 ± 0.13	−0.11 ± 0.27	−0.10 ± 0.26
	Co-occur.	**51.2 ± 5.0****	**19.4 ± 0.23****	**19.9 ± 0.23****	4.56 ± 0.10	4.76 ± 0.09	39.4 ± 0.05**	39.5 ± 0.03**	4.63 ± 0.08**	4.5 ± 0.10**	0.50 ± 0.24	**0.57 ± 0.24**
	Dim.	25.6 ± 1.4**	**20.6 ± 0.21****	**20.1 ± 0.22****	4.49 ± 0.07	4.46 ± 0.07**	39.3 ± 0.06	39.2 ± 0.06	4.35 ± 0.10**	3.8 ± 0.09**	0.20 ± 0.18	**0.16 ± 0.19**
	Co-occur. + Dim.	16.6 ± 2.5**	**18.6 ± 0.20****	**18.4 ± 0.21****	4.59 ± 0.08	4.53 ± 0.07**	39.4 ± 0.04*	39.6 ± 0.04**	4.37 ± 0.14**	4.3 ± 0.11**	0.20 ± 0.18	**0.31 ± 0.20**
	Cold	31.8 ± 2.5	80.9 ± 0.54**	80.0 ± 0.51**	**1.82 ± 0.12****	**2.48 ± 0.18****	**39.5 ± 0.04****	**39.2 ± 0.14**	0.75 ± 0.05**	0.8 ± 0.06**	−0.10 ± 0.55	**−0.29 ± 0.49**
	Cold + Co-occur.	**51.5 ± 3.9****	78.2 ± 0.72**	77.0 ± 0.77**	**2.30 ± 0.13****	**2.49 ± 0.15****	**39.2 ± 0.10**	**39.3 ± 0.12**	4.30 ± 0.11**	NA	0.29 ± 0.48	**0.23 ± 0.53**
	Cold + Dim.	6.5 ± 0.9**	68.1 ± 0.60**	65.6 ± 0.64**	**2.17 ± 0.15****	**2.42 ± 0.17****	**39.1 ± 0.10**	**39.2 ± 0.15**	1.49 ± 0.07**	NA	−0.42 ± 0.44	−0.22 ± 0.45
	Cold + Co-occur. + Dim.[a]	1.3 ± 0.4**	NA	67.7 ± 1.73**	NA	NA	NA	NA	NA	NA	0.95 ± 0.00	−0.99±0.02
	Heat	10.5 ± 2.5**	12.8 ± 0.14**	13.0 ± 0.13**	6.38 ± 0.18**	6.29 ± 0.12**	**39.8 ± 0.09****	**39.9 ± 0.08****	NA	1.3 ± 0.11**	0.01 ± 0.43	−0.19 ± 0.41
	Heat + Co-occur.	14.8 ± 3.4**	13.5 ± 0.22**	13.6 ± 0.19**	6.10 ± 0.09**	6.18 ± 0.14**	**39.9 ± 0.07****	**39.8 ± 0.11****	NA	NA	0.17 ± 0.42	−0.02 ± 0.36
	Heat + Dim.	8.3 ± 1.3**	13.7 ± 0.17**	13.8 ± 0.19**	6.11 ± 0.18**	6.28 ± 0.31**	**39.6 ± 0.14****	**39.7 ± 0.09****	2.74 ± 0.14*	NA	−0.14 ± 0.32	−0.14 ± 0.37
	Heat + Co-occur. + Dim.	29.5 ± 4.4	12.8 ± 0.15**	12.6 ± 0.17**	6.35 ± 0.11**	6.06 ± 0.09**	**39.8 ± 0.05****	**40.0 ± 0.07****	3.37 ± 0.09**	3.2 ± 0.07**	0.18 ± 0.35	**0.28 ± 0.34**
D. melanogaster	Control (25 °C)	80.0 ± 2.5	12.4 ± 0.10	12.3 ± 0.10	6.6 ± 0.07	6.6 ± 0.08	39.9 ± 0.08	40.1 ± 0.04	5.48 ± 0.10	4.0 ± 0.10	0.23 ± 0.17	**0.13 ± 0.15**
	Co-occur.	83.8 ± 2.8	13.8 ± 0.28**	11.8 ± 0.11**	6.4 ± 0.07	6.4 ± 0.10	39.9 ± 0.05	40.0 ± 0.06	5.62 ± 0.10	4.3 ± 0.08*	0.30 ± 0.15	**0.23 ± 0.15**
	Dim.	70.3 ± 4.8	12.4 ± 0.10	12.1 ± 0.10*	6.6 ± 0.11	6.7 ± 0.07	39.9 ± 0.06	40.1 ± 0.04	5.73 ± 0.08*	3.6 ± 0.11**	0.13 ± 0.23	−0.07 ± 0.20
	Co-occur. + Dim.	89.1 ± 1.5**	12.2 ± 0.08*	12.0 ± 0.08**	6.3 ± 0.07**	6.5 ± 0.08	40.0 ± 0.04	40.0 ± 0.05	5.51 ± 0.09	4.6 ± 0.10**	0.42 ± 0.16	**0.34 ± 0.19**
	Cold	79.7 ± 1.9	51.3 ± 0.28**	48.8 ± 0.23**	4.1 ± 0.11**	3.7 ± 0.08**	39.1 ± 0.04**	39.4 ± 0.09**	4.31 ± 0.11**	3.5 ± 0.09**	−0.33 ± 0.43	−0.26 ± 0.42
	Cold + Co-occur.	69.5 ± 3.3*	50.1 ± 0.37**	48.3 ± 0.40**	3.9 ± 0.14**	3.6 ± 0.09**	39.3 ± 0.07**	39.3 ± 0.09**	4.27 ± 0.16**	4.2 ± 0.09	−0.34 ± 0.42	−0.26 ± 0.45
	Cold + Dim.	72.3 ± 1.9*	52.6 ± 0.29**	51.3 ± 0.30**	4.0 ± 0.12**	3.9 ± 0.09**	39.1 ± 0.08**	39.2 ± 0.06**	4.14 ± 0.13**	3.8 ± 0.07**	−0.44 ± 0.43	−0.39 ± 0.44
	Cold + Co-occur. + Dim.	77.3 ± 2.5	49.1 ± 0.23**	46.3 ± 0.20**	4.0 ± 0.10**	3.6 ± 0.09**	39.2 ± 0.07**	39.3 ± 0.11**	3.91 ± 0.13**	3.6 ± 0.09**	−0.34 ± 0.42	−0.27 ± 0.43
	Heat	80.5 ± 2.1	9.5 ± 0.06**	9.3 ± 0.06**	7.3 ± 0.06**	7.2 ± 0.11**	40.2 ± 0.13	40.5 ± 0.06**	5.04 ± 0.08**	4.1 ± 0.09	**0.18 ± 0.28**	**0.23 ± 0.28**
	Heat + Co-occur.	75.0 ± 2.4	10.1 ± 0.11**	9.8 ± 0.10**	7.3 ± 0.08**	7.2 ± 0.07**	39.9 ± 0.13	40.6 ± 0.06**	5.02 ± 0.10**	3.1 ± 0.14**	0.02 ± 0.25	−0.05 ± 0.33
	Heat + Dim.	82.3 ± 1.2	9.4 ± 0.10**	9.2 ± 0.10**	7.5 ± 0.06**	7.3 ± 0.05**	40.2 ± 0.11	40.5 ± 0.07**	4.70 ± 0.12**	4.0 ± 0.08	**0.16 ± 0.32**	**0.22 ± 0.28**
	Heat + Co-occur. + Dim.	78.3 ± 2.4	**9.1 ± 0.09****	**8.8 ± 0.08****	7.1 ± 0.07**	7.5 ± 0.11**	**40.0 ± 0.14**	40.6 ± 0.05**	4.27 ± 0.10**	2.8 ± 0.11**	**0.003 ± 0.26**	−0.01 ± 0.38

Table 1. Results for the effects of each treatment on egg-to-adult viability, developmental time, CT_{min}, CT_{max}, negative geotaxis (RING assay) and overall composite performance in *D. hydei* and *D. melanogaster*. Dim.: dimethoate. Co-occur.: co-occurrence. Values are expressed as means ± S.E. for males (M) and females (F). The direction of the effect of a given treatment on a trait in relation to performance compared to the control environment (25 °C, no dimethoate, no co-occurrence) for that trait is indicated; bold numbers indicate a performance advantage, numbers not in bold a disadvantage. Asterisks indicate significant difference from control: $*p < 0.05$, $**p < 0.01$ (Welch's t-test). Note that these pairwise comparisons only reflect the effects on performance of the treatment, not the interaction between the two or three involved factors. Some treatments did not yield enough live adult flies for assessing all traits or did not exceed the minimum number of flies accepted for assessing a trait. These are indicated as NA. The number of flies (n) that each value is based on, as well as the minimum number of flies accepted for a given trait, can be found in Supplementary Table S2. The composite performance is calculated as the average effect of each treatment after standardising the responses within each trait, and thus represents the average performance effect of the different treatments across the five traits (±S.E., $n = 5$) for each sex. Thus, the direction of the effect of composite performance is not relative to the control. Bold numbers here represent a positive composite performance measure, and numbers that are not in bold represent a negative composite performance measure. Egg-to-adult viability is included in the estimate of composite performance of both males and females as sex-differentiation was not performed for this trait. [a]This treatment yielded only few surviving females and not enough flies to assess thermal tolerance or negative geotaxis.

largely unaffected by co-occurrence (Table 1). Interestingly, in terms of composite performance, combinations of factors involving co-occurrence were mostly positive in both species, albeit highly variable between treatments. The effect of co-occurrence was typically dependent on other factors, e.g. for egg-to-adult viability of *D. hydei* co-occurrence of *D. melanogaster* increased the proportion of surviving adults at both 25 and 13 °C, but not at 31 °C (Fig. 4). However, at both 25 and 13 °C the addition of dimethoate resulted in lower egg-to-adult survival than expected from the effects of dimethoate alone at these temperatures, i.e. there were strong negative synergistic interactions (S−; Fig. 2). Heat (31 °C) reduced egg-to-adult viability but neither co-occurrence or dimethoate alone had a strong effect on survival at this temperature, however, when combined the resulting survival was comparable to that of the control treatment (Fig. 4). The interaction between heat, co-occurrence and dimethoate in *D. hydei* was classified as a negative antagonistic interaction (A−) as it was less negative than predicted additively, i.e. the interaction itself was beneficial in terms of survival (Fig. 2).

Dimethoate decreased the egg-to-adult survival of *D. hydei,* but was beneficial in terms of shortened developmental time and increased negative geotaxis behaviour. The egg-to-adult survival of *D. melanogaster* was not significantly decreased, and most other traits were also unaffected by the presence of dimethoate. Interestingly,

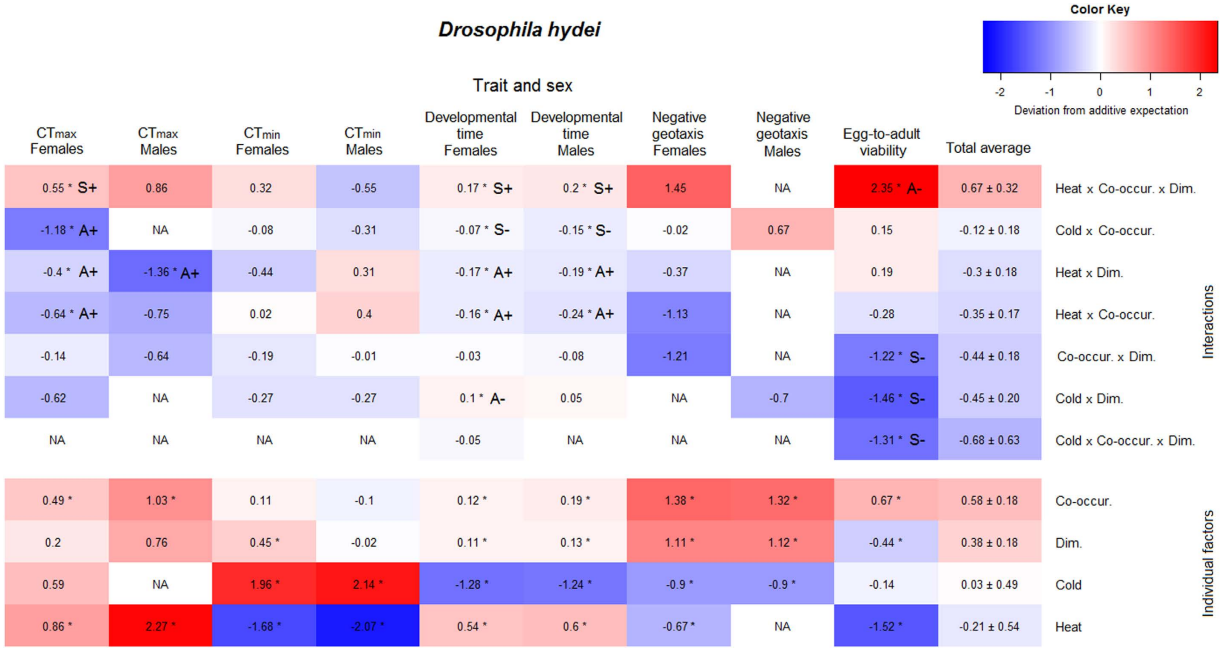


Figure 2. Heat map of interactions in *D. hydei*. Heat map showing the direction and magnitude of the model coefficients reflecting the effects of treatments on egg-to-adult viability, developmental time, CT_{min}, CT_{max} and negative geotaxis (RING assay) in *D. hydei*. The effects are shown for both sexes in all traits except egg-to-adult viability. Positive coefficients represent positive deviation from the additive expectation, and can thus be interpreted as a performance advantage of the interaction itself, regardless of whether the treatment overall was beneficial in terms of performance when compared to the control. Contrary, a negative coefficient implies a negative deviation from additivity and that the interaction itself is detrimental to performance for a given trait. The direction of the effect is illustrated by colour shading from blue (negative) to red (positive) and the values indicate the strength of the effects. The upper part includes all two- and three-way interactions between heat or cold, co-occurrence (Co-occur.), and dimethoate (Dim.). The lower part includes all effects of the individual factors. Within each part the treatments (rows) have been sorted by the average total effect, i.e. the average effect across traits ± S.E., in descending order. An asterisk indicates a significant interaction, or a significant effect of the individual environmental factor. S+ and S− designate interactions that are classified as positive or negative synergistic, respectively, as described in the text. A+ and A− designate interactions that are classified as positive or negative antagonistic, respectively. Some treatments did not yield enough live adult flies for assessing all traits or did not exceed the minimum number of flies accepted for assessing a trait. In a few traits the effect of an individual environmental factor could therefore not be determined, and the interactions involving the particular factor were omitted from the model. Both cases are designated NA.

in *D. melanogaster* the two sexes responded quite differently to dimethoate in their negative geotaxis behaviour response. Males showed an increased activity in the geotaxis assay, whereas females had a lower activity compared to the control treatment (Table 1). Lastly, the consequences of heat and cold were highly species and trait specific. For instance, *D. melanogaster* responded negatively to cold in all traits except egg-to-adult viability and CT_{min}. Developmental temperature seemed to have the greatest impact on developmental time, CT_{min}, and CT_{max} responses regardless of the co-occurrence and dimethoate status of the treatment (Table 1).

Discussion

The main aim of this study was to quantify the frequency, magnitude and direction of interactions between a set of environmental factors. This was investigated by exposing *D. hydei* and *D. melanogaster* to different developmental temperatures, the insecticide dimethoate and co-occurrence of the two species. The effects of these environmental factors were investigated on a range of traits in a full factorial manner.

We found that the effects of individual environmental factors as well as effects of the combinations were highly species specific (Figs 2 and 3). This is in agreement with other studies finding species specific effects of multiple stressors[30,31], suggesting that the exact effects of interactions depend on the experimental setup, as well as on the species investigated. This highlights the potential problems of extrapolating results from studies investigating interactions on one species to, e.g., community scale or to ecological risk assessments[12], and thereby emphasizing the need for further studies using standardised methods and/or multiple species. Interestingly, we found that the effects of the different treatments on composite performance were largely similar in the two species (Table 1), albeit highly variable between treatments. The calculation of a composite overall performance measure allows for an unbiased comparison of treatments in relation to fitness[32,33]. However, care must be taken when interpreting such a measure as a component of fitness as it can be questioned whether some traits contribute more to fitness than others (in our calculations all traits were given the same weight).

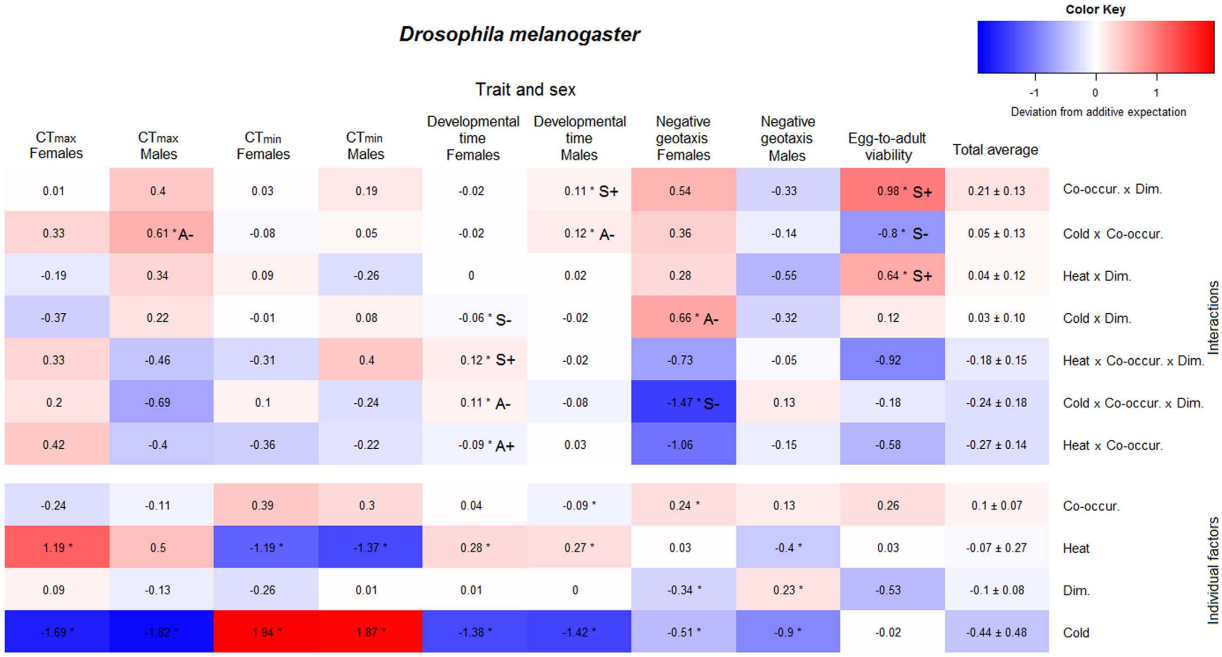


Figure 3. Heat map of interactions in *D. melanogaster*. Heat map showing the direction and magnitude of the model coefficients reflecting the effects of treatments on egg-to-adult viability, developmental time, CT_{min}, CT_{max} and negative geotaxis (RING assay) in *D. melanogaster*. The effects are shown for both sexes in all traits except egg-to-adult viability. Positive coefficients represent positive deviation from the additive expectation, and can thus be interpreted as a performance advantage of the interaction itself, regardless of whether the treatment overall was beneficial in terms of performance when compared to the control. Contrary, a negative coefficient implies a negative deviation from additivity and that the interaction itself is detrimental to performance for a given trait. The direction of the effect is illustrated by colour shading from blue (negative) to red (positive) and the values indicate the strength of the effects. The upper part includes all two- and three-way interactions between heat or cold, co-occurrence (Co-occur.), and dimethoate (Dim.). The lower part includes all effects of the individual factors. Within each part the treatments (rows) have been sorted by the average total effect, i.e. the average effect across traits ± S.E., in descending order. An asterisk indicates a significant interaction, or a significant effect of the individual environmental factor. S+ and S− designate interactions that are classified as positive or negative synergistic, respectively, as described in the text. A+ and A− designate interactions that are classified as positive or negative antagonistic, respectively.

		D. hydei	*D. melanogaster*	Total
Number of interactions	Total	18 (37)	12 (19)	30 (27)
	Two-way	13 (72)	9 (75)	22 (73)
	Three-way	5 (28)	3 (25)	8 (27)
Interaction classification	Synergistic	8 (44)	7 (58)	15 (50)
	S+	3 (16)	4 (33)	7 (23)
	S−	5 (28)	3 (25)	8 (27)
	Antagonistic	10 (56)	5 (42)	15 (50)
	A+	8 (45)	1 (9)	9 (30)
	A−	2 (11)	4 (33)	6 (20)
Performance effect of interaction	Positive	5 (28)	8 (67)	13 (43)
	Negative	13 (72)	4 (33)	17 (57)

Table 2. Number of significant interactions in all combinations of trait and sex for all treatments in both species showing the number of two- and three-way interactions, the classification of the interactions and the fitness effect of the interactions. Classification of interactions into positive and negative synergism (S+ and S−, respectively) and positive and negative antagonism (A+ and A−, respectively) is described in the text. A positive effect of an interaction on performance is defined as when the interaction is more positive than predicted additively and thus beneficial. Similarly, an interaction is defined as having a negative effect on performance when the interaction is more negative than predicted additively and therefore detrimental. The percentage of total interactions is given in parentheses after each number. For the 'total' row the number in parentheses designates the proportion of significant interactions among all tested potential interactions.

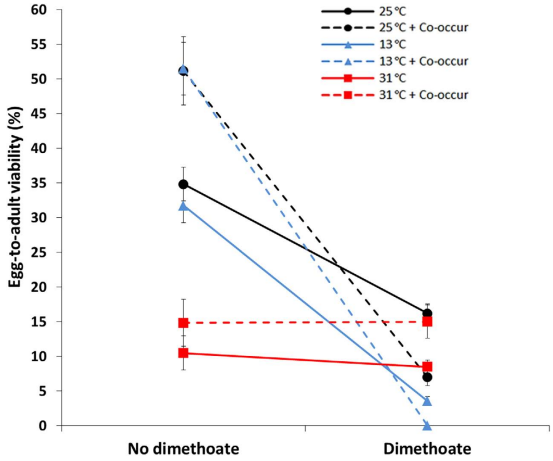


Figure 4. Example of combined effects of treatments. An example of the combined effects of temperature, co-occurrence (co-occur) and dimethoate on egg-to-adult viability (%) in *D. hydei*. Black lines represent 25 °C, blue are 13 °C, and red are 31 °C. Solid lines show the effects of temperature in the absence or presence of dimethoate (75 ppm). Dashed lines show the combined effects of temperature and co-occurrence in the absence or presence of dimethoate. Error bars represent standard error (S.E., $n = 30–40$).

In accordance with other studies our study provides evidence for the interaction of multiple environmental factors in their impact on fitness components[2,4,34] (Figs 2 and 3 and Table 2). However, significant interactions were not the rule and in roughly 70% of all cases interactions were not observed, thus additive effects of multiple factors were the most frequent observation in this study. This is in contrast with the seemingly prevalent notion that interactions are more common than additive effects[2,18,34,35]. In the review by Holmstrup *et al.*[2] of >150 studies investigating two-way interactions between a chemical and another environmental factor, 74% of the studies found interactions, and these were primarily synergistic. In a review by Crain *et al.*[18] of >170 studies manipulating two or more environmental factors in coastal and marine ecosystems, interactions were similarly found in 74% of the studies, approximately equally distributed between synergistic and antagonistic interactions. Furthermore, there seems to be a bias towards investigating and reporting the interactive effects of two or more adverse individual environmental factors, i.e. of the all-negative interaction type (Fig. 1a), in ecological research and toxicology (e.g. reviewed in refs 34, 36 and 37). In general researchers tend to be biased towards publishing "positive" results, i.e. showing interactions rather than additive effects[2], which could incorrectly reflect the frequency of interactions in nature and in laboratory studies.

The majority of studies investigating the effects of environmental factors and their potential interactions have used survival as a metric[34,35]. Less frequently, researchers have examined the effects of interactions on sub-lethal parameters such as growth, reproduction, behaviour or biochemical biomarkers, in which effects can potentially be detected before they affect survival[14,38]. We found that the effects of individual and combined environmental factors varied greatly depending on the trait being investigated (Figs 2 and 3), and consequently the overall performance measure was highly variable. In both species we found examples of significant effects of treatment on developmental time, thermal tolerances and behaviour without a notable effect on egg-to-adult survival. This context dependency raises concerns about drawing conclusions about the severity of interactions in studies that are based solely on mortality assays.

In our study, *D. hydei* benefitted strongly in almost all traits, including composite performance, from co-occurring with *D. melanogaster*, which to a lesser extent benefitted from developing with *D. hydei*. Studies on co-occurrence and interspecific competition in *Drosophila* are relatively scarce (but see e.g. refs 39, 40 and 41) and we have been unable to find studies investigating the interactions between co-occurrence or competition and other environmental variables in *Drosophila*. The egg-to-adult viability data for *D. hydei* suggests that the environmental conditions are suboptimal for full development in the control treatment (35% egg-to-adult survival), but when co-occurring with *D. melanogaster* (except at 31 °C) survival is significantly higher. We propose that *D. melanogaster* enhances the medium for the slower developing *D. hydei*, e.g. by increasing porosity or nutrient availability, thus resulting in positive effects of co-occurrence. These patterns were also evident in the composite performance measure, which was positive in the majority of cases, especially for *D. hydei*.

With the increasing realization of the importance of complex interactions in ecological contexts, it has become clear that despite the common use of the terms synergism and antagonism in the scientific literature to describe interactions, consensus seems to be lacking regarding an operational definition[1,2,16–19,35,42]. Synergism is normally used to define a cumulative effect greater than the additive sum of individual effects, whereas antagonism defines a cumulative effect that is less than additive[1]. Traditionally, the differentiation between synergism and antagonism has been relatively straightforward when individual factors are unidirectional, i.e. all-negative or all-positive (Fig. 1a–c [1,16]), however problems arise when individual factors are of opposite directions (Fig. 1b).

Because of the challenges arising from 1) the typical direction-independent classifications, and 2) opposing individual effects, we employed an alternative approach that systematically defines synergism and antagonism based on the direction and magnitude of the cumulative effect (Fig. 1), as proposed by Piggott *et al.*[19]. To highlight

the specific challenges faced in the traditional interaction framework, we attempted to re-designate interactions in our dataset by the classic definitions, and elaborated on situations where the challenge in using such terms can be circumvented by the re-conceptualised ideas (see Supplementary Figures S4 and S5 and accompanying discussion). In contrast to classic definitions, the proposed re-conceptualised model applies positive or negative to synergism or antagonism, representing situations where cumulative effects are more positive or more negative than additive (for synergistic interactions) or less positive or less negative than additive (for antagonistic interactions) providing additional information of the direction of interactions. We emphasize that the prefixes *positive* and *negative* do not describe the performance or fitness effect of the interaction, e.g. a *positive antagonism* is not necessarily beneficial to the organism, partly because it can be difficult to establish whether an effect direction is beneficial or detrimental to fitness in some traits, e.g. as in the case of developmental time[43]. We expanded on the original reconceptualization of these classic terms by applying them to three-way interactions (Fig. 1). Such second-order interactions of three or more factors have rarely been investigated (but see refs 11, 44 and 45) perhaps due to complicated experimental designs and complex interpretation of such interactions[35]. The new directional interaction type approach overcomes some of the problems of the traditional framework. We argue that despite looking complex at first, this approach translates into more informative descriptions and eases interpretation[18,19].

In this study we found a large proportion of treatments that were beneficial relative to the control treatment. Often treatments had negative effects on some traits, but led to performance benefits in others, i.e. highly context dependent, complicating the traditional assumption of detrimental synergy in situations of opposing individual factors[18]. Our finding of positive effects of interactions on performance is in accordance with other studies showing that the frequency of interactions with a beneficial effect can increase with the amount of detrimental environmental factors and that positive rather than negative effects of interactions dominate in certain natural communities[46,47]. Moreover, positive interactions can maintain the diversity of harsh environments where mutualistic relationships between species often govern survival[48,49]. Thus, our study adds to the growing realization of the importance of positive interactions in ecology which should be taken into account in ecological models and predictions. Furthermore, we identified several situations where an environmental factor, when applied alone, had little or no effect on performance, but when combined with other environmental factors resulted in a significant effect on performance. In (eco)toxicology this is sometimes referred to as potentiation or sensitisation[50]. Such interactions are of great interest, especially in environmental risk assessments, because factors that are seemingly harmless can, when combined, have tremendous unpredictable effects[2,19,35].

Conclusion

In this study we investigated effects of three environmental factors (temperature, dimethoate and species co-occurrence) and their interactions on several life history traits, thermal resistance and a behavioural trait in both sexes of two *Drosophila* species. We expand on the scarce knowledge on consequences of interactions between more than two environmental factors. In doing so, we take novel steps to provide more informative descriptions of consequences of such complex interactions on fitness components. Results suggest that although interactions do occur they are not omnipresent and additivity is more often observed. Further extrapolating results from one species, trait or sex to others might yield misleading results. Lastly, our study also highlights the importance of considering positive interactions in ecological contexts.

Methods

Fly stocks and preparation. *D. hydei* and *D. melanogaster* mass bred populations were established from flies caught at an apple heap in Karensminde orchard at the Danish peninsula of Jutland in September 2014 and November 2013 respectively (for details on location and habitat see ref. 51). Wild caught inseminated females ($n = 25$) contributed with an equal number of offspring to the establishment of mass bred populations. Each population was maintained at a population size of minimum 1000 individuals per generation at 25 °C in a 12:12 light:dark photoperiod for 8 (*D. hydei*) and 37 (*D. melanogaster*) generations prior to the experiment. The medium used for maintenance of flies was a standard *Drosophila* medium consisting of yeast (16 g/L), soy flour (9 g/L), cornmeal (66 g/L), agar (5 g/L), and glucose syrup (100 g/L) mixed with tap water. To control fungal growth, nipagen (9 mL/L) and 80% acetic acid (1 mL/L) were added to the medium. Parental flies were density controlled during development by controlled egg-laying time (24 h period) of 200 flies in five 175 mL bottles with 35 mL medium. To density control the development of experimental flies, we collected eggs produced by parental flies (12–14 days old for *D. hydei* and 3–4 days old for *D. melanogaster*) on three consecutive days using the following approach: Twenty parental flies were distributed into each of 50 vials at 25 °C, each containing a spoon with 1.5 mL of *Drosophila* medium with dry yeast. Approximately 12 h later, eggs were transferred in groups of 40 to vials containing 9 mL Formula 4–24® Instant *Drosophila* Medium Blue (Carolina Biological Supply Company, Burlington, NC, USA). This medium was mixed the day before egg collection and consisted of 1.6 g Formula 4–24® instant medium and 7.5 mL demineralised water (with or without dimethoate), and was kept at 10 °C until use. Instant medium was used in all treatments, regardless of dimethoate status.

Experimental setup. We investigated the effects of three different developmental environmental factors and the potential interactions between these factors on multiple traits, by exposing developing flies to different temperatures, an insecticide and co-occurrence in a full factorial design (Supplementary Table S1). Co-occurrence in this case does not necessarily imply competition between the two species and thus potentially fitness costs, so in this study we simply refer to this situation as co-occurrence. Flies were exposed to either of three constant temperatures during the development from egg to adult; 13 (cold), 25 (intermediate) and 31 °C (warm). This was done by transferring medium vials, immediately after egg collection, to climate incubators (Binder model KBWF 720 E5.3, Binder, Tuttlingen, Germany) maintaining an average (±s.d.) temperature of 13 ± 0.2, 25 ± 0.5

or 31 ± 0.5 °C, 40–60%RH and a 12:12 light:dark photoperiod under cool white-fluorescent light. *D. hydei* and *D. melanogaster* are cosmopolitan species that can be found consistently from latitudes −46.2 °S to +73.4 °N and −54.5 °S to +73.4 °N, respectively (TaxoDros database: v.1.04, http://www.taxodros.uzh.ch). Within these distributional ranges they can experience temperatures in the range of −8.4 °C to +33.5 °C and −9.0 °C to +33.6 °C, for *D. hydei* and *D. melanogaster*, respectively (WorldClim database: v.1.4, http://www.worldclim.org)[52]. Climatic temperature ranges were 10% quantile of minimum temperature in coldest month and 90% quantile of maximum temperature in warmest month. Data was treated and cross-referenced using methods described in Schou *et al.* (in press)[29]. Thus the thermal regimes employed in this study are well within the range of what the two species will experience in their natural habitats.

As a chemical abiotic environmental factor we used dimethoate (analytical grade 99.5%, CAS: 60-51-5, Sigma-Aldrich, Seelze, Germany; for more information on dimethoate e.g. rates of breakdown and acidity see Kristensen *et al.*[53] and references therein). A 10 ng μL^{-1} stock solution of dimethoate was prepared in demineralised water the day before the experiments, and mixed with the developmental medium for a nominal concentration in the medium of 75 $\mu g L^{-1}$ (ppb). This concentration was based on data from Kristensen *et al.*[53] and a series of preliminary range-finding tests assessing egg-to-adult viability and developmental time. *D. hydei* and *D. melanogaster* showed dissimilar responses both in terms of viability and developmental time in preliminary experiments, and we therefore selected an intermediate concentration (Supplementary Figures S1 and S2). Although dimethoate has been banned by the European Union and the US Environmental Protection Agency[54], it is still widely used, especially in developing countries and illegally in Southern Europe[54,55], where concentrations of up to 120 ppb have been found in olives[55]. Thus, the concentration used in this study is considered ecologically relevant, i.e. it is within those concentrations encountered by insects in the field.

Co-occurrence during development was imposed by placing an equal number of eggs from *D. hydei* and *D. melanogaster* simultaneously in the vials (Supplementary Table S1). The total number of eggs was the same as in treatments without co-occurrence, reducing potential density-dependent effects during development. We refer to the 25 °C, no co-occurrence and no dimethoate treatment as the control treatment. For most treatments, we collected eggs into 30 vials per treatment, and for some treatments, which were expected to yield few surviving adult flies, we collected eggs into 40 vials per treatment (Supplementary Table S1). In total 24,000 eggs were distributed to 600 vials. The vials were placed randomly in racks and the racks within each incubator were shuffled randomly every day until emergence of adult flies. The combination of three temperatures (13, 25 and 31 °C), two dimethoate levels (0 and 75 $\mu g L^{-1}$ (ppb)), and two co-occurrence levels (no co-occurrence and co-occurrence) resulted in a total of 12 treatments per species. The vials from each developmental rearing regime were checked daily at 08:00 a.m. for emerged flies. Emerged flies were anaesthetised with CO_2, sexed and counted to estimate egg-to-adult viability and developmental time. Flies needed for phenotypic assessments were transferred in groups of 40 flies to separate vials for each sex and to a common environment two days prior to assessments. In the common environment, adult flies from the two species were kept separately at 7 mL standard *Drosophila* medium (same as used for maintenance) without dimethoate and at 25 °C.

Phenotypes assessed. *Egg-to-adult viability and developmental time.* Egg-to-adult viability was determined as the proportion of eggs in a vial developing successfully into the adult life stage. Flies that had died during emergence from the pupae were not counted as a surviving adult fly. Developmental time was assessed every 24 h as the difference between time of emergence and time of egg collection. The assessment of developmental time of flies of a given treatment ceased when no flies had emerged for two consecutive days (four days for cold treatments). We defined a decreased egg-to-adult viability as a fitness disadvantage and interpreted a decreased developmental time (higher rate of development) as a fitness benefit. The 'faster is better' interpretation is debatable, as fast growth can be associated with trade-offs with other fitness components such as decreased efficiency of the immune system[43], and costs and benefits associated with fast development is likely environment specific.

Thermal limits. Flies from each treatment were assessed for critical thermal minimum (CT_{min}) and critical thermal maximum (CT_{max}), i.e. their ability to tolerate low and high temperatures, respectively. These standardised procedures of gradual cooling or heating have been suggested to be more ecologically relevant than procedures using abrupt temperature changes[56,57]. We defined CT_{min} as the temperature at which absolutely no movement of the body or appendages of flies is observed (see e.g. ref. 56 for details), as a result of the flies entering chill coma. Similarly, CT_{max} or knockdown temperature is defined as the temperature at which a complete cessation of movement of the flies occurs, due to heating. From each sex and each treatment, 20 flies (for exact numbers see Supplementary Table S2) of age 60 ± 12 h were transferred to individual small glass vials (45 × 15 mm) and randomly placed in a metal rack, which was submerged in a water bath pre-set at 25 °C. When assessing CT_{min}, the temperature in the water bath was decreased at a rate of 0.1 °C/min and the temperature, at which the flies was completely immobilised due to chill coma, was recorded. When assessing CT_{max} the temperature in the water bath was increased at a rate of 0.1 °C/min and temperature at which absolutely no movement was observed was recorded. Movement was stimulated by shining a flashlight and gently prodding the vials with a metal rod. We defined an increased CT_{max} i.e. higher heat tolerance as beneficial to performance, as evident in other studies[57,58]. Similarly congruent with other studies, we defined decreased CT_{min} i.e. higher cold tolerance as advantageous in terms of performance, because there seems to be only few costs associated with higher cold tolerance[59–61].

Negative geotaxis. In order to assess the effects of the different environmental factors on behaviour, we investigated the negative geotaxis behaviour of the flies. Negative geotaxis is an innate escape response where flies move in opposite direction of the force of gravity. The behaviour is typically elicited via mechanical stimulation by tapping the flies to the bottom of an empty vial and assessed as the velocity or distance moved by the flies when

ascending the walls of the container. For this purpose, we utilized a modified version of the high-throughput Rapid Iterative Negative Geotaxis (RING) assay, developed by Gargano *et al.*[62]. In this assay digital photography is used to document negative geotaxis behaviour in multiple groups of flies simultaneously (for details on our version of the RING apparatus see Supplementary Figure S3). To assess negative geotaxis, 10 flies were transferred into each of 10 empty vials. Fresh new vials were used for each treatment as Nichols *et al.*[63] found that flies in previously used vials will not climb to the same extent as in new vials. In total 20 flies of each sex from each treatment at age 60 ± 12 h were assessed in the RING assay. Ten minutes after the flies had been transferred to the vials all ten vials were loaded into the apparatus. One minute later the RING apparatus was forcefully knocked down three times in rapid succession to initiate the geotaxis response. A photo of the vertical position of the flies was captured exactly 3 s after eliciting the behaviour. Preliminary tests found the most differentiated response after 3 s, as after 5 s almost all flies had reached the top of the vials. This was performed a total of 5 times with 30 s pause in between. The vials were then rotated within the rack, and a trial of 5 images were captured at each position as described above, resulting in a total of 50 images of each vial. Images of the flies' positions were captured with the camera of an iPhone 5 s with default camera timer options (8 Mp; Apple Inc., Cupertino, CA, USA). The camera was mounted 30 cm from the apparatus in all experiments. The median height of the flies within each vial was measured using ImageJ software (version 1.48; ref. 64). All RING experiments were conducted in a climate controlled room at 25 °C, 50%RH and at constant light. Negative geotaxis behaviour was assessed between 08:00 and 10:00 a.m. on each test day, as the locomotor activity in *Drosophila* exhibits a distinct circadian rhythm[65]. An increased negative geotaxis behaviour, i.e. the flies crawled higher, was interpreted as beneficial, because the ability to escape a potential stressful environment is of key importance for fitness[66].

Composite performance. As a combined measure of overall performance of the treatments we calculated a composite performance measure based on all the traits expressed as a single value. This was done by standardizing the response of the different treatments within each trait, thus expressing the response to a treatment in terms of standard deviations from the mean of all treatments within a specific trait (negative and positive deviations were assigned according to the interpretations of how each trait relates to performance). We then averaged these deviations of each treatment across traits. We did this separately for each sex except for viability where sex-differentiation is not possible. By doing this we give equal weight to all traits, and thus we obtain an unbiased estimate of overall performance, which we assume constitute a component of fitness. This use of a composite performance/fitness measure is similarly employed in other studies[32,33].

Statistical analysis. To estimate the two-way and three-way interactions including the cold treatment, we constructed a linear model (cold model) for each trait with the factorial fixed effects temperature (two levels: benign and cold), dimethoate (two levels: 0 and 75 ppm) and co-occurrence (two levels: presence/absence), as well as all two- and three-way interactions. To compare cold model interactions with interactions involving heat exposure, we constructed parallel models including the heat treatment instead of the cold exposure (heat model). Individual and interaction effects of dimethoate and co-occurrence were included in both models, and thus extracted from one of the two models. In all traits, the response variable was scaled to a z-distribution to ease comparability across traits. The purpose of these models was to obtain a standardized measure of single and interaction coefficients as well as corresponding 95% confidence intervals. Given the large amount of data and models as well as our aim of obtaining a quantification of general patterns, we found this to be an appropriate approach, as opposed to model reduction and p-value estimation. Significance of an interaction was defined as when the confidence interval did not overlap with 0. To ease the comparison of interactions among treatments and traits, we produced heat maps of the estimated coefficients using the R-package 'gplots'[67].

A positive coefficient represents a positive deviation from the additive expectation, and can thus be interpreted as a performance advantage of the interaction itself, regardless of whether the treatment overall was beneficial in terms of performance when compared to the control. Thus although the flies from a particular treatment may be performing worse than the control group flies, the positive interaction is a benefit, compared to the expected value without the interaction. Contrary, a negative coefficient implies that the interaction itself is detrimental to performance.

If no flies had emerged from a given treatment (e.g. "cold + dimethoate + co-occurrence") the corresponding benign temperature treatment ("benign + dimethoate + co-occurrence") was removed from the model to create a balanced model. In this case the model included only two two-way interactions: temperature*dimethoate and temperature*co-occurrence. For CT_{min} and CT_{max}, data were analysed with a general linear model. Egg-to-adult viability data were analysed with a generalised linear model with a logit link function. We detected overdispersion in the model and corrected for this using a quasi-generalised linear model. Developmental time data were analysed with a generalised linear mixed effect model with a Poisson distribution. Replicate vials were included as a random effect, as flies from the same vial were not independent. RING data were analysed with a general linear mixed model with replicate vial, position of the vial in the rack and number of replicate picture (trial number) included as random effects. All statistical analyses were performed in R[68] (v. 3.1.2), and mixed models were performed using the R-package 'lme4'[69]. For a straightforward representation of the effects on performance of each individual treatment relative to the control, we used a simple pairwise comparison (Welch's t-test). These results will serve as background information when interpreting the results of the analysis of interactions presented above. CT_{max} data were anti-log transformed to fulfil assumptions of parametric analysis.

Classification of synergism and antagonism. In contrast to the traditional direction-independent framework[1] we use a classification system based on that of Piggott *et al.*[19], which combines the magnitude and response direction (+ or −) of interaction effects to define synergism and antagonism (Fig. 1). Our definition can be illustrated by assigning a positive effect of an individual effect as +1 and a negative effect as −1. We define

a deviation from additive that is greater than the sum of individual environmental factors and greater than any individual effect in the same direction or an interaction effect greater than both individual effects in absolute terms as synergistic. We classify it as *positive synergistic* (S+) i.e. more positive than predicted additively when $+1+1>+2$ or $-1+1>1$ or $-1+-1>0$, or *negative synergistic* (S−) i.e. more negative than predicted additively when $-1+-1<-2$ or $-1+1<-1$ or $+1+1<0$. If an interaction deviates from additivity and is less than the sum of individual factors or less-than-or-equal-to any individual effect in the same direction we define the interaction as antagonistic. We classify it as *positive antagonistic* (A+) i.e. less positive than predicted additively when $+1+1$ is between 0 and 2 or $-1+1$ is between -1 and 0 (or equal -1), or *negative antagonistic* (A−) i.e. less negative than predicted additively when $-1+-1$ is between -2 and 0 or $-1+1$ is between 0 and 1 (or equal 1). The terms 'more or less positive' and 'more or less negative' also apply to situations where one individual environmental factor has no effect, and the definitions are also easily applied to three-way interactions (Fig. 1). With this definition synergistic and antagonistic does not relate to whether or not the interaction itself constitutes a performance benefit. To assess this, we determine whether the deviation from additivity is positive (performance advantage) or negative (performance disadvantage) based on the model coefficients as described above. In any case a significant interaction can be directly interpreted as an ecological interaction between the individual environmental factors.

References

1. Folt, C., Chen, C., Moore, M. & Burnaford, J. Synergism and antagonism among multiple stressors. *Limnol. Oceanogr.* **44,** 864–877 (1999).
2. Holmstrup, M. *et al.* Interactions between effects of environmental chemicals and natural stressors: a review. *Sci. Total Environ.* **408,** 3746–3762 (2010).
3. Williams, C. M., Henry, H. A. L. & Sinclair, B. J. Cold truths: how winter drives responses of terrestrial organisms to climate change. *Biol. Rev.* **90,** 214–235 (2014).
4. Relyea, R. & Hoverman, J. Assessing the ecology in ecotoxicology: a review and synthesis in freshwater systems. *Ecol. Lett*. **9,** 1157–1171 (2006).
5. Kéfi, S., Holmgren, M. & Scheffer, M. When can positive interactions cause alternative stable states in ecosystems? *Funct. Ecol.* **30,** 88–97 (2016).
6. Schuwirth, N., Dietzel, A. & Reichert, P. The importance of biotic interactions for the prediction of macroinvertebrate communities under multiple stressors. *Funct. Ecol.* **30,** 974–984 (2015).
7. Fox, C. W. & Reed, D. H. Inbreeding depression increases with environmental stress: an experimental study and meta-analysis. *Evolution.* **65,** 246–258 (2011).
8. Liao, W. & Reed, D. H. Inbreeding–environment interactions increase extinction risk. *Anim. Conserv*. **12,** 54–61 (2009).
9. Schou, M. F., Loeschcke, V. & Kristensen, T. N. Inbreeding depression across a nutritional stress continuum. *Heredity* **115,** 56–62 (2015).
10. Bellard, C., Bertelsmeier, C., Leadley, P., Thuiller, W. & Courchamp, F. Impacts of climate change on the future of biodiversity. *Ecol. Lett*. **15,** 365–377 (2012).
11. Kaunisto, S., Ferguson, L. V. & Sinclair, B. J. Can we predict the effects of multiple stressors on insects in a changing climate? *Curr. Opin. Insect Sci.* **17,** 55–61 (2016).
12. Bednarska, A. J., Jevtić, D. M. & Laskowski, R. More ecological ERA: incorporating natural environmental factors and animal behavior. *Integr. Environ. Assess. Manag.* **9,** e39–46 (2013).
13. Osterauer, R. & Köhler, H. R. Temperature-dependent effects of the pesticides thiacloprid and diazinon on the embryonic development of zebrafish (*Danio rerio*). *Aquat. Toxicol.* **86,** 485–494 (2008).
14. Ørsted, M. & Roslev, P. A fluorescence based hydrolytic enzyme activity assay for quantifying toxic effects of Roundup® to *Daphnia magna*. *Environ. Toxicol. Chem.* **34,** 1841–1850 (2015).
15. Haag, J. J., Coupe, M. D. & Cahill, J. F. Antagonistic interactions between competition and insect herbivory on plant growth. *J. Ecol.* **92,** 156–167 (2004).
16. Dunne, R. P. Synergy or antagonism-interactions between stressors on coral reefs. *Coral Reefs* **29,** 145–152 (2010).
17. Vanhoudt, N., Vandenhove, H., Real, A., Bradshaw, C. & Stark, K. A review of multiple stressor studies that include ionising radiation. *Environ. Pollut.* **168,** 177–192 (2012).
18. Crain, C. M., Kroeker, K. & Halpern, B. S. Interactive and cumulative effects of multiple human stressors in marine systems. *Ecol. Lett*. **11,** 1304–1315 (2008).
19. Piggott, J. J., Townsend, C. R. & Matthaei, C. D. Reconceptualizing synergism and antagonism among multiple stressors. *Ecol. Evol.* **5,** 1538–1547 (2015).
20. Heugens, E. H. W., Hendriks, A. J., Dekker, T., Straalen, N. M. van & Admiraal, W. A review of the effects of multiple stressors on aquatic organisms and analysis of uncertainty factors for use in risk assessment. *Crit. Rev. Toxicol.* **31,** 247–284 (2001).
21. Boone, M. D. & Semlitsch, R. D. Interactions of an insecticide with competition and pond drying in amphibian communities. *Ecol. Appl.* **12,** 307–316 (2002).
22. Spadaro, D. A., Micevska, T. & Simpson, S. L. Effect of nutrition on toxicity of contaminants to the epibenthic amphipod *Melita plumulosa*. *Arch. Environ. Contam. Toxicol.* **55,** 593–602 (2008).
23. Forson, D. D. & Storfer, A. Atrazine increases ranavirus susceptibility in the tiger salamander, *Ambystoma tigrinum*. *Ecol. Appl.* **16,** 2325–2332 (2006).
24. Relyea, R. A. & Mills, N. Predator-induced stress makes the pesticide carbaryl more deadly to gray treefrog tadpoles (*Hyla versicolor*). *Proc. Natl. Acad. Sci. USA* **98,** 2491–2496 (2001).
25. MacAlpine, J. L. P., Marshall, K. E. & Sinclair, B. J. The effects of CO_2 and chronic cold exposure on fecundity of female *Drosophila melanogaster*. *J. Insect Physiol.* **57,** 35–37 (2011).
26. Rumschlag, S. L., Boone, M. D. & Fellers, G. The effects of the amphibian chytrid fungus, insecticide exposure, and temperature on larval anuran development and survival. *Environ. Toxicol. Chem.* **33,** 2545–2550 (2014).
27. Hoffmann, A. A., Hallas, R., Anderson, A. R. & Telonis-Scott, M. Evidence for a robust sex-specific trade-off between cold resistance and starvation resistance in *Drosophila melanogaster*. *J. Evol. Biol.* **18,** 804–810 (2005).
28. Sørensen, J. G., Kristensen, T. N., Kristensen, K. V. & Loeschcke, V. Sex specific effects of heat induced hormesis in Hsf-deficient *Drosophila melanogaster*. *Exp. Gerontol.* **42,** 1123–1129 (2007).
29. Schou, M. F., Mouridsen, M. B., Sørensen, J. G. & Loeschcke, V. Linear reaction norms of thermal limits in *Drosophila*: predictable plasticity in cold but not in heat tolerance. *Funct. Ecol.* doi: 10.1111/1365-2435.12782 (2016).
30. Amarasekare, K. G. & Edelson, J. V. Effect of temperature on efficacy of insecticides to differential grasshopper (Orthoptera: Acrididae). *J. Econ. Entomol*. **97,** 1595–1602 (2004).
31. Muturi, E. J., Lampman, R., Costanzo, K. & Alto, B. W. Effect of temperature and insecticide stress on life-history traits of *Culex restuans* and *Aedes albopictus* (Diptera: Culicidae). *J. Med. Entomol.* **48,** 243–250 (2011).

32. Overgaard, J., Kearney, M. R. & Hoffmann, A. A. Sensitivity to thermal extremes in Australian *Drosophila* implies similar impacts of climate change on the distribution of widespread and tropical species. *Global Change Biol.* **20,** 1738–1750 (2014).
33. Hereford, J. A quantitative survey of local adaptation and fitness trade-offs. *Am. Nat.* **173,** 579–588 (2009).
34. Darling, E. S. & Côté, I. M. Quantifying the evidence for ecological synergies. *Ecol. Lett.* **11,** 1278–1286 (2008).
35. Laskowski, R. *et al.* Interactions between toxic chemicals and natural environmental factors - a meta-analysis and case studies. *Sci. Total Environ.* **408,** 3763–3774 (2010).
36. Ban, S. S., Graham, N. A. J. & Connolly, S. R. Evidence for multiple stressor interactions and effects on coral reefs. *Glob. Chang. Biol.* **20,** 681–697 (2014).
37. Klaminder, J., Jonsson, M., Fick, J., Sundelin, A. & Brodin, T. The conceptual imperfection of aquatic risk assessment tests: highlighting the need for tests designed to detect therapeutic effects of pharmaceutical contaminants. *Environ. Res. Lett.* **9,** 84003 (2014).
38. Forbes, V. E., Palmqvist, A. & Bach, L. The use and misuse of biomarkers in ecotoxicology. *Environ. Toxicol. Chem.* **25,** 272–280 (2006).
39. Barker, J. S. F. & Podger, R. N. Interspecific competition between *Drosophila melanogaster* and *Drosophila simulans*: effects of larval density on viability, developmental period and adult body. *Ecology* **51,** 170–189 (1970).
40. Grimaldi, D. & Jaenike, J. Competition in natural populations of mycophagous *Drosophila*. *Ecology* **65,** 1113–1120 (1984).
41. Joshi, A. Variation in the relative magnitude of intraspecific and interspecific competitive effects in novel versus familiar environments in two *Drosophila* species. *J. Genet.* **83,** 179–188 (2004).
42. Chou, T. C. Drug combination studies and their synergy quantification using the Chou-Talalay method. *Cancer Res.* **70,** 440–446 (2010).
43. Niemelä, P. T., Vainikka, A., Hedrick, A. V. & Kortet, R. Integrating behaviour with life history: boldness of the field cricket, *Gryllus integer*, during ontogeny. *Funct. Ecol.* **26,** 450–456 (2012).
44. Chen, C. Y., Hathaway, K. M. & Folt, C. L. Multiple stress effects of Vision herbicide, pH, and food on zooplankton and larval amphibian species from forest wetlands. *Environ. Toxicol. Chem.* **23,** 823–831 (2004).
45. Heugens, E. H. W. *et al.* Population growth of *Daphnia magna* under multiple stress conditions: joint effects of temperature, food, and cadmium. *Environ. Toxicol. Chem.* **25,** 1399–1407 (2006).
46. Callaway, R. M. *et al.* Positive interactions among alpine plants increase with stress. *Nature* **417,** 844–848 (2002).
47. Brooker, R. W. *et al.* Facilitation in plant communities: The past, the present, and the future. *Journal of Ecology* **96,** 18–34 (2008).
48. Cavieres, L. A. & Badano, E. I. Do facilitative interactions increase species richness at the entire community level? *J. Ecol.* **97,** 1181–1191 (2009).
49. He, Q., Bertness, M. D. & Altieri, A. H. Global shifts towards positive species interactions with increasing environmental stress. *Ecol. Lett.* **16,** 695–706 (2013).
50. Odum, E. P. & Barrett, G. W. *Fundamentals of Ecology*. (Thomson/Brooks/Cole, Belmont, CA, USA, 2005).
51. Schou, M. F., Loeschcke, V. & Kristensen, T. N. Strong costs and benefits of winter acclimatization in *Drosophila melanogaster*. *PLOS ONE* **10,** e0130307 (2015).
52. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25,** 1965–1978 (2005).
53. Kristensen, T. N., Dahlgaard, J. & Loeschcke, V. Effects of inbreeding and environmental stress on fitness - using *Drosophila buzzatii* as a model organism. *Conserv. Genet.* **4,** 453–465 (2003).
54. Galt, R. E. Beyond the circle of poison: significant shifts in the global pesticide complex, 1976-2008. *Glob. Environ. Chang.* **18,** 786–799 (2008).
55. Rastrelli, L., Totaro, K. & De Simone, F. Determination of organophosphorus pesticide residues in Cilento (Campania, Italy) virgin olive oil by capillary gas chromatography. *Food Chem.* **79,** 303–305 (2002).
56. Overgaard, J., Kristensen, T. N. & Sørensen, J. G. Validity of thermal ramping assays used to assess thermal tolerance in arthropods. *PLOS ONE* **7,** 1–7 (2012).
57. Terblanche, J. S. *et al.* Ecologically relevant measures of tolerance to potentially lethal temperatures. *J. Exp. Biol.* **214,** 3713–3725 (2011).
58. Lutterschmidt, W. I. & Hutchison, V. H. The critical thermal maximum: history and critique. *Can. J. Zool.* **75,** 1561–1574 (1997).
59. Huey, R., Crill, W., Kingsolver, J. G. & Weber, K. A method for rapid measurement of heat or cold resistance of small insects. *Funct. Ecol.* **6,** 489–494 (1992).
60. Hoffmann, A. A., Sørensen, J. G. & Loeschcke, V. Adaptation of *Drosophila* to temperature extremes: bringing together quantitative and molecular approaches. *J. Therm. Biol.* **28,** 175–216 (2003).
61. Chown, S. L. & Terblanche, J. S. Physiological diversity in insects: ecological and evolutionary contexts. *Adv. In Insect Phys.* **33,** 50–152 (2006).
62. Gargano, J. W., Martin, I., Bhandari, P. & Grotewiel, M. S. Rapid iterative negative geotaxis (RING): a new method for assessing age-related locomotor decline in *Drosophila*. *Exp. Gerontol.* **40,** 386–395 (2005).
63. Nichols, C. D., Becnel, J. & Pandey, U. B. Methods to assay *Drosophila* behavior. *J. Vis. Exp.* **61,** 3795 (2012).
64. Rasband, W. S. ImageJ. U. S. National Institutes of Health (2014).
65. Allada, R., Emery, P., Takahashi, J. S. & Rosbash, M. Stopping time: the genetics of fly and mouse circadian clocks. *Annu. Rev. Neurosci.* **24,** 1091–1119 (2001).
66. Bahrndorff, S., Gertsen, S., Pertoldi, C. & Kristensen, T. N. Investigating thermal acclimation effects before and after a cold shock in *Drosophila melanogaster* using behavioural assays. *Biol. J. Linn. Soc.* **117,** 241–251 (2016).
67. Warnes, G. R. *et al.* gplots: various R programming tools for plotting data. R package version 2.17.0 (2015).
68. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.r-project.org/. (2016).
69. Bates, D., Mächler, M., Bolker, B. M. & Walker, S. C. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67,** 1–48 (2015).

Acknowledgements

We thank Marie B. Mouridsen for providing the *Drosophila* species used in the experiments, Martin Holmstrup for fruitful discussion and comments on an earlier version of the manuscript and Helle Blendstrup, Susan M. Hansen, Henriette C. Jensen, Iben V. Nielsen, Stinne B. Nielsen, and Jens C. Lundgaard for invaluable laboratory assistance. This research was supported by the Danish Natural Science Research Council through a Sapere aude stipend to TNK (DFF − 4002-00036).

Author Contributions

The experimental work was primarily conducted by M.Ø. M.Ø. and M.F.S. analysed the data. All authors planned the research and M.Ø. wrote first draft of the manuscript. T.N.K. and M.F.S. contributed to the scientific analysis and to writing the manuscript. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Ørsted, M. *et al.* Biotic and abiotic factors investigated in two *Drosophila* species – evidence of both negative and positive effects of interactions on performance. *Sci. Rep.* **7**, 40132; doi: 10.1038/srep40132 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

SUPPLEMENTARY INFORMATION FOR PAPER I

Text S1: Supplementary discussion.

Table S1: Details on experimental treatments.

Table S2: Number of flies used for each treatment and each trait.

Figure S1: Egg-to-adult viability results of range-finding test of dimethoate

Figure S2: Developmental time results of range-finding test of dimethoate

Figure S3: Photograph of the modified Rapid Iterative Negative Geotaxis (RING) assay apparatus

Figure S4: Heat map showing direction of model coefficients, and challenges in defining complex interactions in the traditional framework in *D. hydei*.

Figure S5: Heat map showing direction of model coefficients, and challenges in defining complex interactions in the traditional framework in *D. melanogaster*.

SUPPLEMENTARY DISCUSSION

The main point of this study is to present and discuss the results of a set of complex interactions in a multi-trait, multi-species analysis. In doing so, however, we realised certain problems with the classically defined synergism and antagonism terms. Others have discussed problems associated with the traditional definitions struggling to describe the situations of more complex outcomes, which seem to be fairly common when analysing interactions, both in laboratory experiments and in field studies (Crain *et al.* 2008; Darling & Côté 2008; Vanhoudt *et al.* 2012; Piggott *et al.* 2015). Congruent with such studies, we point to challenges with the typical direction-independent classification because of the issues and limitations of the traditional framework outlined here. Also, a large numbers of studies on multiple environmental factors report interactions based on imprecise descriptions or simply the qualitative judgement of the authors (Dunne 2010). Thus, in the scientific literature there is a lack of consensus on operationally robust definitions and quantification of synergism and antagonism (Folt *et al.* 1999; Darling & Côté 2008; Holmstrup *et al.* 2010; Laskowski *et al.* 2010; Vanhoudt *et al.* 2012).

The long-standing scientific classic definitions of synergism and antagonism are valid. We are merely proposing an expansion on the traditional definitions. As first proposed by Piggott *et al.* (2015), and in our work expanded to include three-way interactions, we utilised a system combining the 'interaction effect' (as in the classic effect deviation from the additive model prediction (Folt *et al.* 1999)), with the magnitude and direction of the response (+ or –) relative to individual treatment effects in absolute terms. Thus, the "re-defined" synergism and antagonism still pertain to the classic "more than" and "less than", respectively, as it is traditionally understood. The lack of consensus on definitions is most likely due to the usage of these terms throughout widely different scientific disciplines from ecology to toxicology and medicine. In toxicology and to some extent ecology (and thus ecotoxicology) interactions are frequently regarded as "stressful" and therefore exclusively detrimental to the overall performance of the subject species (Folt *et al.* 1999; Piggott *et al.* 2015). In this context of viewing interactions as always being negative, a synergistic interaction is defined as an interaction causing negative effect greater than predicted by an additive model and an antagonistic interaction as a negative effect that is less than predicted from additivity.

To highlight the problems and limitations of the classic framework in ecological interactions, especially related to the "always-negative" view on interactions, we tried to re-designate classically defined terms to the interactions observed in our dataset (Supplementary Figs. S4-S5). In doing so we identified several issues listed below, and compared our findings to other studies employing the traditional definition framework:

1) In situations where neither individual environmental condition has a significant effect on a trait but the interaction is significantly negative, interactions cannot be properly determined by classic definitions; e.g. when $0 + 0 < 0$. Attempting to classify these situations in the classic paradigm would always result in synergism i.e. "more negative than" the individual effects, as antagonism is interpreted as "less negative than" the individual effects (not pertaining to the result being positive or negative in terms of fitness/performance, but how the interaction relates to the additive expectation). Examples from the present study that represent this sort of challenge in defining interactions in the classic framework are marked with a superscript "*x*" in Supplementary Figs. S4-S5.

2) Even if still pertaining to the "all negative" nature of interactions, in situations where one individual factor has a negative effect while the other has no effect, and their cumulative effect is more negative than additively expected, the classic paradigm is also struggling. In (eco)toxicology this is sometimes referred to as potentiation or sensitisation (Odum & Barrett 2005) and some argue that it is not true "synergism" because it is one-sided and the underlying modes of action are different (Chou 2010). We have not included such further definitions, because it would confuse more than contribute, and we believe all situations are encompassed and informatively described by the re-conceptualised terms used in the present study. Examples from the present study that represent this sort of challenge in defining interactions in the classic framework are marked with a superscript "*y*" in Supplementary Figs. S4-S5.

In contrast to the persistent "all negative" view of interactions and the individual factors assessed, positive effects of individual environmental factors and even positive effects of interactions must be recognised. In ecotoxicology this phenomenon is often referred to as hormesis and is readily observed when assessing the effects of chemicals, e.g. at low dose (Boonstra *et al.* 2005; Holmstrup *et al.* 2010; Laskowski *et al.* 2010). While some "stressors" like chemicals are most frequently investigated as a gradient (concentrations), and thus might result in hormesis being observed at a low dose, other "stressors" are not as easily applied at a continuous scale e.g. biotic factors including co-occurring species or predation/parasitism, which is more of a presence/absence situation. Indeed, one could apply varying levels (densities) of co-occurring species or predators/parasites, but in a full-factorial study on interactions this would quickly scale to non-manageable proportions.

Even if accepting positive effects of interactions on performance or more importantly the positive effect (direction) of one or more individual factors, as

employed in several recent reviews on interactions (Crain *et al.* 2008; Piggott *et al.* 2015), we identify several issues using the classic definitions. While the identification of a synergism or antagonism is generally straightforward when all factors operate in the same direction (Folt *et al.* 1999; Dunne 2010), i.e. all positive (Fig. 1a) or all negative (Fig. 1c), problems arise when individual factors are of opposite directions (Fig. 1b). In such situations, the classic definition of synergism appears paradoxical because what is synergistic to the effect direction of one factor is antagonistic to the effect direction of the other factor(s):

3) By classic definitions it is difficult to classify interactions when the effects of two individual environmental conditions are in opposite direction e.g. $-1 + 1 > 1$ (see below for further discussion). Examples from the present study that represent this sort of challenge in defining interactions in the classic framework are marked with a superscript "*z*" in Supplementary Figs. S4-S5.

4) Crain *et al.* (2008), having accepted the presence of positive effects, assumed that in situations with two opposing individual effect directions, synergy only occurred when the cumulative effect was more negative than the additive sum of the opposing individual effects. Examples from the present study where we have utilised this definition to define synergism are marked with a superscript "*w*" in Supplementary Figs. S4-S5.

The definition of Crain *et al.* (2008) may be appropriate if the effect direction is implicitly negative, e.g. decreased survival rate, but in many other situations such a definition is problematic from an ecological perspective because effect direction can be context dependent (see e.g. discussion of the effect direction on developmental time in main manuscript). This assumption raises another conceptual issue in that the cumulative effect of factors of opposing directions are not necessarily more negative than the single negative stressor acting alone (see "comparative effects" model of Folt *et al.* (1999)). Consider an example of a factor, which when applied alone has a positive effect of +1 and a factor, which when applied alone has a negative effect of –1. The additive cumulative effect of both factors combined is 0, i.e. they neutralise each other. By classic assumptions, as that of Crain *et al.* (2008), we should invoke synergy for any cumulative effect more negative than 0. However, if the cumulative effect is between –1 and 0, this interaction is intuitively antagonistic from the perspective of the negative factor's individual effect (–1), i.e. the cumulative effect of both factors is less negative than the single negative stressor acting alone. In our proposed system, we would classify this as a *positive antagonism*, i.e. it is less positive than predicted from an additive model. Had the cumulative effect been between 0 and 1 we would classify it as a *negative antagonism*, i.e. it is less negative than predicted

additively. Thus, antagonism can be easily interpreted in the traditional sense of "less than" in terms of the cumulative effect relative to the effect of the individual effect size. The positive or negative prefix enables rapid interpretation of the direction relative to the cumulative effect, especially in these situations of opposing individual factors (Fig. 1b). We want to re-enforce that these prefixes does not describe the performance or fitness effect of the interaction, i.e. a *positive antagonism* is not necessarily beneficial to the organism, partly because it can be difficult to establish the relationship between an effect direction and its costs and benefits to performance in some traits, e.g. as in the case of developmental time (Niemelä *et al.* 2012). However this is not a problem specifically pertinent to our system, this is also a problem in the classic definitions framework.

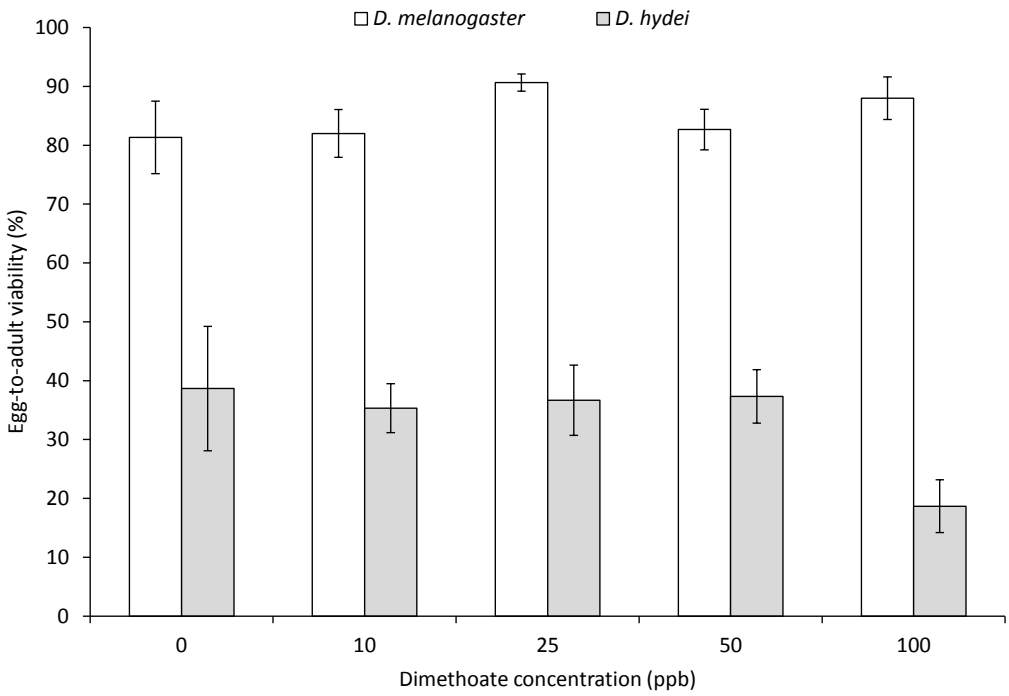
The system also includes a new form of synergy, referred to as "mitigating synergism", when individual environmental factors operating in the same direction interact and result in a cumulative effect in the opposite direction, e.g. two positives make a negative (S–) or two negatives make a positive (S+). Such strong interactions might be of great interest in predicting ecological consequences of multiple environmental factors, because different treatments can synergistically inhibit or mitigate the effect of individual factors (Holmstrup *et al.* 2010; Piggott *et al.* 2015). While we realise that these introduced interaction terms can seem unduly complicated, we believe that the re-conceptualized terms provide more informative descriptions and straightforward interpretations of complex interactions, which would be difficult to even describe in the classic context.

Supplementary Table S1: Environmental treatments in a full factorial design, showing temperature, dimethoate concentration, co-occurrence status, number of vials in a respective treatment and number of eggs per vial. Temp.: temperature. Dim.: dimethoate. Co-occur.: co-occurrence. *D. hydei*: hyd. *D. melanogaster*: mel.

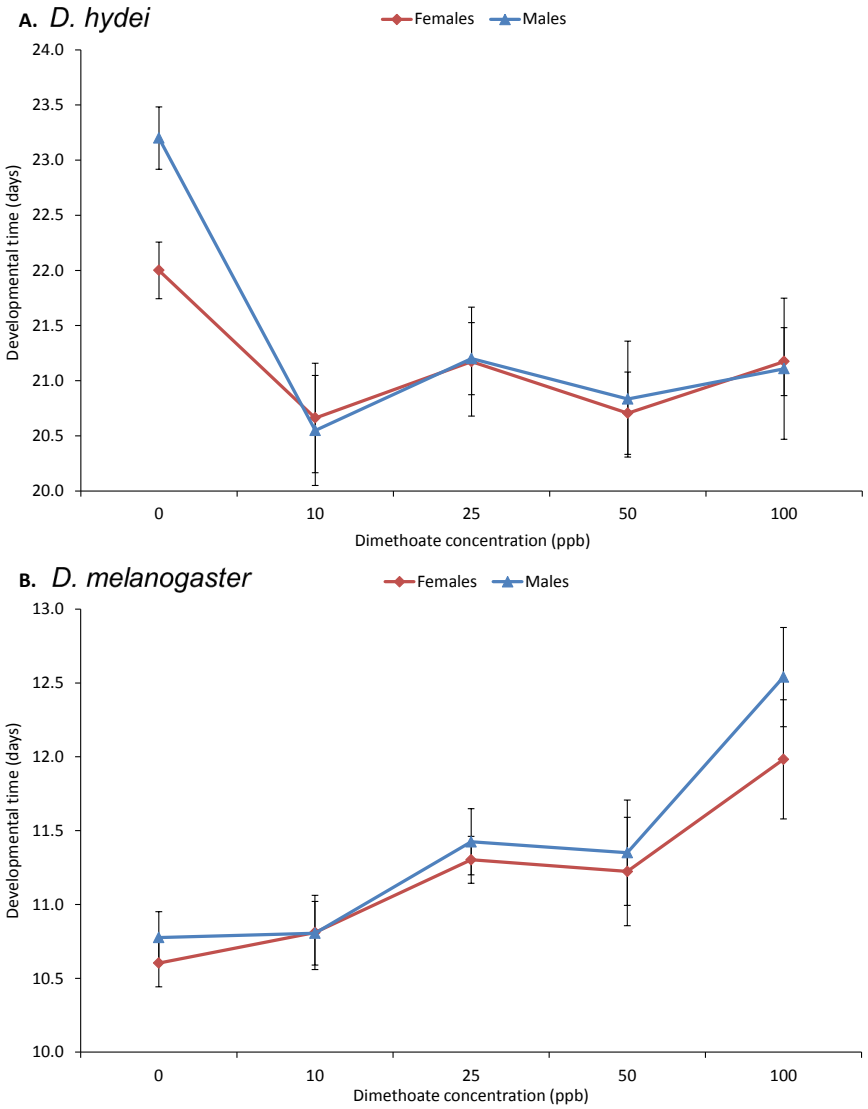
Identification code	Species	Temp. (°C)	Dim. (ppb)	Co-occur.	Number of vials	Number of eggs per vial
h-25-0	hyd	25	0	No	30	40
h-25-75	hyd	25	75	No	40	40
m-25-0	mel	25	0	No	30	40
m-25-75	mel	25	75	No	30	40
h/m -25-0	hyd/mel	25	0	Yes	30	20 of each species
h/m -25-75	hyd/mel	25	75	Yes	40	20 of each species
h-13-0	hyd	13	0	No	30	40
h-13-75	hyd	13	75	No	40	40
m-13-0	mel	13	0	No	30	40
m-13-75	mel	13	75	No	30	40
h/m-13-0	hyd/mel	13	0	Yes	30	20 of each species
h/m -13-75	hyd/mel	13	75	Yes	40	20 of each species
h-31-0	hyd	31	0	No	30	40
h-31-75	hyd	31	75	No	40	40
m-31-0	mel	31	0	No	30	40
m-31-75	mel	31	75	No	30	40
h/m -31-0	hyd/mel	31	0	Yes	30	20 of each species
h/m -31-75	hyd/mel	31	75	Yes	40	20 of each species

Supplementary Table S2: Number of flies used from each species from each treatment for CT_{min}, CT_{max}, developmental time, and RING for each sex (M; males, and F: females) and for egg-to-adult viability for both sexes combined. Average number of flies (and standard deviation (S.D.)) from each trait is also given. NA values indicate that no or too few flies emerged from a given treatment. The minimum number of measurements (limit *n*) for CT_{min}, CT_{max}, and developmental time was 5 and for RING and egg-to-adult viability it was 50 and 30, respectively. Dim.: dimethoate. Co-occur.: co-occurrence.

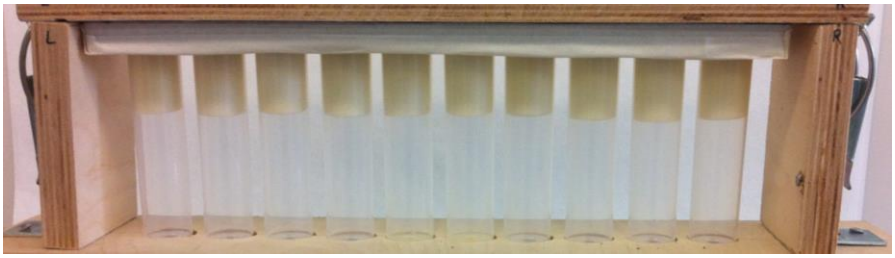
	Trait	**CT_{max}**		**CT_{min}**		**Developmental time**		**RING**		**Viability**
	Sex	**M**	**F**	**M**	**F**	**M**	**F**	**M**	**F**	**Both**
D. hydei	25 °C (Control)	8	17	9	20	119	116	100	100	30
	25 °C + Co-occur.	20	22	26	28	145	163	100	100	30
	25 °C + Dim.	21	19	30	20	162	129	100	100	30
	25 °C + Co-occur. + Dim.	19	17	27	26	68	65	100	100	40
	13 °C	NA	7	10	12	182	199	100	100	30
	13 °C + Co-occur.	22	16	13	16	172	137	100	50	30
	13 °C + Dim.	10	12	14	15	43	35	50	NA	30
	13 °C + Co-occur. + Dim.	NA	NA	NA	NA	NA	6	NA	NA	40
	31 °C	9	10	9	9	66	60	NA	100	30
	31 °C + Co-occur.	6	12	7	14	40	49	50	50	30
	31 °C + Dim.	10	8	10	8	46	54	100	50	30
	31 °C + Co-occur. + Dim.	20	20	19	19	134	102	100	100	40
	Average *n*	15	15	16	17	107	93	90	85	33
	S.D.	6	5	8	6	53	55	20	23	4
D. melanogaster	25 °C (Control)	20	19	20	20	466	494	100	100	30
	25 °C + Co-occur.	19	18	20	20	276	262	100	100	30
	25 °C + Dim.	19	19	19	20	424	434	100	100	30
	25 °C + Co-occur. + Dim.	18	17	20	18	354	368	100	100	40
	13 °C	27	20	18	20	440	516	100	100	30
	13 °C + Co-occur.	19	20	20	20	208	209	50	100	30
	13 °C + Dim.	20	20	19	20	439	428	100	100	30
	13 °C + Co-occur. + Dim.	20	20	19	20	306	313	100	100	40
	31 °C	19	20	20	20	475	491	100	100	30
	31 °C + Co-occur.	19	20	19	20	218	233	100	100	30
	31 °C + Dim.	19	20	20	20	492	495	100	100	30
	31 °C + Co-occur. + Dim.	19	20	20	19	301	326	100	100	40
	Average *n*	20	19	20	20	367	381	96	100	33
	S.D.	2	1	1	1	98	106	14	0	4
	Limit *n*	5	5	5	5	5	5	50	50	30



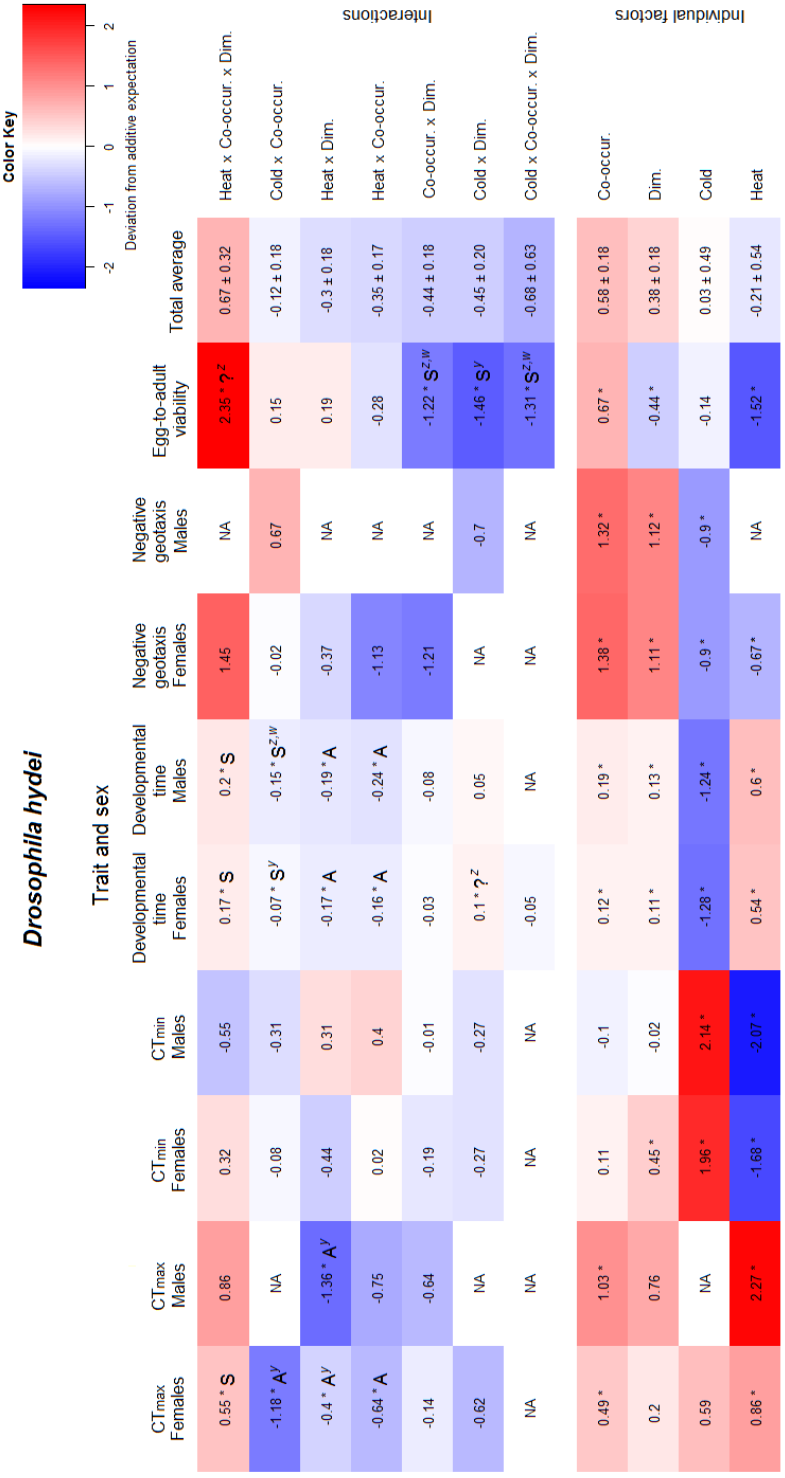
Supplementary Fig. S1: Preliminary screening of egg-to-adult viability (%) in *D. hydei* (grey filled bars) and *D. melanogaster* (white open bars) at a range of dimethoate concentration from 0 to 100 ppb. Error bars represent standard error (n = 5). The media setup, egg collection procedure and subsequent scoring of viability followed the same procedure as described in the main experiment (40 eggs per vial with 9 mL Formula 4-24® Instant *Drosophila* Medium Blue ± dimethoate).



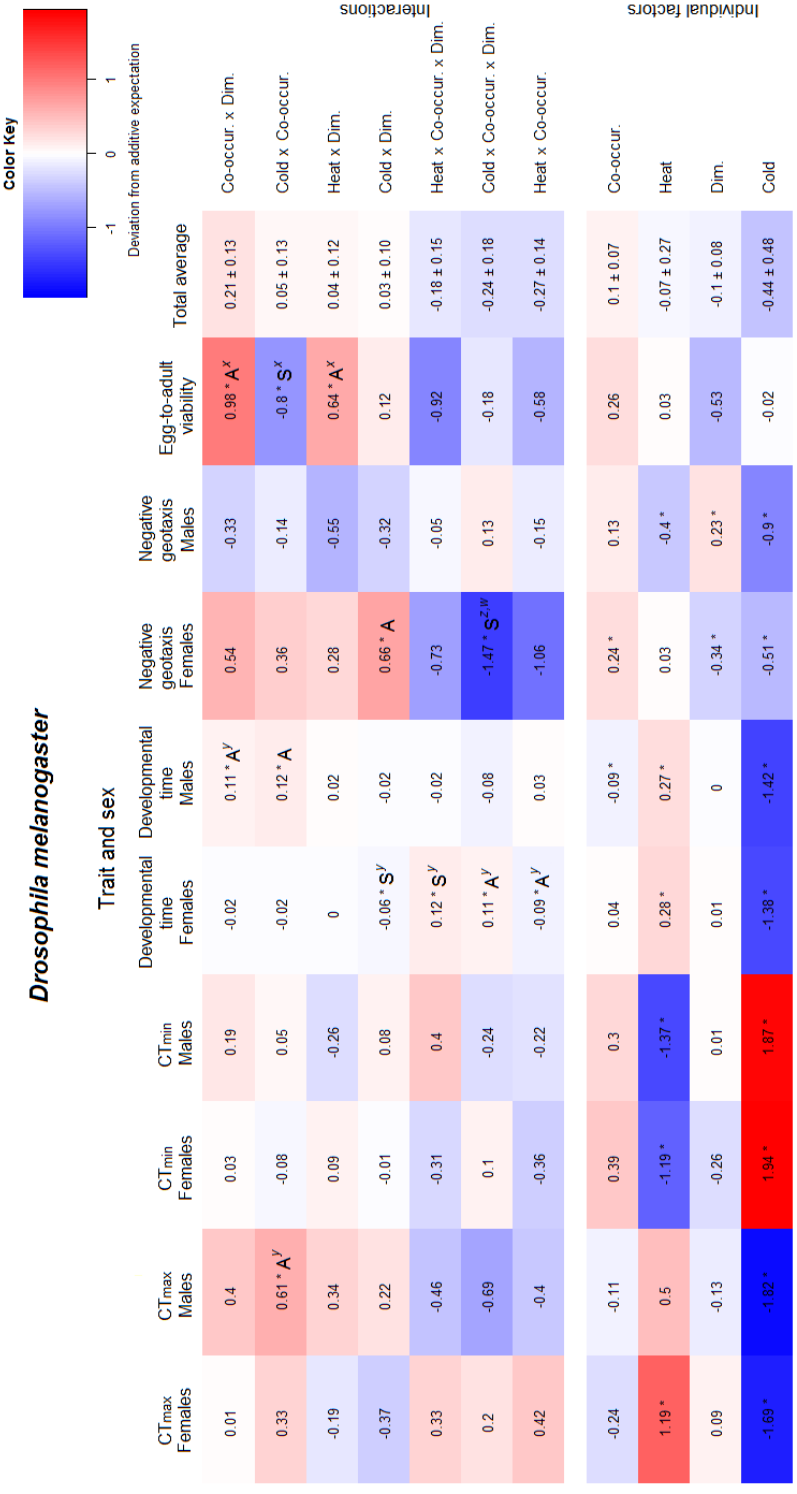
Supplementary Fig. S2: Preliminary screening of egg-to-adult developmental time (in days) in females (red) and males (blue) of *D. hydei* (**A.**) and *D. melanogaster* (**B.**) at a range of dimethoate concentration from 0 to 100 ppb. Error bars represent standard error (n = 5). The media setup, egg collection procedure and subsequent scoring of developmental time followed the same procedure as described in the main experiment (40 eggs per vial with 9 mL Formula 4-24® Instant *Drosophila* Medium Blue ± dimethoate).



Supplementary Fig. S3: Front side view of the Rapid Iterative Negative Geotaxis (RING) assay apparatus modified from Gargano et al. (2005). Our version of the RING apparatus was a custom built open-faced wooden box with a detachable lid held in place with lock hinges. The rack holds 10 empty vertical 27 mL vials. Into the floor of the apparatus was milled a 1 mm indentation as support for 10 empty vertical 27 mL vials. Foam stoppers were inserted to an equal depth across all 10 vials and the lid of the apparatus holds the negative geotaxis vials in place when locked with hinges. A camera was mounted 30 cm from the apparatus to capture photos of the negative geotaxis behaviour.


Drosophila hydei
Trait and sex
CTmax Females
CTmax Males
CTmin Females
CTmin Males
Developmental time Females
Developmental time Males
Negative geotaxis Females
Negative geotaxis Males
Egg-to-adult viability
Total average
Color Key
Deviation from additive expectation
Interactions
Individual factors
Heat x Co-occur. x Dim.
Cold x Co-occur.
Heat x Dim.
Heat x Co-occur.
Co-occur. x Dim.
Cold x Dim.
Cold x Co-occur. x Dim.
Co-occur.
Dim.
Cold
Heat

Supplementary Fig. S4: Heat map showing the direction and magnitude of the model coefficients reflecting the effects of treatments on egg-to-adult viability, developmental time, CT_{min}, CT_{max} and negative geotaxis (RING assay) in *D. hydei*. The effects are shown for both sexes in all traits except egg-to-adult viability. Positive coefficients represent positive deviation from the additive expectation, and can thus be interpreted as a performance advantage of the interaction itself, regardless of whether the treatment overall was beneficial in terms of performance when compared to the control. Contrary, a negative coefficient implies a negative deviation from additivity and that the interaction itself is detrimental to performance for a given trait. The direction of the effect is illustrated by colour shading from blue (negative) to red (positive) and the values indicate the strength of the effects. The upper part includes all two- and three-way interactions between heat or cold, co-occurrence (Co-occur.), and dimethoate (Dim.). The lower part includes all effects of the individual factors. Within each part the treatments (rows) have been sorted by the average total effect, i.e. the average effect across traits ± S.E., in descending order. An asterisk indicates a significant interaction, or a significant effect of the individual environmental factor. The direction of the interaction has been determined based on the traditional definitions of synergism (S) and antagonism (A). In doing so we identified several issues; the nature of these challenges is marked with subscripts $^{w, x, y}$ or ^{z} next to the designation of the interaction. Some interactions might relate to several issues and are given multiple subscript characters, and some interactions simply could not be determined based on contradicting definitions of the classic framework (designated with a question mark). The details of these challenges are described in the text in the Supplementary discussion. Some treatments did not yield enough live adult flies for assessing all traits or did not exceed the minimum number of flies needed for assessing a trait. In a few traits the effect of an individual environmental factor could therefore not be determined, and the interactions involving the particular factor were omitted from the model. Both cases are designated NA.


Color Key
-1
0
1
Deviation from additive expectation
Drosophila melanogaster
Trait and sex
CTmax Females
CTmax Males
CTmin Females
CTmin Males
Developmental time Females
Developmental time Males
Negative geotaxis Females
Negative geotaxis Males
Egg-to-adult viability
Total average
Co-occur. x Dim.
Cold x Co-occur.
Heat x Dim.
Cold x Dim.
Heat x Co-occur. x Dim.
Cold x Co-occur. x Dim.
Heat x Co-occur.
Interactions
Co-occur.
Heat
Dim.
Cold
Individual factors

Supplementary Fig. S5: Heat map showing the direction and magnitude of the model coefficients reflecting the effects of treatments on egg-to-adult viability, developmental time, CT_{min}, CT_{max} and negative geotaxis (RING assay) in *D. melanogaster*. The effects are shown for both sexes in all traits except egg-to-adult viability. Positive coefficients represent positive deviation from the additive expectation, and can thus be interpreted as a performance advantage of the interaction itself, regardless of whether the treatment overall was beneficial in terms of performance when compared to the control. Contrary, a negative coefficient implies a negative deviation from additivity and that the interaction itself is detrimental to fitness. The direction of the effect is illustrated by colour shading from blue (negative) to red (positive) and the values indicate the strength of the effects. The upper part includes all two- and three-way interactions between heat or cold, co-occurrence (Co-occur.), and dimethoate (Dim.). The lower part includes all effects of the individual factors. Within each part the treatments (rows) have been sorted by the average total effect, i.e. the average effect across traits ± S.E., in descending order. An asterisk indicates a significant interaction, or a significant effect of the individual environmental factor. The direction of the interaction has been determined based on the traditional definitions of synergism (S) and antagonism (A). In doing so we identified several issues; the nature of these challenges is marked with subscripts $^{w, x, y}$ or ^{z} next to the designation of the interaction. Some interactions might relate to several issues and are given multiple subscript characters. The details of these challenges are described in the text in the Supplementary discussion.

SUPPLEMENTARY REFERENCES

Boonstra R, Manzon RG, Mihok S, Helson JE (2005) Hormetic effects of gamma radiation on the stress axis of natural populations of meadow voles (*Microtus pennsylvanicus*). *Environmental Toxicology and Chemistry*, **24**, 334–343.

Chou TC (2010) Drug combination studies and their synergy quantification using the Chou-Talalay method. *Cancer Research*, **70**, 440–446.

Crain CM, Kroeker K, Halpern BS (2008) Interactive and cumulative effects of multiple human stressors in marine systems. *Ecology Letters*, **11**, 1304–1315.

Darling ES, Côté IM (2008) Quantifying the evidence for ecological synergies. *Ecology Letters*, **11**, 1278–1286.

Dunne RP (2010) Synergy or antagonism-interactions between stressors on coral reefs. *Coral Reefs*, **29**, 145–152.

Folt C, Chen C, Moore M, Burnaford J (1999) Synergism and antagonism among multiple stressors. *Limnology and Oceanography*, **44**, 864–877.

Gargano JW, Martin I, Bhandari P, Grotewiel MS (2005) Rapid iterative negative geotaxis (RING): A new method for assessing age-related locomotor decline in *Drosophila*. *Experimental Gerontology*, **40**, 386–395.

Holmstrup M, Bindesbøl AM, Oostingh GJ *et al.* (2010) Interactions between effects of environmental chemicals and natural stressors: a review. *Science of the Total Environment*, **408**, 3746–3762.

Laskowski R, Bednarska AJ, Kramarz PE *et al.* (2010) Interactions between toxic chemicals and natural environmental factors - a meta-analysis and case studies. *Science of the Total Environment*, **408**, 3763–3774.

Niemelä PT, Vainikka A, Hedrick A V., Kortet R (2012) Integrating behaviour with life history: boldness of the field cricket, *Gryllus integer*, during ontogeny. *Functional Ecology*, **26**, 450–456.

Odum EP, Barrett GW (2005) *Fundamentals of Ecology*. Thomson/Brooks/Cole, Belmont, CA, USA.

Piggott JJ, Townsend CR, Matthaei CD (2015) Reconceptualizing synergism and antagonism among multiple stressors. *Ecology and Evolution*, **5**, 1538–1547.

Vanhoudt N, Vandenhove H, Real A, Bradshaw C, Stark K (2012) A review of multiple stressor studies that include ionising radiation. *Environmental Pollution*, **168**, 177–192.

PAPER II

METABOLIC AND FUNCTIONAL PHENOTYPIC PROFILING OF *DROSOPHILA MELANOGASTER* REVEAL REDUCED SEX DIFFERENTIATION UNDER STRESSFUL ENVIRONMENTAL CONDITIONS

In print in *Biological Journal of the Linnean Society*

MICHAEL ØRSTED, ANDERS MALMENDAL, JOAQUIN MUÑOZ &
TORSTEN NYGAARD KRISTENSEN

Metabolic and functional phenotypic profiling of *Drosophila melanogaster* reveal reduced sex differentiation under stressful environmental conditions

Michael Ørsted[1], Anders Malmendal[2], Joaquin Muñoz[1],
and Torsten Nygaard Kristensen[1,3]

[1] *Department of Chemistry and Bioscience, Aalborg University, Aalborg E, Denmark*
[2] *Center for Molecular Protein Science, Lund University, Lund, Sweden*
[3] *Department of Bioscience, Aarhus University, Aarhus C, Denmark*

ABSTRACT

Strong sexual dimorphism is commonly observed across species and e.g. trade-offs between reproduction and maintenance are thought to explain this dimorphism. Here we test how the metabolic and functional phenotypic responses to varying types of environmental stress differ in male and female Drosophila melanogaster (Diptera: Drosophilidae), and how stress impacts the magnitude of sexual dimorphism. Experimental stressors that we exposed flies to during development were heat stress, poor nutrition, high acidity, high levels of ammonia and ethanol. Emerged male and female flies from the different rearing regimes were investigated using NMR metabolomics and assessed for body mass and viability. Our results showed that environmental stress leads to reduced sexual dimorphism in both metabolic composition and body mass compared to the level of dimorphism observed at benign conditions. This reduced sexual dimorphism in stressful environments might be caused by a lower investment in sex specific characteristics under such conditions, and our results provide support for the longstanding idea that ecological factors are important for shaping sexual dimorphism and possibly sexual selection.

Keywords: NMR metabolomics - Environmental stress - Sex differentiation - Drosophila melanogaster - Functional phenotypes

INTRODUCTION

Many species show strong sex differentiation in morphological, physiological, behavioural and life-history traits. Commonly observed examples include longer lifespan in females (Shen *et al.* 2009; Niveditha *et al.* 2017), sexual size dimorphism

with females typically being larger in e.g. insects and many bird species and males being larger in many mammals (Dunn *et al.* 2001; Stillwell *et al.* 2010), increased stress resistance in females (Fountain *et al.* 2015; Ørsted *et al.* 2017) and distinct molecular phenotypes between males and females revealed by transcriptomic, proteomic, and metabolomic profiling (Hines *et al.* 2007; Findlay *et al.* 2008; Schou *et al.* 2017). Sexual selection and trade-offs between reproduction and lifespan or between size and stress resistance are thought to be main evolutionary drivers of genetically based sexual dimorphism (Hedrick & Temeles 1989; Dunn *et al.* 2001; Matzkin *et al.* 2009). However, sex differences in phenotypic plasticity in responses to environmental stimuli may also explain variation between sexes (Fischer & Fiedler 2001; Fernández-Montraveta & Moya-Laraño 2007; Stillwell *et al.* 2010; Ketola *et al.* 2012).

Sex specific responses to environmental stress have been shown in numerous studies, especially in arthropods such as bed bugs (Fountain *et al.* 2015), butterflies (Fischer & Fiedler 2001), neriid flies (Cassidy *et al.* 2014), mites (Walzer & Schausberger 2011), and *Drosophila* (Ketola *et al.* 2012; Schou *et al.* 2017; Ørsted *et al.* 2017). However, we still have little knowledge as to whether diverse forms of stress change the degree of sexual dimorphism, i.e. whether male and female responses to environmental stress differ. This would be evident from a stress induced directional increase or decrease in sexual dimorphism when exposed to stress. Such sex specific responses to environmental stress would potentially have strong implications for the evolution of sexual dimorphism and sexual selection. We hypothesize that sex specific abilities to induce plastic responses to environmental stress will alter sexual dimorphism both at the molecular and functional phenotypic levels. We propose that investments in sex specific characteristics such as production of eggs or sperm, mating, fat deposition, ornamentation, size, and sex hormones are reduced when environments gets harsh and resources are scarce.

To assess consequences of environmental stress on sexual dimorphism, we reared a population of *Drosophila melanogaster* (Meigen) at a benign laboratory environment and at a range of ecologically relevant stressful environmental conditions throughout development from egg to adult for one generation and compared male and female metabolomic and functional phenotypic responses. Here we defined functional phenotypes as egg-to-adult viability and body mass as parts of the full organismal phenotype. Stressors represented environmental conditions commonly experienced by some insects i.e. heat stress, poor nutrition, high acidity, and high levels of ammonia and of ethanol. We utilised nuclear magnetic resonance (NMR) metabolomics which is a highly reproducible technique that provides quantitative data on abundances of sugars, nucleotides, amino acids, and lipids (Markley *et al.* 2017). We chose to investigate the metabolome in this study because focus was on detecting sex specific molecular signatures of stress exposure that are more closely related to the organismal

phenotype than for example data obtained from transcriptomic or proteomics platforms (Lankadurai *et al.* 2013). Additionally, it has been suggested that metabolomics is a more sensitive indicator of external environmental stress than other 'omics' techniques, which are subjected to a range of feedback mechanisms and homeostatic controls (Nicholson *et al.* 1999; van Ravenzwaay *et al.* 2007).

MATERIALS AND METHODS

The *D. melanogaster* population originated from flies caught in Denmark (Schou *et al.* 2015), and was maintained as previously described (Ørsted *et al.* 2017) on a standard *Drosophila* sucrose-yeast-agar medium until initiating our experiments. We investigated the effects of exposing flies to six different treatments by varying media contents and thermal environments throughout development from the egg to the adult life stage; control (23°C; standard medium (SM)), heat (32°C; SM), low nutrition (23°C and medium with only 16 g L^{-1} agar and 3 g L^{-1} yeast), high acidity (23°C; SM+0.70 M acetic acid), ammonia (23°C; SM+0.40 M ammonia), and ethanol (23°C; SM+6 vol% ethanol) (Supplementary Table S1). Twenty 3-4 days old parental flies deposited eggs on spoons with 1.5 mL of standard medium at 23°C. Twelve hours later, 40 eggs were transferred to each of 25 replicate vials per treatment containing 10 mL of the respective experimental media, and the vials were transferred to the respective thermal environments.

Vials were checked every 24 hours and emerged flies were counted in an empty plastic vial by flash freezing in liquid nitrogen, and immediately stored in a tube at –80°C. This continued until no flies from a given treatment had emerged for 72 hours, after which all the frozen flies from each treatment were pooled and sexed on an ice block to prevent thawing. Egg-to-adult viability was determined as the proportion of eggs in each vial that successfully developed into an adult fly. Sex specific viability was determined at the treatment level assuming a 1:1 sex ratio of the 1000 eggs per treatment. Dry body mass of ~20 individual flies of each sex per treatment (Table 1) was determined by drying flies at 60°C for 24 h and weighing them on a Sartorius laboratory scale (Quintix35-1S, Göttingen, Germany). The difference in dry body mass is calculated as the difference in means, and the SE of this difference is calculated from the variance sum law: $\sqrt{\frac{\sigma_F^2}{n_F} + \frac{\sigma_M^2}{n_M}}$, where σ^2 are the variances and n are the sample sizes of females (F) and males (M) (see Table 1 for exact sample sizes).

For metabolite extraction, we used five replicates of 15 males and five of 15 females randomly sampled from the pool of flies from each treatment. Sterile glass beads and 1 mL chilled acetonitrile solution (50 vol% in double distilled water) were added to each sample and cooled on ice. The samples were homogenized using a FastPrep®24 tissuelyser (MP Biomedicals, Solon, OH) running 3,800 rpm for 35

seconds, set back on ice, then centrifuged at 14,000 rpm for 10 min at 4°C. The supernatant (900 µL) was transferred to fresh tubes, snap frozen, freeze-dried and stored at –80°C. Immediately before NMR measurements, samples were rehydrated in 200 mL of 50 mM phosphate buffer (pH 7.4) in D_2O, and 180 mL was transferred to a 3 mm NMR tube. The buffer contained 50 mg L^{-1} chemical shift reference 4,4-dimethyl-4-silapentane-1-sulfonic acid-D6, sodium salt (DSS), and 50 mg L^{-1} sodium azide to prevent bacterial growth. NMR measurements were performed at 25°C on a Bruker Avance III HD 800 spectrometer (Bruker Biospin), operating at a ^1H frequency of 799.87 MHz, with a 3 mm TCI cold probe. ^1H NMR spectra were acquired using a 1D NOESY experiment. The water signal was suppressed by presaturation. A total of 64 transients of 32K data points spanning a spectral width of 20 ppm were collected. NMR spectra were processed, normalized to total intensity and pareto scaled as previously described (Schou *et al.* 2017), except that the spectra were referenced to the DSS signal at 0.000 ppm.

Principal Component Analysis (PCA) was performed to assess the effect of environment and sex on the metabolome. To investigate the differentiation in the metabolome between sexes across environments we performed a PCA based on the difference between individual female sample spectra and the median male spectrum and *vice versa* in each environment (Schou *et al.* 2017). Hierarchical Cluster Analysis (HCA) of the Euclidean distance between PCA scores was carried out using Ward's method in R (R Core Team 2017), to enable visualization of the differentiation in a dendrogram. The metabolites that change with different functional phenotypes were efficiently predicted using Orthogonal Projections to Latent Structures (OPLS) models (Trygg & Wold 2002) as previously described (Schou *et al.* 2017). An OPLS model is similar to a PCA but here the first predictive component describes the metabolite variation that shows the highest correlation with a certain phenotypic measure. Likewise, OPLS Discriminant Analysis (OPLS-DA) (Bylesjö *et al.* 2006) was used to predict metabolites that vary between groups of flies from different rearing environments. The significant correlations were calculated with sequential Bonferroni correction ($P<0.05$) for an assumed total number of 100 metabolites. NMR chemical shifts and the primary range used for identification of correlations can be seen in Supplementary Table S2. Egg-to-adult viability data was analysed with a quasi-binomial generalised linear model (to correct for over-dispersion) with a logit link function. Multivariate analysis was performed using the SIMCA14 software (Umetrics, Malmö, Sweden).

RESULTS AND DISCUSSION

We considered all the experimental treatments to be stressful as they all significantly reduced body mass (one-way ANOVA; $F_{(5,111)}$=28.3; P<0.001), and all except the ethanol treatment significantly reduced egg-to-adult viability (GLM; t<-3.91; P<0.001) compared to the control treatment (Table 1; Supplementary Fig.S1A-B).

To describe the metabolic responses to the varying stressful environmental conditions, metabolite NMR spectra were analysed by PCA. A qualitative inspection of the PCA scores reveals a differential metabolite response (MANOVA; P<3x10^{-22}) of males and females across the different treatments (Fig.1A). Specifically, females from the control treatment are different from both male control flies and all stressed flies (Fig.1A). This suggests that one major effect of stress treatment is that metabolites specific for females are attenuated and the metabolome gets more 'male-like'. Separate analyses of the sexes show that in males, reduced nutrition and heat resulted in the largest differences in metabolite composition, while the remaining

Table 1. Functional phenotypic results. Results of dry body mass and egg-to-adult viability for males and females (means, SE, and n) of the six experimental treatments. Asterisks denote significant difference from the control, as determined by one-way ANOVA and Welch's t-tests for dry body mass and by generalised linear models for egg-to-adult viablity. The effects of the treatments are also expressed as the difference in % from the benign control condition, which is indexed as 100%.

Trait	Sex	Treatment	Control	Heat	Low nutrition	Acid	Ammonia	Ethanol
Dry body mass	Males	Mean (mg)	0.283	0.212*	0.155*	0.257*	0.231*	0.246*
		SE	0.008	0.008	0.007	0.010	0.009	0.006
		N	20	17	20	20	20	20
		Difference from control (%)		-24.8	-45.3	-9.0	-18.2	-13.1
	Females	Mean (mg)	0.461	0.283*	0.202*	0.371*	0.324*	0.365*
		SE	0.018	0.011	0.017	0.014	0.012	0.012
		N	20	20	20	20	20	20
		Difference from control (%)		-38.5	-56.2	-19.4	-29.8	-20.7
Egg-to-adult viability	Males	Mean (%)	87.4	19.4*	56.0*	81.6*	52.4*	86.2
		SE	1.42	2.20	1.97	1.35	2.01	1.07
		N	25	25	25	25	25	25
		Difference from control (%)		-77.8	-35.9	-6.6	-40.0	-1.4
	Females	Mean (%)	96.6	18.0*	53.2*	86.2*	44.8*	93.8
		SE	1.42	2.20	1.97	1.35	2.01	1.07
		N	25	25	25	25	25	25
		Difference from control (%)		-81.4	-44.9	-10.8	-53.6	-2.9

treatments; acid, ammonia and ethanol are clustered with the control (Fig. 1B). In females, however, the controls are different from the acid, ammonia and ethanol treated flies (Fig. 1C). When examining how the difference between male and female metabolomes changes with stress, all stressors result in responses that are distinct from the control condition in a largely linear trajectory from benign control over acid, ammonia, ethanol and nutrition to heat stress (Fig.1D). The metabolite changes associated with Fig. 1A and 1D are listed in Supplementary Table S3.

To investigate how the sex-specific changes in metabolites were associated with the functional phenotypes we correlated PC1 explaining 33% of the variation in the sex-differentiation PCA (Fig.1D) with viability (a proxy for stress severity). Interestingly, the PC1 scores correlate significantly with stress severity (Fig.2A; R=-0.73; $t_{(58)}$=-8.10; P<0.001). It is noteworthy that the sign of the sex differentiation along PC1 is reversed for the heat stressed flies relative to the rest. The fact that the difference between male and female metabolomes continue changing along an almost linear trajectory, even after the sex difference under benign conditions has disappeared, suggests that it is the same mechanism that is adjusted by stress in all the stressful environments. The total metabolite difference between males and females was lower for all stressful conditions; total sex difference was 58, 68, 59, 50 and 89% of that of control flies for ethanol, acid, ammonia, low nutrition, and heat stress, respectively (calculated as the square root of the sum of squares). Assuming that viability is a good predictor of the severity of the stress induced by the treatments, our results suggest that sexual dimorphism at the metabolite level decrease in stressful environments compared to a benign environment. The sex difference in metabolite composition in the heat treatment was higher than for the other stress treatments. It was reversed in sign relative to the control resulting in a reversed sign of PC scores (Fig. 1A, 1D, 2A).

Although sex specific responses to environmental stress have been shown in numerous arthropods (Fischer & Fiedler 2001; Walzer & Schausberger 2011; Ketola *et al.* 2012; Cassidy *et al.* 2014; Fountain *et al.* 2015; Ørsted *et al.* 2017), we are not aware of studies that have investigated the actual magnitude of the sexual dimorphism in the metabolite response across different forms of stress. Schou *et al.* (2017) investigated the metabolite response across a developmental temperature gradient and found, in accordance with our findings, the largest sex differentiation at benign temperatures compared to either cold or warm extremes. At the functional phenotypic level, females were not only larger but they also showed a more plastic stress response in dry body mass compared to males (Two-way ANOVA; $F_{(1,230)}$=6.96; P=0.008; Supplementary Fig.S1C). This was not simply because females had more to lose because of their bigger body mass, in fact relative to their respective controls females weigh less than males under stressful conditions (Supplementary Fig.S1D). The difference in male and female body mass was correlated with viability: the lower the

survival, the lower the sex difference in body mass (R=-0.74; $t_{(115)}$=-11.85; P<0.001; Fig.2B). This means that the observations at the functional phenotypic level concur with the results of the metabolite analyses in that stress not only reduces sexual dimorphism compared to benign environments, but also that this reduction is linearly correlated with the severity of the stressor. Thus, our results suggest, in concordance with other studies, that NMR metabolomics provide results that are closely linked to the organismal phenotype and that it can be utilised as a sensitive indicator of environmental stress (van Ravenzwaay *et al.* 2007; Lankadurai *et al.* 2013). In many species of insects, body size is more plastic in females than in males (Stillwell *et al.* 2010), but only few studies have explicitly investigated the relationship between the magnitude of sex differentiation and stress severity. Walzer & Schausberger (2011) observed a linear decrease in sexual size dimorphism with increasing stress (lower prey densities) in mites, congruent with our findings.

We used OPLS-DA and OPLS models (Supplementary Table S4) to identify metabolites that were different between males and females under benign control conditions, between female control flies and female stressed flies, and to identify the trajectory along which this difference changes between stresses (Fig.1D), as well as which metabolites correlated with different functional phenotypes. The identified metabolite changes are listed in Table 2. Interestingly, some of the metabolites that are found in highest concentrations in the control females compared to both male control flies and to stressed females (Table 2) seem to be related to reproduction. These include betaine, choline and methionine which are important for neural development and reproductive success (Zeisel 2009; Dick *et al.* 2011). Glucose on the other hand was found in lower concentrations in control females compared to females under stress. Increased levels of circulating glucose has been associated with increased resistance to oxidative and starvation stress, while reducing female fecundity in *Drosophila* (Broughton *et al.* 2005). Although of descriptive character, this suggest that the reduced sexual dimorphism in stressful environments is caused by a trade-off between reproduction and stress resistance; lower investment in sex specific characteristics such as reproduction and more investment in stress response mechanisms that improve chances of survival. Such trade-offs between reproduction and life-span, and often also stress resistance are commonly observed (Broughton *et al.* 2005; Partridge *et al.* 2005; Shen *et al.* 2009; Niveditha *et al.* 2017). Interestingly, none of the identified metabolites were unique for either sex, so the stress responses were more related to differences in concentrations rather than the presence or absence of sex specific metabolites.

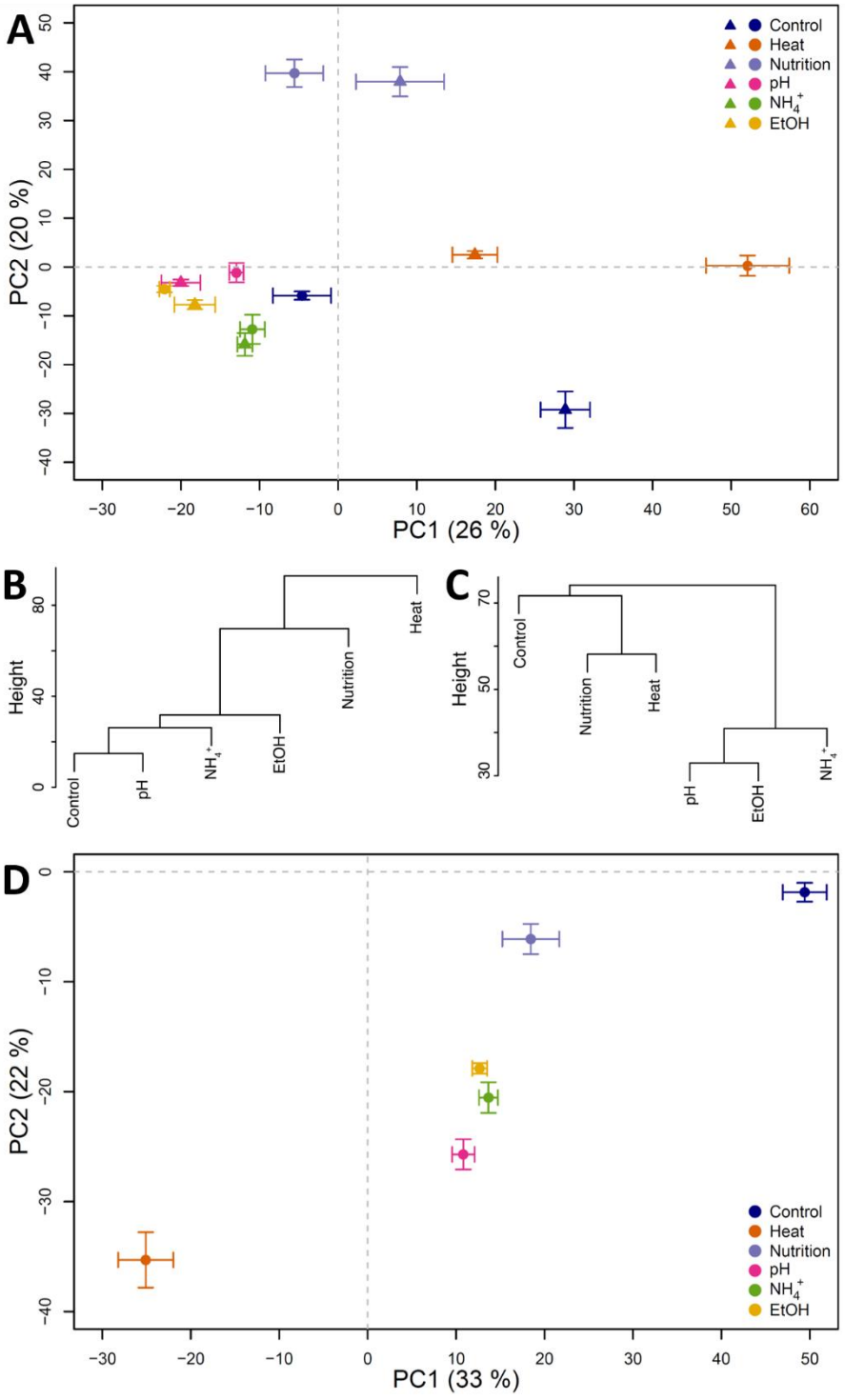

A
Control
Heat
Nutrition
pH
NH4+
EtOH
PC2 (20 %)
PC1 (26 %)
B
Height
Heat
Nutrition
NH4+
EtOH
Control
pH
C
Height
Control
Nutrition
Heat
pH
EtOH
NH4+
D
PC2 (22 %)
PC1 (33 %)
Control
Heat
Nutrition
pH
NH4+
EtOH

Fig. 1. Principal Component Analysis (PCA) scores plots and dendrograms describing the overall metabolite variation across environmental conditions; control, heat, low nutrition, acid (pH), ammonia (NH_4^+), and ethanol (EtOH). (A) PCA scores for all samples (circles; males, triangles; females). Error bars represent SE (n=5). Dendrograms based on hierarchical cluster analysis of PCA scores for males (B) and females (C). The length of the vertical axis is a measure of the dissimilarities between clusters of treatments. (D) PCA scores for the sex difference between females and males based on a PCA of the difference between individual female sample spectra and the median male spectrum and *vice versa* in each environment (Schou *et al.* 2017). Error bars represent SE (n=8, i.e. 10 measurements – 2 medians). The loadings associated with the scores displayed in panel A and D are shown in Supplementary Table S3.

At the functional phenotypic level, females were not only larger but they also showed a more plastic stress response in dry body mass compared to males (Two-way ANOVA; $F_{(1,230)}$=6.96; P=0.008; Supplementary Fig.S1C). This was not simply because females had more to lose because of their bigger body mass, in fact relative to their respective controls females weigh less than males under stressful conditions (Supplementary Fig.S1D). The difference in male and female body mass was correlated with viability: the lower the survival, the lower the sex difference in body mass (R=-0.74; $t_{(115)}$=-11.85; P<0.001; Fig.2B). This means that the observations at the functional phenotypic level concur with the results of the metabolite analyses in that stress not only reduces sexual dimorphism compared to benign environments, but also that this reduction is linearly correlated with the severity of the stressor. Thus, our results suggest, in concordance with other studies, that NMR metabolomics provide results that are closely linked to the organismal phenotype and that it can be utilised as a sensitive indicator of environmental stress (van Ravenzwaay *et al.* 2007; Lankadurai *et al.* 2013). In many species of insects, body size is more plastic in females than in males (Stillwell *et al.* 2010), but only few studies have explicitly investigated the relationship between the magnitude of sex differentiation and stress severity. Walzer & Schausberger (2011) observed a linear decrease in sexual size dimorphism with increasing stress (lower prey densities) in mites, congruent with our findings.

Table 2. Metabolite changes correlated to stress regime, sex, and functional phenotypes. We calculated the correlation coefficient (R) between individual metabolite intensities and the OPLS predictive components of models (Supplementary Table S4) to identify metabolites that are different between males and females under benign control conditions, different between female control flies and female stressed flies, linear change in the sex difference with stress (Fig.1D). We also performed OPLS modelling of egg-to-adult viability and the difference in body mass between males and females, based on the sex specific metabolite changes. A plus sign (+) denotes an significant increase in concentration in female control flies relative to male control flies, in female control flies relative to female stressed flies, in stressed flies along the linear trajectory in Fig. 1D, a positive correlation with egg-to-adult viability, and with the sex difference in dry weight. A minus sign denotes the opposite. Significant spectral correlations were identified by applying sequential Bonferroni correction ($P<0.05$) for an assumed total number of 100 metabolites. Only significant correlations are presented. *Full name of the galactoside is 1-O-(4-O-(2-aminoethyl phosphate)-ß-D-galactopyranosyl)-glycerol.

Metabolite	**Female control vs. male control**	**Female control vs. stressed females**	**Linear stress trajectory for sex difference**	**Egg to adult viability**	**Sex difference in dry weight**
Alanine		+	-		
Aspartate	+		-	+	+
Betaine	+	+	-	+	+
Choline		+	-	+	+
Fatty acid	+	+	-	+	+
Glucose	-	-	+	-	-
Glutamate		+	-	+	+
Glutamine	-		+	-	-
Galactoside*	-				+
Isoleucine		+	-	+	+
Leucine		+	-	+	+
Maltose	-	-	+	-	-
Methionine sulphoxide	+	+	-	+	+
Methionine	+	+	-	+	+
Phenylalanine		+	-	+	+
Phosphocholine			-	+	
Proline			-	+	+
Threonine	+	+	-	+	
Tyrosine			+		
Uridine	+	+	-	+	+
Valine		+	-		+

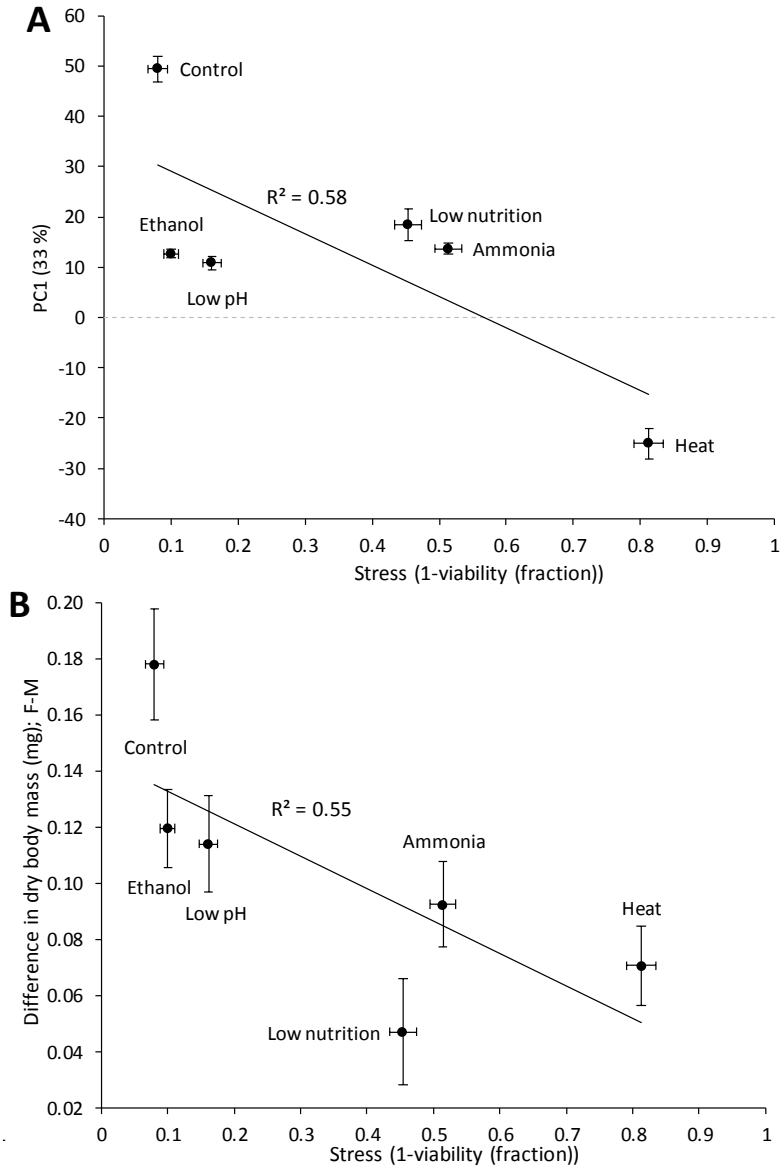


Figure 2. Relationship between stress severity (measured by egg-to-adult viability) and PC1 of the sex differentiation metabolite model (A), and difference in dry body mass between females and males (B). The mean viability for both sexes is used here for each vial (n=25). PC1 scores of the sex differentiation metabolite model are from Fig. 1D (n=8). Difference in dry body mass is calculated as the difference between mean female and male body mass. The SE of this difference is calculated from the variance sum law as: $\sqrt{\frac{\sigma_F^2}{n_F} + \frac{\sigma_M^2}{n_M}}$, where σ^2 are the variances and n are the sample sizes of females (F) and males (M), see Table 1 for exact sample sizes.

In conclusion, our results suggest that under stressful conditions the male metabolome is rather robust, while females are more sensitive, and become more similar to males in their metabolic composition. This is also evident in the different responses in body mass to increasing stress, where females show a more plastic response, and become more similar to males. Thus, we show a very high degree of concordance between stress responses on both the metabolic and a functional phenotypic level, where we observed a decrease in sexual dimorphism under stressful conditions. Numerous suggestions are put forward in the literature to explain why sexual dimorphism is commonly observed in natural and domestic populations, and sexual selection and ecological factors are commonly proposed explanations for differences between males and females, thoughts that go all the way back to Darwin and Wallace (Darwin 1874; Wallace 1889). Ketola *et al.* (2012) found that the heritability for heat tolerance was differentially affected by developmental temperature for males and females, and that most of the genetic variation for the trait was genetically uncorrelated in the two sexes, allowing independent evolution. More sophisticated designs than ours are needed to enable discriminating between these and other non-mutually exclusive alternatives (see e.g. Cooper, 2010), but we do show rather convincingly, that environmental factors have strong impacts on the degree of sexual dimorphism and that molecular and functional phenotypic differences between male and female flies are diminished with increasing levels of environmental stress, providing support for Wallace's ideas about the importance of ecological factors in shaping sexual dimorphism (Wallace 1889). Further studies are needed to fully elucidate the evolutionary implications of these results and we suggest that the impact of environmental stress is considered more in future empirical and theoretical work within this fascinating research field.

ACKNOWLEDGEMENTS

We thank Neda Moghadam, Helle Blendstrup and Susan Hansen for laboratory assistance and Göran Karlsson and Anders Pedersen at the Swedish NMR Center at the University of Gothenburg for help with and use of their spectrometer. We also thank Luis E. Castañeda and two anonymous reviewers for their helpful comments. The Danish Natural Science Research Council supported this research (Sapere aude stipend to TNK; DFF-4002-00036). We declare no conflict of interest regarding the publication of this work.

REFERENCES

Broughton SJ, Piper MDW, Ikeya T *et al.* (2005) Longer lifespan, altered metabolism, and stress resistance in *Drosophila* from ablation of cells making insulin-like ligands. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 3105–3110.

Bylesjö M, Rantalainen M, Cloarec O *et al.* (2006) OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *Journal of Chemometrics*, **20**, 341–351.

Cassidy EJ, Bath E, Chenoweth SF, Bonduriansky R (2014) Sex-specific patterns of morphological diversification: evolution of reaction norms and static allometries in neriid flies. *Evolution*, **68**, 368–383.

Cooper IA (2010) Ecology of sexual dimorphism and clinal variation of coloration in a damselfly. *The American Naturalist*, **176**, 566–572.

Darwin C (1874) *The Descent of Man, and Selection in Relation to Sex*. Crowell, New York, NY.

Dick KB, Ross CR, Yampolsky LY (2011) Genetic variation of dietary restriction and the effects of nutrient-free water and amino acid supplements on lifespan and fecundity of *Drosophila*. *Genetics Research*, **93**, 265–73.

Dunn PO, Whittingham L a, Pitcher TE (2001) Mating systems, sperm competition, and the evolution of sexual dimorphism in birds. *Evolution*, **55**, 161–75.

Fernández-Montraveta C, Moya-Laraño J (2007) Sex-specific plasticity of growth and maturation size in a spider: implications for sexual size dimorphism. *Journal of Evolutionary Biology*, **20**, 1689–1699.

Findlay GD, Yi X, MacCoss MJ, Swanson WJ (2008) Proteomics reveals novel *Drosophila* seminal fluid proteins transferred at mating. *PLOS Biology*, **6**, 1417–1426.

Fischer K, Fiedler K (2001) Dimorphic growth patterns and sex-specific reaction norms in the butterfly *Lycaena hippothoe sumadiensis*. *Journal of Evolutionary Biology*, **14**, 210–218.

Fountain T, Butlin RK, Reinhardt K, Otti O (2015) Outbreeding effects in an inbreeding insect, *Cimex lectularius*. *Ecology and Evolution*, **5**, 409–418.

Hedrick A V., Temeles EJ (1989) The evolution of sexual dimorphism in animals: Hypotheses and tests. *Trends in Ecology and Evolution*, **4**, 136–138.

Hines A, Yeung WH, Craft J *et al.* (2007) Comparison of histological, genetic, metabolomics, and lipid-based methods for sex determination in marine mussels. *Analytical Biochemistry*, **369**, 175–186.

Ketola T, Kristensen TN, Kellermann VM, Loeschcke V (2012) Can evolution of sexual dimorphism be triggered by developmental temperatures? *Journal of Evolutionary Biology*, **25**, 847–855.

Lankadurai BP, Nagato EG, Simpson MJ (2013) Environmental metabolomics: an emerging approach to study organism responses to environmental stressors. *Environmental Reviews*, **21**, 180–205.

Markley JL, Brüschweiler R, Edison AS *et al.* (2017) The future of NMR-based metabolomics. *Current Opinion in Biotechnology*, **43**, 34–40.

Matzkin LM, Watts TD, Markow TA (2009) Evolution of stress resistance in *Drosophila*: interspecific variation in tolerance to desiccation and starvation. *Functional Ecology*, **23**, 521–527.

Nicholson JK, Lindon JC, Holmes E (1999) "Metabonomics": understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica*, **29**, 1181–1189.

Niveditha S, Deepashree S, Ramesh SR, Shivanandappa T (2017) Sex differences in oxidative stress resistance in relation to longevity in *Drosophila melanogaster*. *Journal of Comparative Physiology B*, 1–11. e-pub ahead of print: DOI 10.1007/s00360-017-1061-1.

Partridge L, Gems D, Withers DJ (2005) Sex and death: what is the connection? *Cell*, **120**, 461–472.

R Core Team (2017) R: A language and environment for statistical computing. Version 3.4.0. R Foundation for Statistical Computing, Vienna, Austria. http://www.r-project.org/.

van Ravenzwaay B, Cunha GCP, Leibold E *et al.* (2007) The use of metabolomics for the discovery of new biomarkers of effect. *Toxicology Letters*, **172**, 21–28.

Schou MF, Kristensen TN, Pedersen A *et al.* (2017) Metabolic and functional characterization of effects of developmental temperature in *Drosophila melanogaster*. *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology*, **312**, R211–R222.

Schou MF, Loeschcke V, Kristensen TN (2015) Strong costs and benefits of winter acclimatization in *Drosophila melanogaster*. *PLOS ONE*, **10**, e0130307.

Shen J, Ford D, Landis GN, Tower J (2009) Identifying sexual differentiation genes that affect *Drosophila* life span. *BMC Geriatrics*, **9**, 56.

Stillwell RC, Blanckenhorn WU, Teder T, Davidowitz G, Fox CW (2010) Sex differences in phenotypic plasticity affect variation in sexual size dimorphism in insects: from physiology to evolution. *Annual Review of Entomology*, **55**, 227–245.

Trygg J, Wold S (2002) Orthogonal projections to latent structures (O-PLS). *Journal of Chemometrics*, **16**, 119–128.

Wallace AR (1889) *Darwinism: an Exposition of the Theory of Natural Selection with Some of its Applications*. MacMillan, London, UK.

Walzer A, Schausberger P (2011) Sex-specific developmental plasticity of generalist and specialist predatory mites (Acari: Phytoseiidae) in response to food stress. *Biological Journal of the Linnean Society*, **102**, 650–660.

Zeisel SH (2009) Importance of methyl donors during reproduction. *American Journal of Clinical Nutrition*, **89**, 673S–677S.

Ørsted M, Schou MF, Kristensen TN (2017) Biotic and abiotic factors investigated in two *Drosophila* species – evidence of both negative and positive effects of interactions on performance. *Scientific Reports*, **7**, 40132.

SUPPLEMENTARY INFORMATION FOR PAPER II

Supplementary Table S1. Protocol for preparing experimental media. Ingredients listed for 1 litre of media, notes and references related to the specific treatments. Temperature in the climate rooms was monitored with data loggers (iButton DS1923-F5 with software OneWireViewer x64 version 0.3.15.50, Maxim, Sunnyvale, CA). All treatments were maintained at an average ± SD temperature of 23.01±0.07 °C and 50-70 %RH, except the heat treatment, which was maintained at 32.01±0.27 °C and 50-70 %RH.

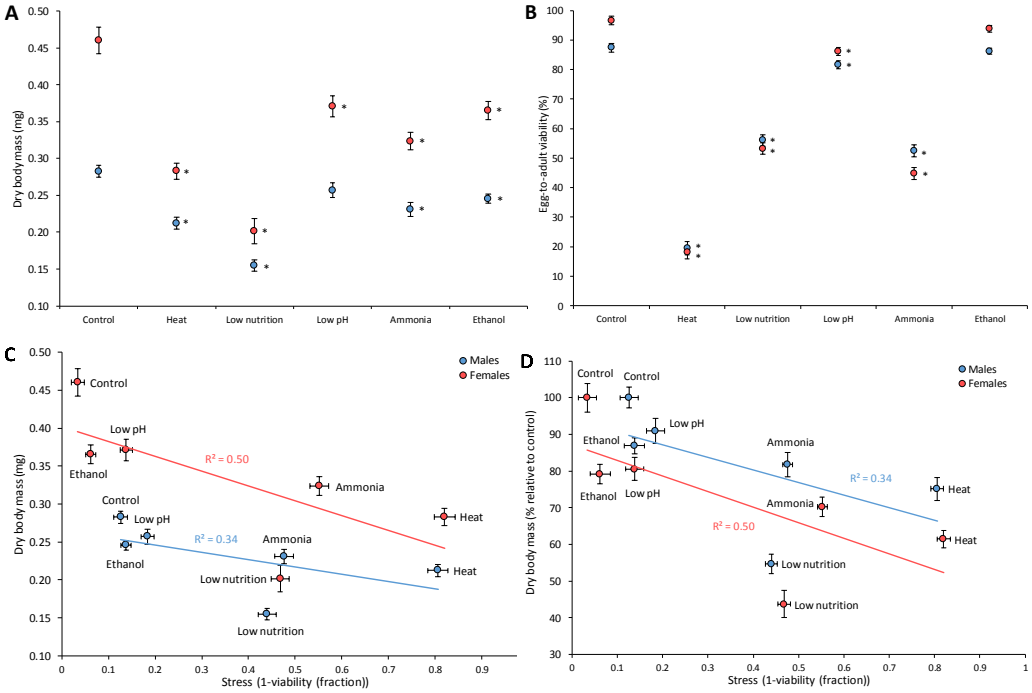
Treatment	Ingredients (per L of media)	Notes	References
Control (standard)	30 g ground oatmeal 40 g sucrose 16 g agar 60 g dry yeast 1 mL acetic acid 12 mL Nipagen	Nipagen is 95 g/L methyl-4-hydroxybenzoate in ethanol (70 vol% in water). Nipagen and acetic acid is added to the media after it has cooled to 70°C to control fungal growth. All standard based media was prepared the day before egg-collection, and stored at room temperature overnight.	-
Heat	Standard	As above except vials are maintained at 32°C.	e.g. Kristensen *et al.* (2015); Ørsted, Schou, & Kristensen, (2017).
Low nutrition	3 g dry yeast 16 g agar	Yeast was dissolved in water prior to adding to boiling water with agar	Schou, Loeschcke, & Kristensen (2015); Urquhart-Cronish & Sokolowski (2014)
Acid	Standard (969 mL) + 31 mL acetic acid	Acetic acid is added slowly and whisked in to ensure a homogenous batch. Nominal concentration of acetic acid: 0.7 M.	Clark & Fucito (1998); Hodge, Campbell-Smith, & Wilson (1996); Hodge (2001)
Ammonia	Standard (900 mL) + 21.4 g ammonium chloride (NH_4Cl)	Ammonium chloride was dissolved in 100 mL water and added to the media after it had cooled to 70°C. Nominal concentration of ammonia: 0.4 M.	Borash *et al.* (2000)
Ethanol	Standard (940 mL) + 60 mL absolute ethanol	Ethanol was whisked carefully into the media after it had cooled to below 50°C. This prevents excessive evaporation, and should result in no more than 10-15 % loss (Yampolsky, Glazko, & Fry, 2012). Furthermore, this media was prepared immediately before use on the day of egg-collection.	Yampolsky *et al.* (2012); Logan-Garbisch *et al.* (2014)

Supplementary Table S2. ^{1}H NMR chemical shifts and integration range used for correlations. Identity of the metabolites was verified using all the listed chemical shifts. Signal intensities used for the correlations presented in Table 2 were calculated as the total intensity within the integration range. Note that some of the signals used to estimate intensities of different metabolites overlap with other signals. *Full name of the galactoside is 1-O-(4-O-(2-aminoethyl phosphate)-ß-D-galactopyranosyl)-glycerol.

Metabolite	chemical shifts (ppm)	Integration range (ppm)
Alanine	1.47	1.476 – 1.462
Asparate	2.80, 2.65	2.799 – 2.787
Betaine	3.88, 3.26	3.260 – 3.252
Choline	4.05, 3.19	3.196 – 3.189
Fatty acid	5.31, 1.28, 0.89	0.883 – 0.823
Glucose	5.22, 4.64, 3.83	4.635 – 4.625
Glutamate	2.34, 2.12, 2.05	2.359 – 2.335
Glutamine	2.44, 2.13	2.454 – 2.416
Galactoside*	4.46, 4.18, 3.93, 3.76, 3.61	4.466 – 4.451
Isoleucine	1.45, 1.00, 0.93	1.009 – 0.993
Leucine	1.69, 0.95	0.965 – 0.946
Maltose	5.40, 4.65	5.411 – 5.397
Methilnine sulphoxide	3.88, 3.01, 2.74, 1.67	2.744 – 2.739
Methionine	3.86, 2.63, 2.15	2.640 – 2.622
Phenylalanine	7.42, 7.37, 7.32	7.422 – 7.404
Phosphocholine	4.17, 3.21	3.215 – 3.208
Proline	2.12, 2.01	2.039 – 1.995
Threonine	4.25, 3.58, 1.32	1.326 – 1.323
Tyrosine	7.18, 6.89	7.194 – 7.174
Uridine	7.87, 5.90, 4.34	5.912 – 5.882
Valine	2.27, 1.03, 0.98	1.042 – 1.023

Supplementary Table S3. OPLS model statistics for parameter prediction from metabolite data. Capability of the *D. melanogaster* metabolome to predict different parameters was tested using orthogonal projections to latent structures (OPLS) models. The test was done both with continuous parameters, as well as categorical parameters. A† describes the number of model components where the first number accounts for the predictive component(s) correlating with the predicted variable and the second the orthogonal component(s). N‡ describes the number of observations included in the model. R^2§ describes how much of the total metabolite variation that is explained by the model. Q^2* represents the predictability of the total model and is related to the statistical validity of the model. $Q^2 > 0.5$ is considered significant and is bold in the table. Q^2 was calculated using cross-validation with all measurements for one condition left out at a time. §§ The Q^2 of this model refers to its ability to separate different stress treatments. The model is still useful for characterization of the linear stress trajectory.

Parameter	Metabolome	Type	A†	N‡	R^2§	Q^2*
Female control vs. male control	Control flies	OPLS-DA	1+0	10	0.57	0.88
Female control vs. stressed females	Female flies	OPLS-DA	1+0	30	0.26	0.85
Linear stress trajectory for sex difference	Sex difference	O2PLS-DA	3+0+0	60	0.59	0.48§§
Egg-to-adult viability	Sex difference	OPLS	1+2	60	0.52	0.83
Sex difference in dry weight	Sex difference	OPLS	1+2	60	0.54	0.76



Supplementary Fig. S1. (A) Effects of the six experimental treatments on dry body mass. In all panels, red circles are females and blue circles are males. Error bars represent SE (n=17-20 flies), and asterisks denote significant difference (as determined by one-way ANOVA and Welch's t-tests). (B) Effects of the six experimental treatments on egg-to-adult viability. Error bars represent SE (n=25 vials), and asterisks denote significant difference (as determined by GLM). (C) Correlations between dry body mass and severity of the stressful treatment measured as 1 - viability. (D) Correlations between dry body mass measured as % relative to the respective control for each sex and severity of the stressful treatment measured as 1 - viability. Error bars represent SE (n=17-20 for body mass difference and n=25 for stress severity (viability)). The lines represent linear regressions; red is for females and blue is for males, and R^2 values are presented in the same sex specific colours.

SUPPLEMENTARY REFERENCES

Borash DJ, Pierce VA, Gibbs AG, Mueller LD (2000) Evolution of ammonia and urea tolerance in *Drosophila melanogaster*: resistance and cross-tolerance. *Journal of Insect Physiology*, **46**, 763–769.

Clark AG, Fucito CD (1998) Stress tolerance and metabolic response to stress in *Drosophila melanogaster*. *Heredity*, **81**, 514–527.

Hodge S (2001) The effect of pH and water content of natural resources on the development of *Drosophila melanogaster* larvae. *Drosophila Information Service*, **84**, 38–43.

Hodge S, Campbell-Smith R, Wilson N (1996) The effects of resource acidity on the development of *Drosophila* larvae. *The Entomologist*, **115**, 129–139.

Kristensen TN, Overgaard J, Lassen J, Hoffmann AA, Sgrò C (2015) Low evolutionary potential for egg-to-adult viability in *Drosophila melanogaster* at high temperatures. *Evolution*, **69**, 803–814.

Logan-Garbisch T, Bortolazzo A, Luu P *et al.* (2014) Developmental ethanol exposure leads to dysregulation of lipid metabolism and oxidative stress in *Drosophila*. *G3*, **5**, 49–59.

Schou MF, Loeschcke V, Kristensen TN (2015) Inbreeding depression across a nutritional stress continuum. *Heredity*, **115**, 1–7.

Urquhart-Cronish M, Sokolowski MB (2014) Gene-environment interplay in *Drosophila melanogaster*: chronic nutritional deprivation in larval life affects adult fecal output. *Journal of Insect Physiology*, **69**, 95–100.

Yampolsky LY, Glazko G V, Fry JD (2012) Evolution of gene expression and expression plasticity in long-term experimental populations of *Drosophila melanogaster* maintained under constant and variable ethanol stress. *Molecular ecology*, **21**, 4287–99.

Ørsted M, Schou MF, Kristensen TN (2017) Biotic and abiotic factors investigated in two *Drosophila* species – evidence of both negative and positive effects of interactions on performance. *Scientific Reports*, **7**, 40132.

PAPER III

ENVIRONMENTAL VARIATION PARTITIONED INTO SEPARATE HERITABLE COMPONENTS

In review after first revision in *Evolution*

MICHAEL ØRSTED, PALLE DUUN ROHDE, ARY ANTHONY HOFFMANN, PETER SØRENSEN & TORSTEN NYGAARD KRISTENSEN

Environmental variation partitioned into separate heritable components

Michael Ørsted[1,2], Palle Duun Rohde[3,4,5], Ary Anthony Hoffmann[1,2], Peter Sørensen[3], and Torsten Nygaard Kristensen[1,6]

[1] *Department of Chemistry and Bioscience, Aalborg University, Aalborg E, Denmark*

[2] *Bio21 Institute, University of Melbourne, Parkville, Victoria, Australia*

[3] *Center for Quantitative Genetics and Genomics, Aarhus University, Tjele, Denmark*

[4] *iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research, Aarhus C, Denmark*

[5] *iSEQ, Center for Integrative Sequencing, Aarhus University, Aarhus C, Denmark*

[6] *Department of Bioscience, Aarhus University, Aarhus C, Denmark*

ABSTRACT

Trait variation is normally separated into genetic and environmental components, yet genetic factors also control the expression of environmental variation, encompassing plasticity across environmental gradients and within-environment responses. We defined four components of environmental variation: plasticity across environments, variability in plasticity, variation within environments, and differences in within-environment variation across environments. We assessed these components for cold tolerance across five rearing temperatures using the Drosophila melanogaster Genetic Reference Panel (DGRP). The four components were found to be heritable, and genetically correlated to different extents. By whole genome single marker regression, we detected multiple candidate genes controlling the four components and showed limited overlap in genes affecting them. Using the binary UAS-GAL4 system, we functionally validated the effects of a subset of candidate genes affecting each of the four components of environmental variation and also confirmed the genetic and phenotypic correlations obtained from the DGRP in distinct genetic backgrounds. We delineate selection targets associated with environmental variation and the constraints acting upon them, providing a framework for evolutionary and applied studies on environmental sensitivity. Based on our results we suggest that the traditional quantitative genetic view of environmental variation and genotype-by-environment interactions needs revisiting.

Keywords: Environmental variation, plasticity, genetic control, cold tolerance, DGRP

INTRODUCTION

Phenotypic trait variation can be partitioned into genetic (V_G) and environmental (V_E) variation. However, the expression of environmental variation is under heritable genetic control (Ros *et al.* 2004; Ibáñez-Escriche *et al.* 2008; Ayroles *et al.* 2015; Morgante *et al.* 2015; Sørensen *et al.* 2015; Blasco *et al.* 2017). While environmental variation has been studied for many decades, the underlying genetic mechanisms controlling V_E are poorly understood, particularly as different forms of V_E are recognized (Hill & Mulder 2010). The different forms of V_E may be genetically independent, thereby having distinct impacts on evolutionary outcomes (Hill & Mulder 2010). Knowledge of the genetic basis of V_E is important for understanding evolutionary trajectories in variable and potentially stressful environments (Zhang & Hill 2005; Lande 2014), when undertaking breeding programs aiming at generating homogenous phenotypes across different production conditions (Mulder *et al.* 2008; Hill & Mulder 2010), and for identifying the impact of environmental conditions on genotype specific responses to diseases and treatments in humans (Hunter 2005; Vasseur & Quintana-Murci 2013). Furthermore, the existence of several forms of V_E may challenge the traditional view of genotype-by-environment (GxE) interactions.

The ability of genotypes to produce a multitude of phenotypes when exposed to different environmental conditions is termed phenotypic plasticity. Environmental variation resulting from plasticity (here termed V_{plast}) is perhaps the most studied type of V_E (DeWitt & Scheiner 2004; Valladares *et al.* 2006), and is usually assessed by measuring individuals of a given genotype under different environmental conditions (Table 1 and Fig. 1A). Trait plasticity is often regarded as being separate from trait means, thus, being controlled by a different set of genes capable of responding to selection (Stearns 1989; Scheiner & Lyman 1991). It is considered particularly important from an adaptive perspective because V_{plast} may allow populations to rapidly utilise novel ecological opportunities in temporally and spatially heterogeneous environments or respond to stressful conditions, including those emerging as a consequence of anthropogenic climate change (Teplitsky *et al.* 2008; Hoffmann & Sgrò 2011; Anderson *et al.* 2012).

A second source of environmental variation is the within-environment variation (here termed V_{WE}), which is a measure of the extent to which a genotype can produce the same phenotype within an environment (Table 1 and Fig. 1B). Like plasticity, within-environment variation is typically considered to be under genetic control (Ros *et al.* 2004; Ibáñez-Escriche *et al.* 2008; Ayroles *et al.* 2015; Morgante *et al.* 2015; Sørensen *et al.* 2015), and have recently been shown to respond to selection, without affecting trait mean (Blasco *et al.* 2017). This form of variation is also an important component of adaptive responses. For instance, in a fluctuating environment, stabilising selection will favour individuals, which produce high levels of variability in their offspring to maximise fitness (Zhang & Hill 2005; Devaux & Lande 2010).

This is often referred to as a diversified bet hedging strategy, and represent a genetically controlled form of variation expressed within an environment (Simons & Johnston 1997; Marshall *et al.* 2008). Improving our knowledge of the underlying genetic basis of V_{WE} and its relationship to other components of environmental variation helps to understand and predict species responses in a more variable future climate (Chown *et al.* 2010). In applied breeding programs, where a homogenous phenotype is often desirable, in-depth knowledge of the genetic basis of V_{WE} may have important economic value (Mulder *et al.* 2008; Hill & Mulder 2010).

Two other forms of environmental variation impacting phenotypic variation can be recognized, but are not commonly characterised. The first involves developmental stability across environments; that is variation of within-environment variation across environments (here termed V_{AE}). It measures the extent to which the expression of phenotypic variation within an environment is constant across the range of environments considered (Table 1 and Fig. 1B). The second involves variation of the plastic response (here termed $V_{v.plast}$), i.e., a measure of the extent to which the same genotype can consistently produce the same plastic response across environments (Table 1 and Fig. 1A).

All four components of environmental variation are likely to be under genetic control, although for the latter two (V_{AE} and $V_{v.plast}$) this has not been documented. Control of within-environment variation (V_{WE}) is thought to be linked to heat shock proteins and several other classes of environmental stress responsive genes (Bergman & Siegal 2003; Sangster *et al.* 2008). The control of plastic responses is thought to reside with processes like acclimation and hardening (Schlichting & Pigliucci 1993; Beaman *et al.* 2016) and the genetic architecture of these processes have been investigated especially in *Drosophila* (Sørensen *et al.* 2005). However, it is currently unknown whether the different components of V_E have the same underlying genetic basis, given that quantitative genetic and genomic analyses of the different components of V_E have rarely been performed.

Here we investigated the genetic control of the four components of environmental variation for cold tolerance in *Drosophila melanogaster*. Cold tolerance is an important trait for understanding evolutionary processes in natural populations because the ability to withstand low temperatures limits the geographic distribution of many species (Sunday *et al.* 2011; Williams *et al.* 2015) and cold tolerance in particular is a good predictor of present and future geographical distributions (Kimura 2004; Overgaard *et al.* 2011; Araújo *et al.* 2013). Cold tolerance is strongly impacted by environmental variation; in ectotherms the trait is highly plastic (Schou *et al.* 2017b), and plays a key role in determining adaptation to thermal extremes (Hoffmann *et al.* 2003; Chown *et al.* 2010; Overgaard *et al.* 2011).

We conducted a comprehensive investigation of the genetic basis for each of the four components of environmental variation impacting cold tolerance, as well as the correlation among these, using 166 inbred lines from the *Drosophila melanogaster* Genetic Reference Panel (DGRP) (Mackay *et al.* 2012). Individuals from each DGRP line were reared at five thermal environmental condition; 17, 20, 23, 26, and 29 °C, for one generation and the cold tolerance of adult males was then assessed using the dynamic measure critical thermal minimum (CT_{min}; the temperature of chill coma onset under gradual cooling (Overgaard *et al.* 2012)). The four components of V_E were determined (Table 1) for each DGRP line, and to investigate the underlying genetic basis, we performed whole genome single marker regression for each component of V_E and validated our results by gene expression knockdown using the *UAS*-GAL4 system. This allowed four questions about the genetic control of different components of V_E affecting cold tolerance within and between environments to be answered: 1) Are the different components of V_E heritable, i.e. have adaptive potential? 2) Is genetic variation of the different components of V_E correlated? 3) Are the components of V_E related to inherent (non-plastic) cold tolerance? and 4) What is the genetic architecture controlling the different components of V_E?

Table 1. Overview of the four components of environmental variation. Overview of four components of environmental variation as applied here to cold tolerance (CT_{min}); plasticity (V_{plast}), within-environment variation (V_{WE}), across environment variation (V_{AE}), and variation of the plastic response ($V_{v.plast}$) and how they are computed. Sample studies or reviews are included.

Component of environmental variation	Computation	Definition	References
Plasticity, V_{plast}	Regression coefficient (i.e. slope) of a linear regression of CT_{min} data as a function of rearing temperature.	The ability of a genotype to produce a multitude of phenotypes when exposed to different environmental conditions. Norm-of-reactions are widely used in the study of phenotypic plasticity.	Different indices of phenotypic plasticity are reviewed in e.g. Valladares *et al.* (2006) and DeWitt and Scheiner (2004).
Within-environment variation, V_{WE}	Coefficient of variation (CV = standard deviation/mean) within each environment. Line mean CV ($\overline{CV}$) was also computed.	How consistently a genotype produces the same phenotype within a given environment. Similar measures include micro-environmental variation.	Ayroles *et al.* (2015); Morgante *et al.* (2015); Sørensen *et al.* (2015).
Across environment variation, V_{AE}	Highest CV in any given environment minus lowest CV in any other given environment (ΔCV).	Developmental stability across environments; the consistency of within-environment variation across environments.	Tuljapurkar and Istock (1993).
Variation in plastic response, $V_{v.plast}$	Standard error of the slope.	A measure of the extent to which the same genotype can produce the same plastic response consistently.	Present study.

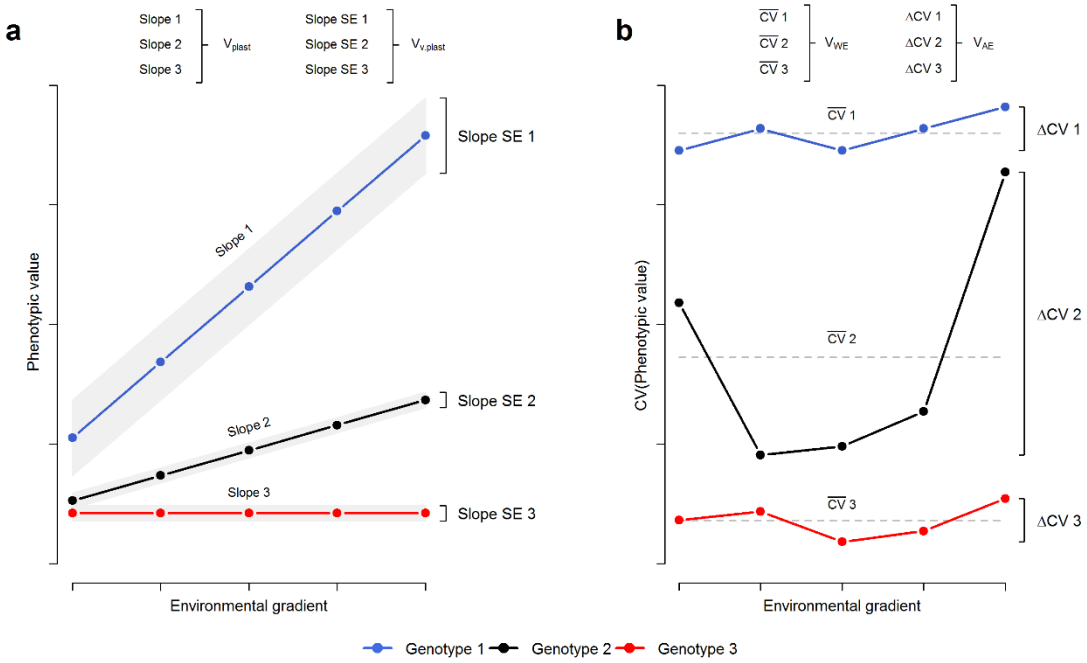


Fig. 1 Illustration of the four components of environmental variation. (a) Average values of any phenotype measured across an environmental gradient in individuals from three hypothetical genotypes differing in plasticity (V_{plast}), i.e. slope of the reaction norm, and thus in the contribution of plasticity to the overall environmental variation of the genotypes. Genotypes 1 and 2 have different phenotypic plasticity with genotype 1 being the most plastic. Genotype 3 is canalized and has no environmental variation as a result of plasticity. The shading represents the variation of the plastic response ($V_{v.plast}$), i.e. differences in the standard error of the slope. Here genotype 1 has the highest plasticity and simultaneously the highest variation in the plastic response, whereas genotypes 2 and 3 have the same slope SE despite having different slopes. (b) Environmental variance (measured as the coefficient of variation, CV) of three hypothetical genotypes (separate from panel (a)) across an environmental gradient. Points represent the within-environment variation (V_{WE}) measured as within-environment CV. Dashed grey lines represent a measure of the mean within-environment variation ($\overline{CV}$) across environments. On the right-hand side the difference in variation across environments (V_{AE}) measured as ΔCV is shown. The genotypes differ in both variation within-environments and across environments. Genotype 1 has a consistently high CV in all environments, i.e. the phenotypic variation within each environment is large, and consistently so across environments (low ΔCV). Genotype 2 has low CVs in some environments but high in others, i.e. an intermediate $\overline{CV}$, but a high variation in the expression of phenotypic variation across environments (high ΔCV). Genotype 3 has low CVs in all environments, i.e. low developmental instability both within and across environments.

MATERIALS AND METHODS

Fly stocks and maintenance

The DGRP consists of a collection of lines that through full-sib mating has been inbred to an expected inbreeding coefficient of F∼1 (Mackay *et al.* 2012), resulting in the large majority of genomic sites being homozygous within line. The study included 166 lines from the *Drosophila melanogaster* Genetic Reference Panel (DGRP) (Mackay *et al.* 2012) obtained from Bloomington *Drosophila* Stock Center (NIH P40OD018537), 40 *UAS*-RNAi lines and the two corresponding genetic background strains from Vienna *Drosophila* Resource Center (Dietzl *et al.* 2007), and the GAL4 line act5-*GAL4*/CyO, which was donated to us from colleagues at Copenhagen University, Denmark. All stocks were maintained at 23 °C, 50 %RH and a 12:12 h photoperiod on a standard oatmeal-sugar-agar-yeast medium for two generations in our laboratory prior to initiating the study.

Thermal rearing conditions

To assess the effect of thermal rearing condition on cold tolerance, the 166 DGRP lines were reared at five thermal conditions: 17, 20, 23, 26, and 29 °C. For each line, three replicates of approximately 20 adult flies (age 3-4 days), reproduced for 12 h at 23 °C in vials containing 7 mL medium after which the flies were discarded and the vials were transferred to incubators (KBWF 720 E5.3, Binder, Tuttlingen, Germany) at the respective rearing temperature. The temperature within the incubators was continuously monitored using data loggers (iButton DS1923-F5 with software OneWireViewer x64 version 0.3.15.50, both from Maxim, Sunnyvale, CA, USA). Incubators maintained an average temperature (±SD) of 17±0.14 °C, 20±0.35 °C, 23±0.15 °C, 26±0.37 °C, and 29±0.39 °C.

Assessment of critical thermal minimum, CT_{min}

Individuals from each of the five rearing conditions were assessed for their ability to tolerate low temperatures determined using a dynamic measure of critical thermal minimum (CT_{min}), which is a standardised procedure of gradual cooling (Overgaard *et al.* 2012). CT_{min} is defined as the temperature at which a complete cessation of movement is observed as a result of the flies entering chill coma.

Within one day after eclosion, a maximum of 20 adult males were collected under CO_2 anaesthesia, transferred to a new vial with 3 mL medium, and returned to the respective temperature regime for 48 h. Assessment of CT_{min} was performed on approximately 10 male flies (age 60±12 h) per line per rearing temperature (for exact numbers see Supplementary Table S1). Flies were transferred to individual screw-cap glass vials (45 x 15 mm) and randomly placed in a metal rack, which was submerged into a water bath with ethylene glycol and water (1:3 vol.) pre-set to 23 °C. The temperature of the water bath was decreased at a rate of 0.1 °C/min, and the

temperature at which all movement of a fly ceased (gently tapping the vials with a metal rod and shining a flashlight) due to chill coma was recorded. In all experiments, 12,612 flies were scored for CT_{min}. Because the rearing temperature affects developmental time and due to the scale of the experiment, the assays were performed on multiple days, at different time of the day and in two different water baths. To minimize potential bias due to these experimental factors, all assays were performed with the exact same setup, in the same location, and throughout the experiment all flies were scored for CT_{min} by the same person. The temperature in the water bath was continuously monitored during the assays with iButton data loggers. In the temperature range comparable among all assays (23.0 to 8.5 °C), the average cooling rates (±SD) of temperature change in the two baths were -0.0970±0.0016 °C/min and -0.0971±0.0017 °C/min, respectively.

Measures of environmental variation

Four types of environmental variation were computed (Table 1, Fig. 1). In order to retain a replicated structure of the data, the CT_{min} from each DGRP line and rearing temperature were grouped into three groups of individuals according to day of assay, time of day, and water bath, in this order. This way we also accounted for some of the variation due to experimental factors potentially impacting V_E components, especially V_{WE} (Supplementary Fig. S1). Based on the three replicate groups, the four components of environmental variation were computed (i.e. obtaining three replicates per measure per line) using log-transformed CT_{min} to ensure normality of the data. To test the robustness of the V_E estimates, we performed 10,000 random samplings of three replicate groups of the same size as the experimental groups, with replacement of observations in each line and in each environment. The four V_E components were calculated based on each of these samplings. We found a high degree of concordance between the estimates based on the three experimental groups and estimates based on randomly sampled groups across all the DGRP lines (see Supplementary information for details).

Phenotypic plasticity (V_{plast}) was estimated as the regression coefficient of a linear regression (i.e., slope) of CT_{min} as a function of rearing temperature (Table 1, Fig. 1A). We chose a linear fit because previous studies have shown that the relationship between CT_{min} and developmental temperature is linear (Schou *et al.* 2017a), and because linear norm-of-reaction analysis is the most commonly used in studies of phenotypic plasticity (Valladares *et al.* 2006). The variation of the plastic response ($V_{v.plast}$) was computed as the standard error of the regression coefficient (Table 1, Fig. 1A), as this defined the uncertainty of determination of the plastic response.

Line specific variation, both within (V_{WE}) and across (V_{AE}) environments, was estimated using the coefficient of variation (CV = SD/mean) (Table 1, Fig. 1B). To enable the comparison of a single measure of within-environment variation with the other three components of environmental variation, we computed an average V_{WE} as

the mean of the individual CVs from each of the five rearing conditions ($\overline{CV}$, Table 1, Fig. 1B). As a measure of developmental stability, i.e. environmental variation across environments (V_{AE}), we computed the difference between the highest CV and the lowest CV across environments (ΔCV, Table 1, Fig. 1B).

DGRP genotypes

Only segregating biallelic SNPs with a minor allele frequency ≥ 0.05, a Phred quality score ≥ 500, and a genotype call ≥ 0.8 were included. This resulted in a total of 1,725,755 SNPs distributed on the six chromosome arms (2L, 2R, 3L, 3R, 4 and X). Sequence data including information on major inversions and *Wolbachia* infection status were downloaded from the DGRP2 facility (http://dgrp2.gnets.ncsu.edu). SNPs were annotated to genes using FlyBase annotation v5.49 (flybase.org). In addition, genes were linked to gene ontology (GO) categories (The Gene Ontology Consortium *et al.* 2000). Only GO terms with at least 10 directly evidenced genes were included. In total, SNPs were annotated to 10,517 genes and 1,117 GO terms.

Quantitative genetic parameters

The proportion of phenotypic variance explained by genetic variance (H^2) and additive genetic variance captured by SNPs (h^2) were estimated using the average information restricted maximum-likelihood (REML) procedure (Madsen *et al.* 1994; Johnson & Thompson 1995). H^2 was estimated as:

$$H^2=\frac{\sigma^2_{g_L}}{\sigma^2_{g_L}+\sigma^2_e}$$

where $\sigma^2_{g_L}$ and σ^2_e were obtained by fitting $\mathbf{y} = \mathbf{Xb} + \mathbf{Zg_L} + \mathbf{e}$, where $\mathbf{y}$ was a vector of phenotypic observations (CT_{min} at 23 °C and slope SE (log transformed), slope (untransformed), ΔCV and $\overline{CV}$ (both square root transformed)), $\mathbf{b}$ was a vector of fixed effects (day, bath, time on day, and *Wolbachia* infection status for CT_{min}, and for environmental variation phenotypes only *Wolbachia* infection status was included), $\mathbf{g}_L$ was a vector of random line effects assuming the DGRP lines to be independent, i.e. $\mathbf{g}_L \sim N(0, \mathbf{I}_L\sigma^2_{g_L})$, where I_L was an identity matrix, and $\mathbf{e}$ was a random residual, $\mathbf{e} \sim N(0, \mathbf{I}\sigma^2_e)$. The estimate of h^2 was obtained as:

$$h^2=\frac{\sigma^2_g}{\sigma^2_g+\sigma^2_e}$$

where σ^2_g and σ^2_e were obtained by fitting $\mathbf{y} = \mathbf{Xb} + \mathbf{Zg} + \mathbf{e}$, where $\mathbf{y}$ was a vector of phenotypic observations (as defined above), $\mathbf{b}$ was a vector of fixed effects (as defined above, including five major polymorphic inversions, *I2Lt*, *I2Rns*, *I3Rp*, *I3Rk* and *I3RMo*), $\mathbf{g}$ was a vector of random genetic effects, $\mathbf{g} \sim N(0, \mathbf{G}\sigma^2_g)$, where $\mathbf{G}$ was the genomic relationship matrix, and $\mathbf{e}$ was a random residual, $\mathbf{e} \sim N(0, \mathbf{I}\sigma^2_e)$. The $\mathbf{G}$ matrix was computed as $\mathbf{G}=\mathbf{WW}'/m$ (VanRaden 2008), where m was the number of

genetic markers, and **W** was a centred and scaled genotype matrix. Each column vector of **W**, $\mathbf{w}_i = \frac{\mathbf{a}_i - 2p_i}{\sqrt{2p_i(1-p_i)}}$, p_i was the frequency of the minor allele at locus *i*, and $\mathbf{a}_i$ was the *i*th column vector of the allele count matrix, **A**, containing the genotypes encoded as 0, 1, 2, counting the number of the minor alleles. Standard errors of the heritability estimates were obtained using the inverse of Fisher's information measure multiplied by minus one (Lynch & Walsh 1998; Sorensen & Gianola 2002).

Phenotypic (ρ_p) and genetic correlations (ρ_g) were computed among inherent cold tolerance (defined as CT_{min} at 23 ℃), V_{plast}, $V_{v.plast}$, V_{AE}, V_{WE}. The genetic correlation was approximated as the correlation between SNP effects from the whole genome single marker regression (see below). In both cases, the correlations were obtained as $\rho = \mathrm{cov}(x, y) \Big/ \sqrt{\sigma_x^2 \sigma_y^2}$, with standard errors computed as $SE_\rho = \sqrt{\frac{1-\rho^2}{n-2}}$, where *n* was the sample size.

Whole genome single marker regression

For each component of V_E and inherent cold tolerance (CT_{min} at 23 ℃) we performed whole genome single marker regression to test individual SNPs for association. In order to account for the fixed effects (i.e. experimental conditions, chromosomal inversions and *Wolbachia* infection) and polygenic inheritance among the DGRP lines, the response variables were the estimated genetic values (i.e., $\hat{\mathbf{g}}$) obtained from linear mixed models described in 'Quantitative genetic parameters'. Thus, the degree of association between the i^{th} SNP and the component of V_E was determined using a *t*-test of the regression of the genetic values on the i^{th} SNP. Significance level was set to p-value < 1×10^{-5}. This arbitrary significance level was based on a range of other DGRP studies (Durham *et al.* 2014; Montgomery *et al.* 2014; Vaisnav *et al.* 2014; Gaertner *et al.* 2015).

Functional categories, here GO categories, were tested for enrichment of associated SNPs, i.e., SNPs with *p*-values $< 1 \times 10^{-5}$. We used a count based set test approach (previously described by (Rohde *et al.* 2016; Sørensen *et al.* 2017), which counts the number of genetic markers in a particular GO term that are associated with the trait. The enrichment score was computed as:

$$T_{\text{count}} = \sum_{i=1}^{m_f} \mathbb{I}(t_i > t_0)$$

where m_f is the number of markers within the GO term, t_i is the i-th single marker p-value, t_0 is the significance threshold (here $t_0 = 1 \times 10^{-5}$), and $\mathbb{I}$ is an indicator function that takes the value one if the argument is satisfied. Under the competitive null hypothesis, i.e., individually associated SNPs are distributed randomly across the genome, the significance can be determined by obtaining an empirical distribution of T_{count}. Here, the empirical distribution was obtained using the circular permutation

approach as described by Cabrera et al. (2012), where the genome was considered to be circular. Then, the set of SNP p-values were permuted by rotating with respect to their genomic position, i.e., a random number between one and the total number of SNPs was drawn, and the observed SNP p-value of the first SNP in the genome rotates to that of the random number-th SNP, and all other SNP p-values rotate to the same degree to the corresponding SNPs. Thus, the SNPs retain the original order but, at each permutation, gain new random p-values. This uncouples the association between SNP and GO term, while retaining the correlation pattern (due to linkage disequilibrium) among p-values. For each permutation (here 10,000) a new T_{count} statistic was computed based on the original position of the GO term, and a empirical p-value was obtained as a one-tailed test of the proportion of permuted test statistics that were larger than the observed.

A permutation approach was used to test if the number of genes and GO terms in common between traits was larger than expected. Initially, an incidence matrix with *n* rows (number of genes or GO terms) and *m* columns (the number of traits, i.e., five) was constructed: genes or GO terms associated with a trait was set to one, otherwise to zero. The true overlap between traits was compared to an empirical distribution obtained by permuting the elements within columns and computing the overlap (x10,000). The probability of the observed overlap was estimated under the null hypothesis of independent association among traits. An empirical *p*-value was obtained as the fraction of all permutations with an overlap equal/higher to the observed overlap among traits.

Across environment predictions

In order to support the estimated genetic correlations, we used the results from the whole genome single marker regression to predict the observed values of inherent cold tolerance and values of the four V_E components. This was achieved by fitting series of genomic best linear unbiased prediction (GBLUP) models (Meuwissen *et al.* 2001) and genomic feature best linear unbiased prediction (GFBLUP) models (Sarup *et al.* 2016; Edwards *et al.* 2016). The GBLUP models serves as a NULL model fitting all SNPs within one random genetic component assuming all SNP effects to be drawn from the same distribution:

$$\tilde{\mathbf{y}} = \boldsymbol{\mu} + \mathbf{Zg} + \mathbf{e}$$

where $\boldsymbol{\mu}$ is a vector of the overall mean, $\mathbf{Z}$ are design matrices linking adjusted phenotypic values to genetic values ($\mathbf{g}$) captured by all SNPs; $\mathbf{g} \sim N(0, \mathbf{G}\sigma_g^2)$. $\tilde{\mathbf{y}}$ is a vector of adjusted phenotypic values, i.e. $\tilde{\mathbf{y}} = \mathbf{g}+\mathbf{e}$, where the genetic and residual effects were obtained from the linear mixed model from the 'Quantitative genetic parameters'. The GFBLUP model includes an additional random genetic effect, that allows one to assess the importance (in terms of e.g., variance explained or predictive ability) of a selected set of SNPs:

$$\tilde{\mathbf{y}} = \boldsymbol{\mu} + \mathbf{Zf} + \mathbf{Zr} + \mathbf{e}$$

with same specifications, as in the GBLUP model, except **f** is a vector of line-specific genetic effects for the SNPs within the genomic feature, and **r** is a vector of line-specific genetic values for the SNPs not within the genomic feature (i.e. the remaining SNPs). The random genetic and residual effects were assumed to be independent normally distributed: $\mathbf{f} \sim N(0, \mathbf{G}_f\sigma_f^2)$, $\mathbf{r} \sim N(0, \mathbf{G}_r\sigma_r^2)$, $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$, where $\mathbf{G}_f$ and $\mathbf{G}_r$ were computed as described in the paragraph above for the subset of SNPs. Here, the feature group contained SNPs that in the whole genome single marker regression with a *p*-value bin, i.e., $p <$ 1E-5, 1E-4, 0.001. 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 and 0.9. Specifically, the feature groups were based on the *p*-values obtained from one trait (inherent cold tolerance or a component of V_E), and the predictions were performed for another trait. A total of 20 cross-trait predictions were performed. For each genomic feature (i.e., *p*-value bin) a GFBLUP model was fitted, and the performance was measured as the predictive ability, which is the correlation between observed and predicted genomic effects (see below), compared to the predictive ability of the GBLUP model using 10-fold cross validation, i.e., 9:10 as training set (t), and 1:10 as a validation set (v) with 100 training and validation sets. The predicted genetic effects for GFBLUP of lines in the validation set were computed as:

$$\hat{\mathbf{g}}_v = \left(\mathbf{G}_{f_{v,t}}\hat{\sigma}_f^2 + \mathbf{G}_{r_{v,t}}\hat{\sigma}_r^2\right)\left[\mathbf{G}_{f_{t,t}}\hat{\sigma}_f^2 + \mathbf{G}_{r_{t,t}}\hat{\sigma}_r^2 + \mathbf{I}_{t,t}\hat{\sigma}_e^2\right]^{-1}\left(\tilde{\mathbf{y}}_t - \hat{\mu}_t\right)$$

Similarly, for the GBLUP model, the genetic effects were computed as:

$$\hat{\mathbf{g}}_v = \left(\mathbf{G}_{v,t}\hat{\sigma}_g^2\right)\left[\mathbf{G}_{t,t}\hat{\sigma}_g^2 + \mathbf{I}_{t,t}\hat{\sigma}_e^2\right]^{-1}\left(\tilde{\mathbf{y}}_t - \hat{\mu}_t\right)$$

Validation of candidate genes

For each V_E measure we selected top 10-11 associated genes to be used in gene expression knockdown using the binary *UAS*-GAL4 system. For each rearing temperature, virgin females from the *UAS*-RNAi lines where crossed to actin-GAL4 (*y*^{1} *w***; P{Act5C-GAL4}25FO1/CyO,*) males to ubiquitous knockdown gene expression of individual candidate genes in the F_1 adult male offspring. CT_{min} of ca. 20 flies from each *UAS*-GAL4 cross and rearing condition were assessed. Each measure of V_E was then estimated based on seven replicates, each consisting of ca. 3 flies. The observed phenotype (i.e., the four V_E components) of the *UAS*-GAL4 F_1 males was compared to a corresponding control line with the same genetic background crossed to the GAL4 line, and significance was assessed using a one-tailed Mann-Whitney U-test. The *p*-values were corrected for multiple testing using Bonferroni correction. The standard error of the difference of means were determined as $\sigma_{M_1-M_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$, where M_1 and M_2 are means, σ_1^2 and σ_2^2 are the variance among replicate values, and n_1 and n_2 are number of replicates of the *UAS*-GAL4 line and its control, respectively.

Assessment of inbreeding effects using diallel crosses

To test whether differences in CT_{min} between DGRP lines was due to line-specific inbreeding effects, we evaluated potential inbreeding depression by comparing CT_{min} of parental lines and their hybrids. In addition, the contribution of dominance and overdominance affecting the inheritance of CT_{min} was evaluated. A set of half diallel crosses of selected DGRP lines was used (Supplementary Table S3). Five DGRP lines with high average CT_{min} and four lines with low average CT_{min} (averaged across the five rearing temperatures), all free of *Wolbachia*, were used in the crosses. Less than 8 h after eclosion, five to ten virgin males and females were collected under CO_2 anaesthesia and put into separate vials, and left for 48 h to recover before the crosses were established. Hereafter, they were transferred to a new vial for egg-laying for four time periods of 12 h at 23 °C using the same setup as in the main experiment. All possible crosses between the selected DGRP lines were performed, resulting in a total of 36 hybrid crosses (approximately equal amount of crosses were done with males and females from each of the high and low CT_{min} lines). At an age of 60±12 h 8 males from the 36 hybrid crosses and 9 parental lines were tested using the same setup as in the main experiment. Measures of additive and non-additive components were obtained by computing the general combining ability (GCA) and specific combining ability (SCA). The GCA/SCA ratio evaluates the contribution of additive vs. non-additive gene action responsible for the inheritance of the trait. GCA/SCA < 1 indicates larger effects of the non-additive gene action, whereas, GCA/SCA > 1 indicates greater importance of additive gene action (Griffing 1956).

Statistical software

All analyses were done within the R programming environment (R Core Team 2017). All quantitative genetic and genomic analyses were performed with the "qgg" package freely available at http://psoerensen.github.io/qgg/. In particular, the mixed models are efficiently solved using the average information REML implemented in DMU (Madsen *et al.* 1994). The AI-REML function in the "qgg" package provides an R interface to DMU, which can be downloaded from http://dmu.agrsci.dk/DMU/. Genes were linked to GO categories, using the BioConductor package 'org.Dm.eg.db' v. 2.14 (Carlson 2017). Estimates of GCA and SCA were obtained following the methodology of Griffing (1956) for analysis of parental and hybrid genotypes (Method II) with a random effects model using the R package "DiallelAnalysisR" (Yaseen 2016).

RESULTS

All four components of environmental variation had a heritable component

Plasticity (V_{plast}) was estimated as the slope of a linear regression of CT_{min} as a function of rearing temperature, and the variation of the plastic response ($V_{v.plast}$) was computed as the standard error of the slope (Table 1). Within-environment variation (V_{WE}) for CT_{min} was estimated using the coefficient of variation (CV = SD/mean). To enable the comparison of a single measure of within-environment variations (CVs) with the other three components of environmental variation, we computed a line mean V_{WE} (Table 1). As a measure of environmental variation across environments (V_{AE}), we computed the DGRP line specific difference between the highest CV and lowest CV across environments (Table 1). The robustness of all four measures was confirmed with resampling as described above (Supplementary Fig. S1). We observed large phenotypic differences among the DGRP lines in the four measures of environmental variation (Fig. 2) and in CT_{min} at the individual temperatures (Supplementary Table S1). The broad and narrow sense heritability estimates (H^2 and h^2) of the four components of environmental variation were all significant and varied considerably with V_{WE} and V_{AE} having low estimates, $V_{v.plast}$ an intermediate estimate, and V_{plast} having the highest estimate (Fig. 2). We observed significant positive phenotypic (ρ_p) and genetic correlations (ρ_g) among $V_{WE,}$ V_{AE} and $V_{v.plast}$, and between V_{plast} and $V_{v.plast}$ (Fig. 2), however, V_{plast} was only weakly positively correlated with the other measures of environmental variation (Fig. 2). The correlations between the rank orders of the DGRP lines for the four measures of environmental variation resembled these patterns, with the highest correlation between the rank order of V_{WE} and V_{AE} (Supplementary Fig. S2).

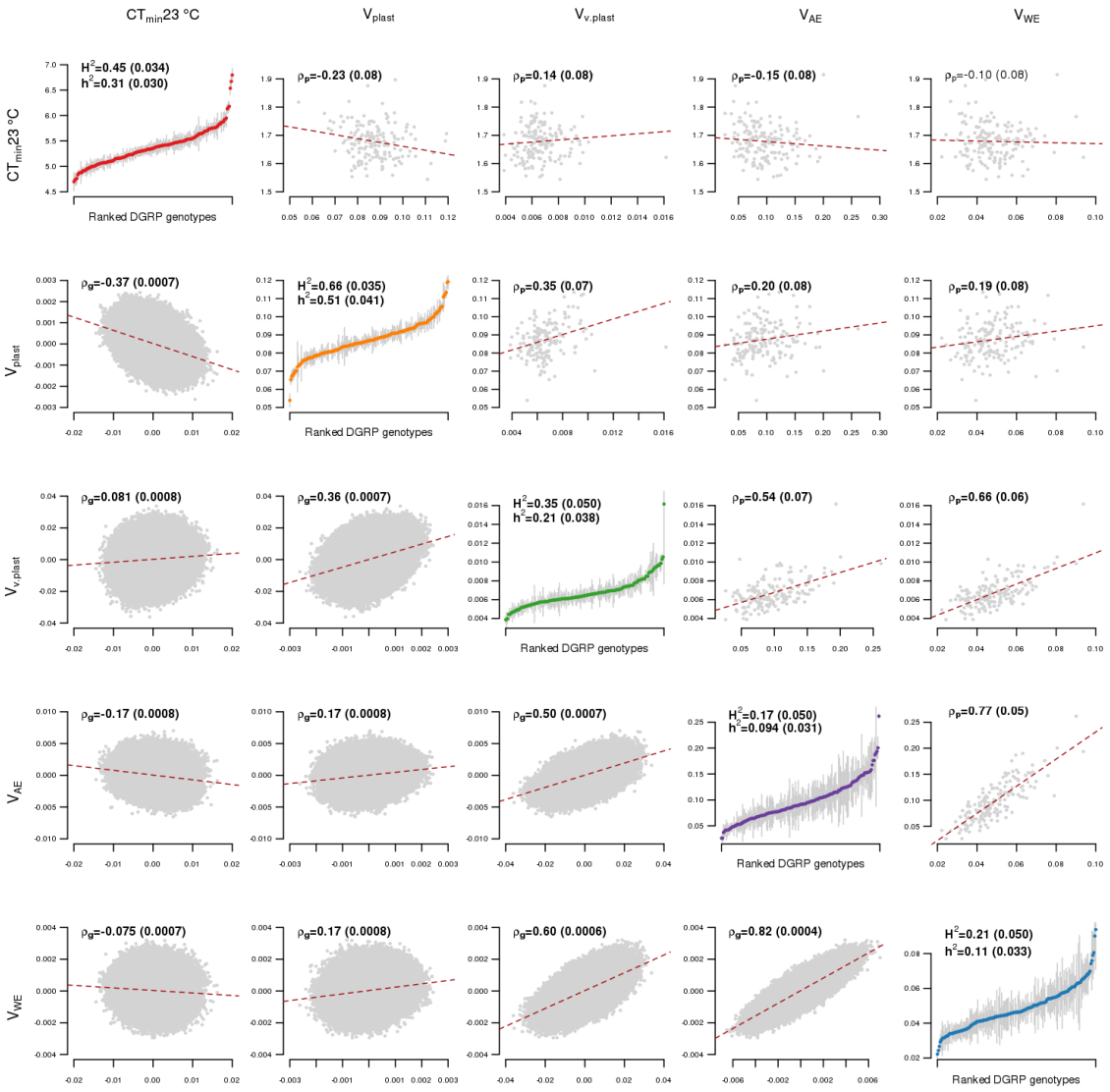


Fig. 2 Estimated genetic parameters and correlations between cold tolerance and the components of environment variance. Estimated genetic parameters for inherent cold tolerance (i.e., CT_{min} at 23 °C) and the components of environment variance (V_{plast}, $V_{v.plast}$, V_{AE}, V_{WE}). Diagonal elements show the DGRP line mean (±SE) sorted by increasing values within trait. Broad sense (H^2) and narrow sense (h^2) heritability estimates with their approximated SEs are shown. Plots above the diagonal show the phenotypic correlations (ρ_p, SE in parenthesis) between traits. Plots below the diagonal show the genetic correlations (ρ_g, SE in parenthesis). Here genetic correlations were approximated as SNP-correlations. Dashed lines are the regression lines visualizing the correlations. Genetic parameters in bold indicate values significantly different from zero, i.e. the estimate deviates more than 1.645 x SE from 0 ($p < 0.05$). Note that the y-axis across the top row changes because the correlations have been computed based on log(y), but CT_{min} values (i.e., top left panel) are shown in their original scale.

Single nucleotide polymorphisms (SNPs) associated with the four components of V_E

To disentangle the underlying genetic architecture of the four components of environmental variance, we applied whole genome single marker regression on individual genetic markers. At a nominal *p*-value threshold of $p < 1 \times 10^{-5}$ we found several SNPs associated with the four components of environmental variation (Supplementary Fig. S3 and Supplementary Table S4). In total, 148 genes (132 unique) were significantly associated with one or more of the four components of environmental variation (30 for V_{plast}, 18 for $V_{v.plast}$, 35 for V_{WE}, and 65 for V_{AE}). The majority of genes were not in common between V_E components, however, a significant overlap of 16 genes between V_{WE} and V_{AE} was found (Supplementary Table S4). In line with $\rho_p$we found significant positive genetic correlations (ρ_g) for the V_E measures (Fig. 2), supported by increased predictive ability across traits when causal variants were used in the prediction models (Supplementary Fig. S4). To investigate whether there was overlap in the functional categories associated with each type of environmental variation, we performed a gene ontology (GO) enrichment test with sets of SNPs with a *p*-value < $1x10^{-5}$. We only found a significant overlap in GO terms between V_{AE} and V_{WE} (Supplementary Fig. S5) supporting the notion that there is some degree of similarity in the genetic architecture of these two V_E components, and that the rest are decoupled on a functional level as well.

Components of environmental variation are genetically distinct from inherent cold tolerance

To assess if components of environmental variation were linked to inherent (non-plastic) cold tolerance, we used CT_{min} of DGRP lines reared at 23 °C as a measure of inherent cold tolerance. While 23 °C is arbitrary, this is a benign temperature and normally used to rear the DGRP lines. We found a weak but significant negative phenotypic correlation between CT_{min} at 23 °C and V_{plast} supported by a negative genetic correlation (ρ_g) (Fig. 2). Thus, there was a tendency that for individuals reared at 23 °C the most cold tolerant DGRP lines showed the highest degree of plasticity. Despite these correlations, we found no overlap of genes associated with inherent cold tolerance and V_{plast}. Similarly, we found no overlapping genes between inherent cold tolerance and $V_{v.plast}$, V_{WE} and V_{AE} (Supplementary Table S4) and inherent cold tolerance was a poor predictor of the four components of environmental variation (Supplementary Fig. S4).

Functional validation of candidate genes

To phenotypically validate a set candidate genes, we selected 10-11 top candidate genes for each V_E component and performed gene expression knockdown using the binary *UAS*-GAL4 system (Supplementary Table S5). The knockdown lines were reared under the same five thermal regimes as the DGRP lines, and then characterized for CT_{min} to allow estimation of the different components of V_E for each candidate gene. Each *UAS*-GAL4 line was compared to a control line with the same genetic background. These comparisons provided evidence that 16 of the 40 tested candidate genes exhibited a significantly altered phenotypic response compared to the controls in the four components of V_E (Fig. 3 and Supplementary Table S5).

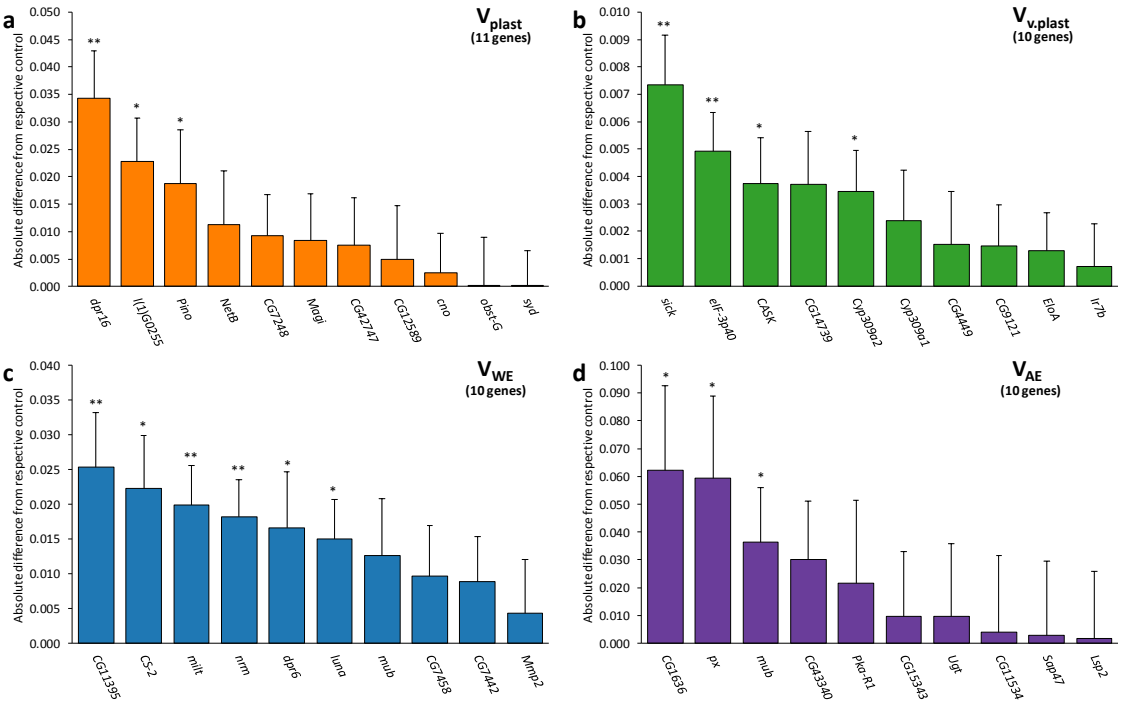


Fig. 3 Effects of gene expression knockdown of the candidate genes for the four environmental variation components. Effects of gene expression knockdown of the candidate genes for (**a**) V_{plast} in orange (11 genes), (**b**) $V_{v.plast}$ in green (10 genes), (**c**) V_{WE} in blue (10 genes), and (**d**) V_{AE} in purple (10 genes). Bars represent the absolute difference between the mean of a given gene assessed in the *UAS*-GAL4 lines and its respective genetic background. Asterisks denote significant difference from control as determined by a one-tailed Mann-Whitney *U* test between replicate values of each V_E component compared to the control (n=14): * *p*-value < 0.05, ** *p*-value < 0.01. The *p*-values has been corrected for multiple testing using the Bonferroni approach. Error bars represent standard error of the difference of means determined as $\sigma_{M_1-M_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$, where M_1 and M_2 are means, σ_1^2 and σ_2^2 are the variance among replicate values, and n_1 and n_2 are number of replicates of the *UAS*-GAL4 line and its control, respectively.

We further tested whether the genetic (ρ_g) and phenotypic (ρ_p) correlations between the four components of V_E and inherent cold tolerance (CT_{min} 23 °C) obtained using the DGRP (Fig. 2) were evident in the *UAS*-GAL4 lines, providing validation that the genes contributed to the genetic architecture of the traits in ways expected from the DGRP set. We compared the difference of each *UAS*-GAL4 line from its respective control (signed effect sizes) for all pairwise components of V_E and inherent cold tolerance (CT_{min} 23 °C) using both the 10-11 candidate genes in each of the four V_E components and all 40 candidate genes together (Supplementary Figs. S6-15). These comparisons showed that the highly significant correlations between V_{WE}, V_{AE}, and $V_{v.plast}$ in the DGRP lines were also found in the distinct genetic backgrounds of the *UAS*-GAL4 lines. This was regardless of whether only the candidate genes for the individual V_E components or all 40 candidate genes were used in the pairwise correlations (Supplementary Figs. S6-8). Similarly, the weak or non-significant correlations of ρ_g and ρ_p between V_{plast} and both V_{WE} and V_{AE}, as well between CT_{min} 23 °C and $V_{v.plast}$, V_{WE} and V_{AE} in the main experiment was confirmed using the *UAS*-GAL4 system (Supplementary Figs. S9-13). The positive correlation between V_{plast} and $V_{v.plast}$ and the negative relationship between inherent cold tolerance and V_{plast} was also confirmed with the *UAS*-GAL4 system (Supplementary Figs. S14-15). Overall, the pairwise relationships between the four components of V_E and inherent cold tolerance in the *UAS*-GAL4 lines were highly consistent with the observations in the main experiment using the DGRP (Fig. 4).

Diallel crosses show additive gene action and no inbreeding effects

Because inbreeding effects may confound interpretations of correlation patterns, we performed diallel crosses with selected DGRP lines to test if inbreeding depression affected CT_{min} (Supplementary Table S3). Out of 36 line crosses, 6 F_1's had significantly lower average CT_{min} than the mean of the parental genotypes (one sample t-test, $t_{14} < -2.70$, $p < 0.05$). We found no F_1's with lower average CT_{min} than the best performing parental genotype, i.e. with lowest CT_{min} (two-sample t-test, $t_7 > -1.75$, $p > 0.11$). Overall, there was no significant difference in CT_{min} between parental genotypes and hybrid crosses ($\chi_1^2 = 1.48$, $p = 0.22$, as determined by maximum likelihood). Based on one-way ANOVAs, a significant effect of general combining ability (GCA) was found ($F_8 = 11.55$, $p < 0.001$), whereas no significant effect of specific combining ability (SCA) ($F_{36} = 1.38$, $p = 0.075$) was detected. The GCA/SCA ratio was 3.42, suggesting that inbreeding effects are minor.

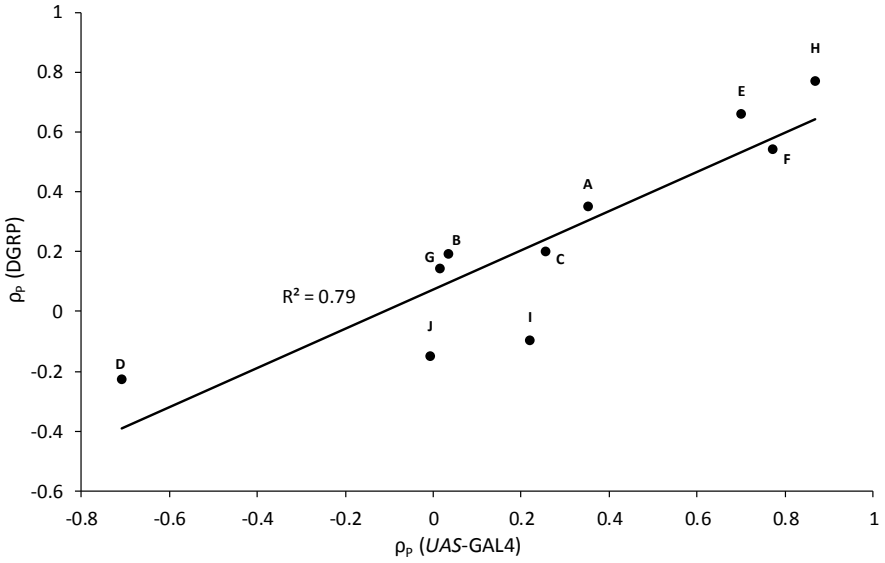


Fig. 4 Comparison of pairwise trait correlations in the DGRP with correlations obtained in the *UAS*-GAL4 genetic background. Comparison of phenotypic (ρ_P) correlations between the four V_E components and inherent cold tolerance (i.e. CT_{min} 23 °C) obtained from the DGRP (Figure 2) with the same pairwise correlations obtained when investigating top candidate genes for each component using the *UAS*-GAL4 system ($\rho_{\mathbf{P}}$ (*UAS*-GAL4)). Estimates of $\rho_{\mathbf{P}}$ (*UAS*-GAL4) were obtained as correlations for all combinations of the four V_E components and inherent cold tolerance, and using all 40 tested candidate genes in calculating the correlation, regardless of which V_E component these genes were associated with (see Supplementary Figs. S6-15). The coefficient of determination, R^2, is shown. Letters denote correlations between the following combinations of V_E components and inherent cold tolerance: A: V_{plast} and $V_{v.plast}$, B: V_{plast} and V_{WE}, C: V_{plast} and V_{AE}, D: V_{plast} and CT_{min} 23 °C, E: $V_{v.plast}$ and V_{WE}, F: $V_{v.plast}$ and V_{AE}, G: $V_{v.plast}$ and CT_{min} 23 °C, H: V_{WE} and V_{AE}, I: V_{WE} and CT_{min} 23 °C, and J: V_{AE} and CT_{min} 23 °C.

DISCUSSION

Here we undertook a comprehensive analysis of the genetic control of environmental variation (V_E) of cold tolerance in *Drosophila melanogaster*. This was obtained by partitioning environmental variation into four components: plasticity (V_{plast}), within-environment variation (V_{WE}), across environment variation (V_{AE}), and variation of plasticity ($V_{v.plast}$; Fig. 1 and Table 1). We investigated the genetic basis for each component, the independence of these components from each other and from inherent cold tolerance, and the genes governing these patterns.

We found considerable variation among the DGRP lines for within-environment variation (V_{WE}) (Fig. 2 and Supplementary Table S1) suggesting genetic control of the trait which is consistent with recent studies (Ayroles *et al.* 2015; Morgante *et al.* 2015; Sørensen *et al.* 2015). The finding that V_{WE} is under genetic control suggests that this trait will be impacted by selection and genetic drift. Such effects have implications for breeding programs where consistent trait values are often desirable (Hill & Mulder 2010; Rönnegård *et al.* 2013). For decades, V_E has been seen as separate from genetic variation in classic quantitative theory (Falconer & Mackay 1996). In traditional analyses, V_E has typically been lumped into a residual variance component as part of the variance that cannot be explained by other components in the model. Only recently, have researchers started to acknowledge that V_E also has a genetic basis (Ros *et al.* 2004; Ibáñez-Escriche *et al.* 2008; Ayroles *et al.* 2015; Morgante *et al.* 2015; Sørensen *et al.* 2015; Blasco *et al.* 2017). We expand on this with our novel proposed partitioning of V_E into four components, which has implications for evolutionary biologists and breeders alike.

We observed substantial genetic variation for plasticity (V_{plast}) (Fig. 2) in line with high heritability of plasticity in fitness related traits in *Daphnia magna* (Stoks *et al.* 2016), but inconsistent with some quantitative genetic models pointing to low genetic variation for plasticity, and thereby slow evolution of plastic responses (Berrigan & Scheiner 2004). The phenotypic and genetic correlations between V_{plast} and V_{WE} were rather low but significant (Fig. 2) suggesting that the genetic mechanisms affecting trait plasticity and variation within environments share some characteristics. This implies that it will be difficult to fully disentangle the effects of plastic and non-plastic effects as organisms respond to changing environments (Via *et al.* 1995; van Kleunen & Fischer 2005; Stoks *et al.* 2016).

We considered two novel components of environmental variation; across environment variation (V_{AE} calculated as ΔCV) representing the extent to which within-environment variation changes across environments, and variation in the plastic response ($V_{v.plast}$ calculated as standard error of the slope). The heritability estimate for $V_{v.plast}$ was relatively high, suggesting a large degree of genetic variation for this trait (Fig. 2). The estimated H^2 and h^2 for V_{AE} and V_{WE} were significant and similar, suggesting potential for evolutionary change in both within- and across

environment variation. Selection for increased V_{AE} would be beneficial for a population of similar genotypes if the optimum fluctuates in time and space such that some individuals would benefit from large V_{WE}, whereas others would increase fitness by decreased V_{WE}, resulting in a large V_{AE}. As an example of genetic variation in V_{AE} likely to be under selection, differentiated diapause or diapausing egg formation in insects can be considered, i.e. the situation where a fraction of genotypes always show diapause, a fraction of individuals diapause only at certain conditions, and a fraction that never diapause (Tuljapurkar & Istock 1993; Simons & Johnston 1997). This leads to situations where genotypes produce different levels of phenotypic variation both within and across environments.

Partitioning of V_E into four components is novel and genetic variation in these components is not captured by classic analyses of environmental variation and genotype by environment (GxE) interactions. Genotype specific plasticity can be regarded as a GxE interaction. To elaborate, our proposed partitioning of environmental variation into different components is not covered by a classic GxE paradigm, as we investigate both environmental variation within an environment (V_{WE}) and environmental variation across environments (V_{AE}). This results in potential interactions between genotype and V_{WE} and between genotype and V_{AE}, which cannot be distinguished under a GxE approach. The GxE term does not capture genotypic effects on environmental variation due to variation in the plastic response ($V_{v.plast}$), as this encompasses both variation within and across environments. Therefore, our partitioning provides new insights into the nature of genetic contributions to environmental variances and potential interactions among them.

Phenotypic and genetic correlations between inherent cold tolerance (CT_{min} for flies developed at 23 °C) and $V_{v.plast}$, V_{WE}, and V_{AE} (Fig. 2) were rather weak and differed in direction, supported by a lack of overlapping genes between inherent cold tolerance and the four components of V_E identified by marker regressions. This was confirmed in the validation using the *UAS*-GAL4 system. Gerken et al. (2015) similarly found little overlap between genes controlling basal and plastic cold tolerance. Other studies with *D. melanogaster* show variable associations between trait means and environmental variation (Harbison *et al.* 2013; Ayroles *et al.* 2015; Morgante *et al.* 2015). In the case of a positive correlation between trait means and overall V_E, directional selection for increased trait means will simultaneously increase V_E. This has been suggested as a reason for the high environmental variation typically observed in fitness related traits (Morgante *et al.* 2015). On the other hand, traits with a negative correlation between trait mean and trait variance can be of particular interest, especially in breeding programs where selecting for both increased trait value and trait homogeneity can be of economic value (Mulder *et al.* 2008; Hill & Mulder 2010). We found some evidence of similarity in the genetic mechanisms controlling V_{WE} and V_{AE} as indicated by shared candidate genes (Supplementary Table S4) and the significant genetic correlation observed between these two V_E components.

Similarly, we found an overlap of GO terms between V_{WE} and V_{AE} (Supplementary Fig. S5), which could suggest some shared characteristics on a higher functional level between these two V_E components.

We identified no overlapping genes between the V_{plast} and inherent cold tolerance (CT_{min} 23 °C) (Supplementary Table S4), suggesting partly separate genetic mechanisms for plasticity and inherent cold tolerance. The existence of regulatory genes that control plasticity but not trait means is often debated (Schlichting & Pigliucci 1993; Via *et al.* 1995), but there is supporting evidence (Schlichting & Levin 1986; Scheiner & Lyman 1991). We did, however, find significant negative correlations between inherent cold tolerance and V_{plast} (ρ_p and ρ_g) i.e. genotypes with low cold tolerance will on average tend to have low plasticity. This finding argues against a trade-off between plasticity and trait value (Murren *et al.* 2015). The debate about whether plasticity (hardening or acclimation) constrains an organism's basal temperature tolerance is ongoing (Stillman 2003; Calosi *et al.* 2008; Chown *et al.* 2010), and although trade-offs have been observed in some taxa (Stillman 2003; Gerken *et al.* 2015), a recent meta-analysis across taxa found a lack of support for the trade-off hypothesis (Gunderson *et al.* 2015). The results presented here were not confounded by inbreeding effects, as the diallel crosses showed no evidence of overdominance and only a very small degree of dominance. Furthermore, the general- and specific combining ability (GCA/SCA) ratio was well above 1, implying a greater relative importance of additive gene action compared to non-additive gene action (Griffing 1956). In accordance with other authors (Morgante *et al.* 2015), we found no association between any of the four V_E measures and residual heterozygosity across the DGRP lines (results not shown).

The majority of genes associated with V_{plast} with known biological functions were involved in regulatory processes of development (*Adgf-A* (Dolezal *et al.* 2005)), components of the nervous system (*NetB* (Labrador *et al.* 2005) and *cno* (Pérez-Gómez *et al.* 2013)), immune responses (*dpr16* (Vogel *et al.* 2003)), and in DNA repair (*Pino* (Park & Song 2008)), where we functionally validated the latter two (Fig. 3). Regulatory genes are believed to be a crucial part of the genetic control of plasticity (Schlichting & Pigliucci 1993), and might therefore also be expected to be involved in expression variation in $V_{v.plast}$. Several of the candidate genes for $V_{v.plast}$ have been shown to be involved in regulatory processes such as neurotransmitter regulation (*CASK* (Zordan *et al.* 2005)), regulation of transcription (*EloA* (Gerber *et al.* 2004)), regulation within biological pathways such as MAPK (*alph* (Baril & Therrien 2006)) and Notch (*eIF-3p40* (Zhang *et al.* 2012)), and some of the candidate genes for $V_{v.plast}$ have been identified in *Drosophila* as general environmental stress responsive genes (King-Jones *et al.* 2006; Baril *et al.* 2009). Here, we showed functionally that *eIF-3p40* and *CASK* have an effect on $V_{v.plast}$ (Fig. 3). The gene *mub* was associated with both V_{AE} and $V_{WE,}$ which we validated functionally for V_{AE} (Fig. 3). This gene has been shown to be involved in thermosensory behaviour (Hong *et al.* 2008). Similarly,

we identified candidate genes that have previously been associated with variation in cold tolerance, e.g. *Dys* (Gerken *et al.* 2015), and *Cad87A* which has been associated with thermal stress (DeSalvo *et al.* 2008).

The validation data strongly support the results from the DGRP for the individual gene functions (Fig. 3). Furthermore, the overall patterns of phenotypic relationships among the four V_E components and between V_E components and inherent cold tolerance was similar in the *UAS*-GAL4 and DGRP lines; in fact the phenotypic correlations among the V_E components obtained from each of the two distinct sets of genotypes was highly correlated (Fig. 4 and Supplementary Figs. S6-15), despite the *UAS*-GAL4 lines coming from distinct genetic backgrounds. This further strengthens our confidence in the conclusions drawn on the basis of the DGRP comparison. It also represents a novel way of utilizing this gene expression knockdown system, beyond simply functionally validating individual genes.

In conclusion, CT_{min} and all four components of environmental variation were under genetic control. Some of these components of V_E were weakly genetically correlated to each other, suggesting some shared genetic architecture. Despite being correlated, there was little overlap in genes controlling different components of environmental variation, suggesting separate selection targets. Using genomic resources we identified new candidate genes that are shared between the different components of V_E and genes specific to particular components.

Empirical studies on the genetic basis of V_E are still in their initial phase, with most studies aiming simply to document its existence (Mulder *et al.* 2008; Ayroles *et al.* 2015; Sørensen *et al.* 2015). We have conceptualized the partitioning of overall environmental variation into separate identifiable components to an extent well beyond what has been attempted before, and shown substantial genetic variance for each component. We propose that our study can work as a hypothesis-generating platform motivating future studies elucidating the nature of evolutionary forces maintaining variation for V_E components and their interactions.

AUTHOR CONTRIBUTIONS

MØ, PDR, AAH and TNK framed the conceptual ideas. The experimental work was primarily conducted by MØ and TNK. All authors contributed to the data analysis. MØ, PDR, and TNK wrote the first draft of the manuscript. All authors reviewed the final manuscript.

ACKNOWLEDGEMENTS

DGRP and *UAS* stocks were obtained from the Bloomington *Drosophila* Stock Center (NIH P40OD018537) and from Vienna *Drosophila* Resource Center (VDRC), respectively. We thank Héctor Herranz Muñoz from Copenhagen University, Denmark, for providing us with the act5-Gal4/CyO line. We thank Iben R. Jensen, Helle Blendstrup, Susan M. Hansen, and Iben V. Nielsen for invaluable laboratory assistance. This research was financed by the Danish Natural Science Research Council through a Sapere aude stipend to TNK (DFF – 4002-00036), the Science Industry Endowment Fund and a grant from the Velux Visiting Professor programme to AAH, by the Otto Mønsted Foundation to MØ, The Lundbeck Foundation, Denmark to PDR (grant number R155-2014-1724), and by the Danish Strategic Research Council (GenSAP: Centre for Genomic Selection in Animals and Plants, contract 12-132452) to PS. The authors declare no conflict of interests regarding the publication of this paper.

REFERENCES

Anderson JT, Inouye DW, McKinney AM, Colautti RI, Mitchell-Olds T (2012) Phenotypic plasticity and adaptive evolution contribute to advancing flowering phenology in response to climate change. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 3843–3852.

Araújo MB, Ferri-Yáñez F, Bozinovic F *et al.* (2013) Heat freezes niche evolution. *Ecology Letters*, **16**, 1206–1219.

Ayroles JF, Buchanan SM, O'Leary C *et al.* (2015) Behavioral idiosyncrasy reveals genetic control of phenotypic variability. *Proceedings of the National Academy of Sciences*, **112**, 6706–6711.

Baril C, Sahmi M, Ashton-Beaucage D, Stronach B, Therrien M (2009) The PP2C alphabet is a negative regulator of stress-activated protein kinase signaling in *Drosophila*. *Genetics*, **181**, 567–579.

Baril C, Therrien M (2006) Alphabet, a Ser/Thr phosphatase of the protein phosphatase 2C family, negatively regulates RAS/MAPK signaling in *Drosophila*. *Developmental Biology*, **294**, 232–245.

Beaman JE, White CR, Seebacher F (2016) Evolution of plasticity: mechanistic link between development and reversible acclimation. *Trends in Ecology and Evolution*, **31**, 237–249.

Bergman A, Siegal ML (2003) Evolutionary capacitance as a general feature of complex gene networks. *Nature*, **424**, 549–552.

Berrigan D, Scheiner SM (2004) Modeling the evolution of phenotypic plasticity. In: *Phenotypic plasticity: functional and conceptual approaches* (eds DeWitt TJ, Scheiner SM), pp. 82–97. Oxford University Press, Oxford.

Blasco A, Martínez-Álvaro M, García M-L, Ibáñez-Escriche N, Argente M-J (2017) Selection for environmental variance of litter size in rabbits. *Genetics Selection Evolution*, **49**, 48.

Cabrera CP, Navarro P, Huffman JE *et al.* (2012) Uncovering networks from genome-wide association studies via circular genomic permutation. *G3*, **2**, 1067–1075.

Calosi P, Bilton DT, Spicer JI (2008) Thermal tolerance, acclimatory capacity and vulnerability to global climate change. *Biology Letters*, **4**, 99–102.

Carlson M (2017) org.Dm.eg.db: Genome wide annotation for Fly. R package version 3.4.1.

Chown SL, Hoffmann AA, Kristensen TN *et al.* (2010) Adapting to climate change: a perspective from evolutionary physiology. *Climate Research*, **43**, 3–15.

DeSalvo MK, Voolstra CR, Sunagawa S *et al.* (2008) Differential gene expression during thermal stress and bleaching in the Caribbean coral *Montastraea faveolata*. *Molecular Ecology*, **17**, 3952–3971.

Devaux C, Lande R (2010) Selection on variance in flowering time within and among individuals. *Evolution*, **64**, 1311–1320.

DeWitt T, Scheiner S (2004) *Phenotypic plasticity, functional and conceptual approaches*. Oxford University Press, Oxford.

Dietzl G, Chen D, Schnorrer F *et al.* (2007) A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. *Nature*, **448**, 151–6.

Dolezal T, Dolezelova E, Zurovec M, Bryant PJ (2005) A role for adenosine deaminase in *Drosophila* larval development. *PLOS Biology*, **3**, e201.

Durham MF, Magwire MM, Stone EA, Leips J (2014) Genome-wide analysis in *Drosophila* reveals age-specific effects of SNPs on fitness traits. *Nature Communications*, **5**, 1–8.

Edwards SM, Sørensen IF, Sarup P, Mackay TFC, Sørensen P (2016) Genomic prediction for quantitative traits is improved by mapping variants to gene ontology categories in *Drosophila melanogaster*. *Genetics*, **203**, 1871–1883.

Falconer DS, Mackay TFC (1996) *Introduction to Quantitative Genetics*. Longman Publishing Group, Harlow, UK.

Gaertner BE, Ruedi E a, McCoy LJ *et al.* (2015) Heritable variation in courtship patterns in *Drosophila melanogaster*. *G3*, **5**, 531–9.

Gerber M, Eissenberg JC, Kong S *et al.* (2004) In vivo requirement of the RNA polymerase II elongation factor Elongin A for proper gene expression and development. *Molecular and cellular biology*, **24**, 9911–9919.

Gerken AR, Eller OC, Hahn DA, Morgan TJ (2015) Constraints, independence, and evolution of thermal plasticity: probing genetic architecture of long- and short-term thermal acclimation. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 4399–4404.

Griffing B (1956) Concept of general and specific combining ability in relation to diallel crossing systems. *Australian Journal of Biological Sciences*, **9**, 463–493.

Gunderson AR, Stillman JH, Gunderson AR (2015) Plasticity in thermal tolerance has limited potential to buffer ectotherms from global warming. *Proceedings of the Royal Society B: Biological Sciences*, **282**, 1–8.

Harbison S, McCoy L, Mackay TF (2013) Genome-wide association study of sleep in *Drosophila melanogaster*. *BMC genomics*, **14**, 281.

Hill WG, Mulder HA (2010) Genetic analysis of environmental variation. *Genetics research*, **92**, 381–395.

Hoffmann A, Sgrò C (2011) Climate change and evolutionary adaptation. *Nature*, **470**, 479–485.

Hoffmann AA, Sørensen JG, Loeschcke V (2003) Adaptation of *Drosophila* to temperature extremes: bringing together quantitative and molecular approaches. *Journal of Thermal Biology*, **28**, 175–216.

Hong ST, Bang S, Hyun S *et al.* (2008) cAMP signalling in mushroom bodies modulates temperature preference behaviour in *Drosophila*. *Nature*, **454**, 771–775.

Hunter DJ (2005) Gene-environment interactions in human diseases. *Nature Reviews Genetics*, **6**, 287–298.

Ibáñez-Escriche N, Varona L, Sorensen D, Noguera JL (2008) A study of heterogeneity of environmental variance for slaughter weight in pigs. *Animal*, **2**, 19–26.

Johnson DL, Thompson R (1995) Restricted maximum likelihood estimation of variance components for univariate animal models using sparse matrix techniques and average information. *Journal of Dairy Science*, **78**, 449–456.

Kimura MT (2004) Cold and heat tolerance of drosophilid flies with reference to their latitudinal distributions. *Oecologia*, **140**, 442–449.

King-Jones K, Horner MA, Lam G, Thummel CS (2006) The DHR96 nuclear receptor regulates xenobiotic responses in *Drosophila*. *Cell Metabolism*, **4**, 37–48.

van Kleunen M, Fischer M (2005) Constraints on the evolution of adaptive phenotypic plasticity in plants. *The New phytologist*, **166**, 49–60.

Labrador JP, O'Keefe D, Yoshikawa S *et al.* (2005) The homeobox transcription factor even-skipped regulates netrin-receptor expression to control dorsal motor-axon projections in *Drosophila*. *Current Biology*, **15**, 1413–1419.

Lande R (2014) Evolution of phenotypic plasticity and environmental tolerance of a labile quantitative character in a fluctuating environment. *Journal of Evolutionary Biology*, **27**, 866–875.

Lynch M, Walsh B (1998) *Genetics and analysis of quantitative traits*. Sinauer Associates, Sunderland.

Mackay TFC, Richards S, Stone EA *et al.* (2012) The *Drosophila melanogaster* Genetic Reference Panel. *Nature*, **482**, 173–178.

Madsen P, Jensen J, Thompson R (1994) Estimation of (co)-variance components by REML in multivariate mixed linear models using average of observed and expected information. In: *Fifth World Congress of Genetics Applied to Livestock Production*, pp. 455–462. Guelph, Ontario, Canada.

Marshall DJ, Bonduriansky R, Bussière LF (2008) Offspring size variation within broods as a bet-hedging strategy in unpredictable environments. *Ecology*, **89**, 2506–2517.

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, **157**, 1819–1829.

Montgomery SL, Vorojeikina D, Huang W *et al.* (2014) Genome-wide association analysis of tolerance to methylmercury toxicity in *Drosophila* implicates myogenic and neuromuscular developmental pathways. *PLOS ONE*, **9**, e110375.

Morgante F, Sørensen P, Sorensen D, Maltecca C, Mackay TFC (2015) Genetic architecture of micro-environmental plasticity in *Drosophila melanogaster*. *Scientific Reports*, **5**, 9785.

Mulder HA, Bijma P, Hill WG (2008) Selection for uniformity in livestock by exploiting genetic heterogeneity of residual variance. *Genet. Sel. Evol.*, **40**, 37–59.

Murren CJ, Auld JR, Callahan H *et al.* (2015) Constraints on the evolution of phenotypic plasticity: limits and costs of phenotype and plasticity. *Heredity*, **115**, 293–301.

Overgaard J, Kristensen TN, Mitchell KA, Hoffmann AA (2011) Thermal tolerance in widespread and tropical *Drosophila* species: does phenotypic plasticity increase with latitude? *The American naturalist*, **178**, 80–96.

Overgaard J, Kristensen TN, Sørensen JG (2012) Validity of thermal ramping assays used to assess thermal tolerance in arthropods. *PLOS ONE*, **7**, 1–7.

Park S-Y, Song Y-H (2008) Genetic screen for genes involved in Chk2 signaling in *Drosophila*. *Molecules and cells*, **26**, 350–355.

Pérez-Gómez R, Slováková J, Rives-Quinto N, Krejci A, Carmena A (2013) A Serrate-Notch-Canoe complex mediates essential interactions between glia and neuroepithelial cells during *Drosophila* optic lobe development. *Journal of cell science*, **126**, 4873–84.

R Core Team (2017) R: A language and environment for statistical computing. Version 3.4.0. R Foundation for Statistical Computing, Vienna, Austria. http://www.r-project.org/.

Rohde PD, Demontis D, Cuyabano BCD *et al.* (2016) Covariance association test (CVAT) identify genetic markers associated with schizophrenia in functionally associated biological processes. *Genetics*, **203**, 1901–1913.

Ros M, Sorensen D, Waagepetersen R *et al.* (2004) Evidence for genetic control of adult weight plasticity in the snail *Helix aspersa*. *Genetics*, **168**, 2089–2097.

Rönnegård L, Felleki M, Fikse WF, Mulder HA, Stransberg E (2013) Variance component and breeding value estimation for genetic heterogeneity of residual variance in Swedish Holstein dairy cattle. *J. Dairy Sci*, **96**, 2627–2636.

Sangster TA, Salathia N, Undurraga S *et al.* (2008) HSP90 affects the expression of genetic variation and developmental stability in quantitative traits. *Proc Natl Acad Sci U S A*, **105**, 2963–2968.

Sarup P, Jensen J, Ostersen T, Henryon M, Sørensen P (2016) Increased prediction accuracy using a genomic feature model including prior information on quantitative trait locus regions in purebred Danish Duroc pigs. *BMC Genetics*, **17**, 11.

Scheiner SM, Lyman RF (1991) The genetics of phenotypic plasticity II. Response to selection. *Journal of Evolutionary Biology*, **4**, 23–50.

Schlichting CD, Levin DA (1986) Effects of inbreeding on phenotypic plasticity in cultivated *Phlox*. *Theoretical and applied genetics*, **72**, 114–119.

Schlichting C, Pigliucci M (1993) Control of phenotypic plasticity via regulatory genes. *The American Naturalist*, **142**, 336–370.

Schou MF, Kristensen TN, Pedersen A *et al.* (2017a) Metabolic and functional characterization of effects of developmental temperature in *Drosophila melanogaster*. *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology*, **312**, R211–R222.

Schou MF, Mouridsen MB, Sørensen JG, Loeschcke V (2017b) Linear reaction norms of thermal limits in *Drosophila*: predictable plasticity in cold but not in heat tolerance. *Functional Ecology*, **31**, 934–945.

Simons AM, Johnston MO (1997) Developmental instability as a bet-hedging strategy. *Oikos*, **80**, 401–406.

Sorensen D, Gianola D (2002) *Likelihood, bayesian, and MCMC methods in quantitative genetics*. Springer-Verlag, New York.

Stearns SC (1989) The evolutionary significance of phenotypic plasticity. *BioScience*, **39**, 436–445.

Stillman JH (2003) Acclimation capacity underlies susceptibility to climate change. *Science*, **301**, 65.

Stoks R, Govaert L, Pauwels K, Jansen B, De Meester L (2016) Resurrecting complexity: the interplay of plasticity and rapid evolution in the multiple trait response to strong changes in predation pressure in the water flea *Daphnia magna*. *Ecology Letters*, **19**, 180–190.

Sunday JM, Bates AE, Dulvy NK (2011) Global analysis of thermal tolerance and latitude in ectotherms. *Proceedings of the Royal Society B: Biological Sciences*, **278**, 1823–1830.

Sørensen IF, Edwards SM, Rohde PD, Sørensen P (2017) Multiple trait covariance association test identifies gene ontology categories associated with chill coma recovery time in *Drosophila melanogaster*. *Scientific Reports*, **7**, 2413.

Sørensen P, de los Campos G, Morgante F, Mackay TFC, Sorensen D (2015) Genetic control of environmental variation of two quantitative traits of *Drosophila melanogaster* revealed by whole-genome sequencing. *Genetics*, **201**, 487–497.

Sørensen JG, Nielsen MM, Kruhøffer M, Justesen J, Loeschcke V (2005) Full genome gene expression analysis of the heat stress response in *Drosophila melanogaster*. *Cell stress & chaperones*, **10**, 312–328.

Teplitsky C, Mills JA, Alho JS, Yarrall JW, Merilä J (2008) Bergmann's rule and climate change revisited: disentangling environmental and genetic responses in a wild bird population. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 13492–13496.

The Gene Ontology Consortium, Ashburner M, Ball CA *et al.* (2000) Gene Ontology: tool for the unification of biology. *Nature Genetics*, **25**, 25–29.

Tuljapurkar S, Istock C (1993) Environmental uncertainty and variable diapause. *Theoretical Population Biology*, **43**, 251–280.

Vaisnav M, Xing C, Ku H-C *et al.* (2014) Genome-wide association analysis of radiation resistance in *Drosophila melanogaster*. *PLOS ONE*, **9**, e104858.

Valladares F, Sanchez-Gomez D, Zavala MA (2006) Quantitative estimation of phenotypic plasticity: bridging the gap between the evolutionary concept and its ecological applications. *Journal of Ecology*, **94**, 1103–1116.

VanRaden PM (2008) Efficient methods to compute genomic predictions. *Journal of Dairy Science*, **91**, 4414–4423.

Vasseur E, Quintana-Murci L (2013) The impact of natural selection on health and disease: uses of the population genetics approach in humans. *Evolutionary Applications*, **6**, 596–607.

Via S, Gomulkiewicz R, De Jong G *et al.* (1995) Adaptive phenotypic plasticity: consensus and controversy. *Trends in Ecology & Evolution*, **10**, 212–217.

Vogel C, Teichmann SA, Chothia C (2003) The immunoglobulin superfamily in *Drosophila melanogaster* and *Caenorhabditis elegans* and the evolution of complexity. *Development*, **130**, 6317–6328.

Williams CM, Henry HAL, Sinclair BJ (2015) Cold truths: how winter drives responses of terrestrial organisms to climate change. *Biological Reviews*, **90**, 214–235.

Yaseen M (2016) DiallelAnalysisR: diallel analysis with R. R package version 0.1.1.

Zhang X-S, Hill WG (2005) Evolution of the environmental component of the phenotypic variance: stabilizing selection in changing environments and the cost of homogeneity. *Evolution*, **59**, 1237–1244.

Zhang J, Liu M, Su Y, Du J, Zhu AJ (2012) A targeted in vivo RNAi screen reveals deubiquitinases as new regulators of Notch signaling. *G3*, **2**, 1563–1575.

Zordan MA, Massironi M, Ducato MG *et al.* (2005) *Drosophila* CAKI/CMG protein, a homolog of human CASK, is essential for regulation of neurotransmitter vesicle release. *Journal of Neurophysiology*, **94**, 1074–1083.

SUPPLEMENTARY INFORMATION FOR PAPER III

Text S1: Resampling of V_E estimates.

Figure S1: Correlations between V_E measures based on experimental groups, and on resamplings.

Figure S2: Pairwise correlations between rank order of the DGRP lines for all V_E measures.

Table S1: Line mean, standard error and number of replicates per trait for all DGRP lines. (supplied online: goo.gl/Ba1vum)

Table S2: Phenotypic correlations between V_{plast}, $V_{v.plast}$, V_{WE}, and V_{AE} based on resampled observations.

Table S3: List of DGRP lines used for assessment of inbreeding effects in diallel crosses.

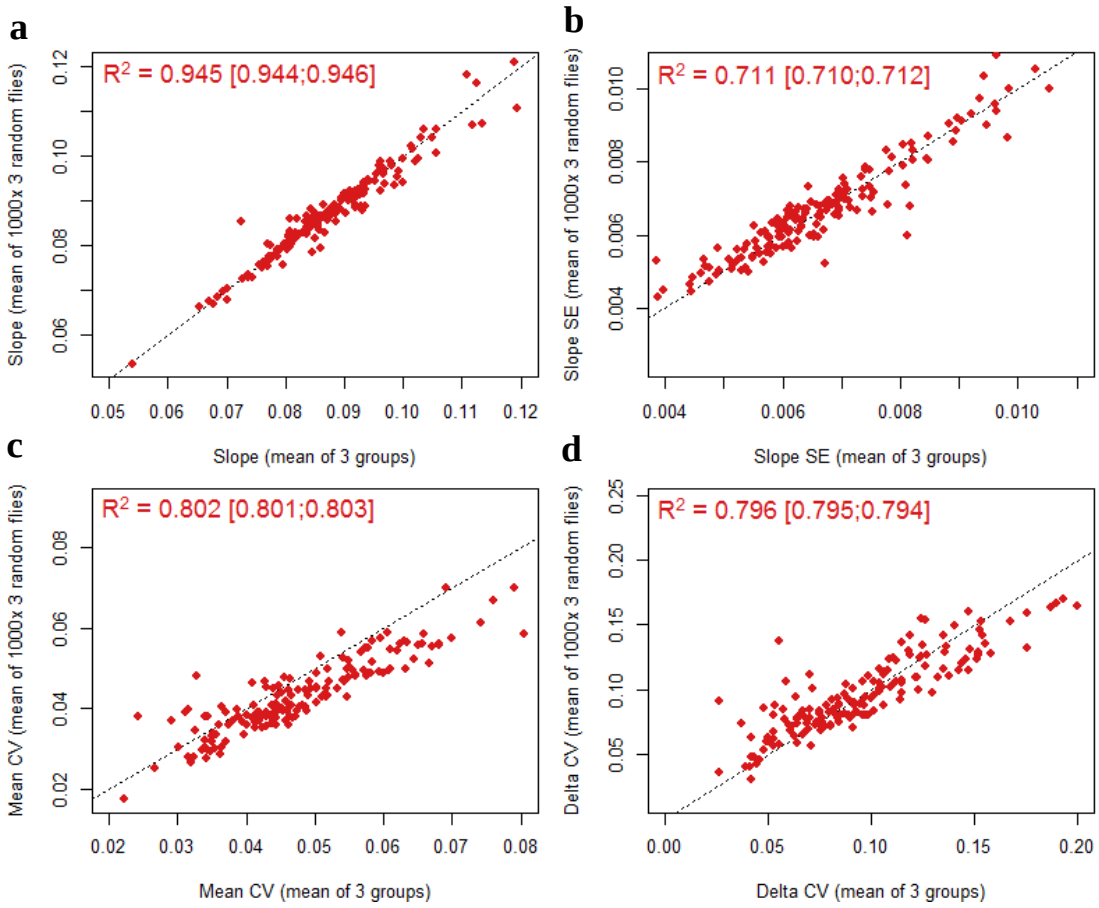
Table S4: Top SNPs associated with V_{plast}, $V_{v.plast}$, V_{WE}, and V_{AE}. (supplied online: goo.gl/ZrKKM6)

Table S5: Results of functional validation of candidate genes for V_{plast}, $V_{v.plast}$, V_{WE}, and V_{AE} using the *UAS*-GAL4 system.

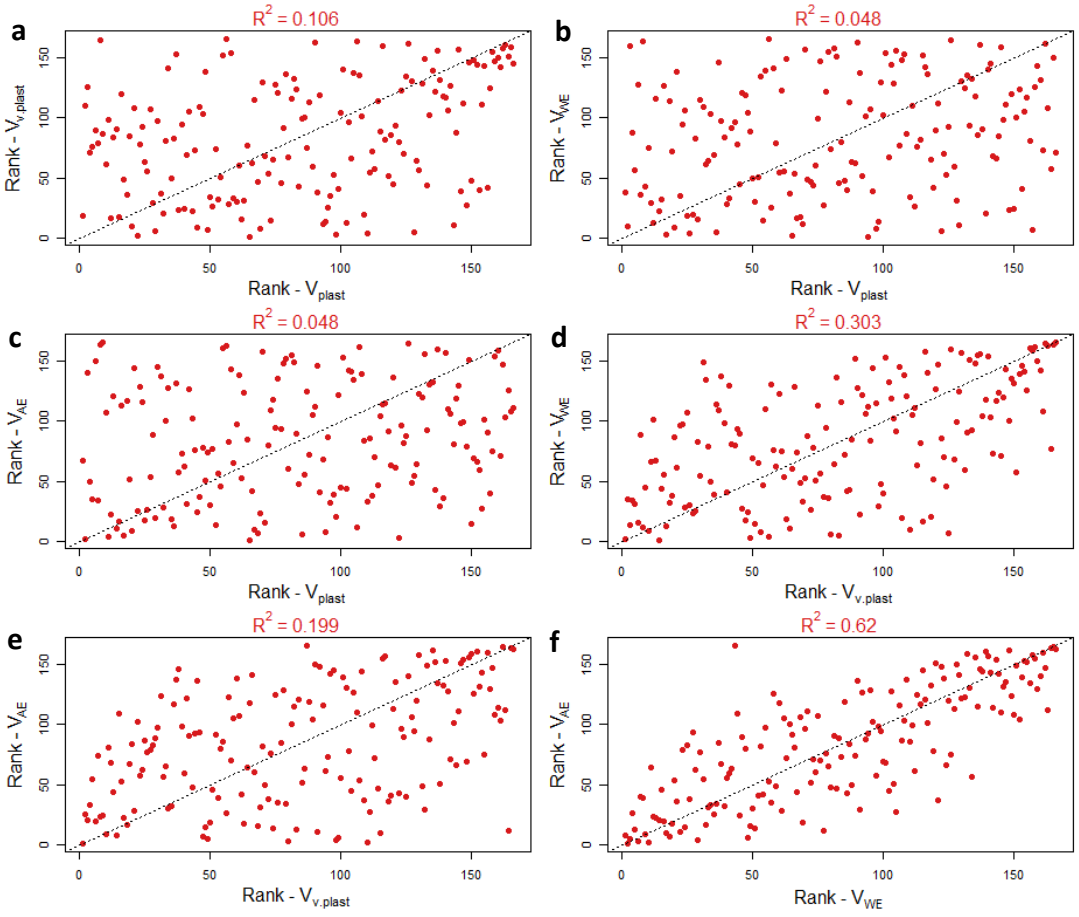
RESAMPLING OF V_E ESTIMATES

In order to retain a replicated structure of the data, the CT_{min} from each DGRP line and rearing temperature were grouped into three groups of individuals according to day of assay, time of day, and water bath, in this order. This way we also accounted for some of the variation due to experimental factors, potentially impacting V_E components, especially V_{WE}. The grouping was done because it has previously been shown that the power to detect causal variants increases with within-line replicates rather than using line means (Edwards *et al.* 2016; Sørensen *et al.* 2017). Based on the three replicate groups, the four components of environmental variation were computed (i.e. obtaining three replicates per measure per line) using log-transformed CT_{min} to ensure normality of the data. To avoid pseudo-replication, we only used each observation once, but to test the robustness of the V_E estimates, we performed 10,000 random samplings of three replicate groups of the same size as the experimental groups with replacement of observations in each line and in each environment, and calculated the four V_E components based on each of these samplings. We found high correlations between the estimates based on the three experimental groups and estimates based on randomly sampled groups across all the DGRP lines (R^2 = 0.95, 0.71, 0.80, 0.80 for V_{plast}, $V_{v.plast}$, V_{WE}, V_{AE}, respectively; Supplementary Fig. S1 below). Only V_{WE} estimates were consistently lower in the experimental groups as compared to randomly sampled groups, which is not surprising given the experimental replicates were grouped based on experimental design to account for known variance.

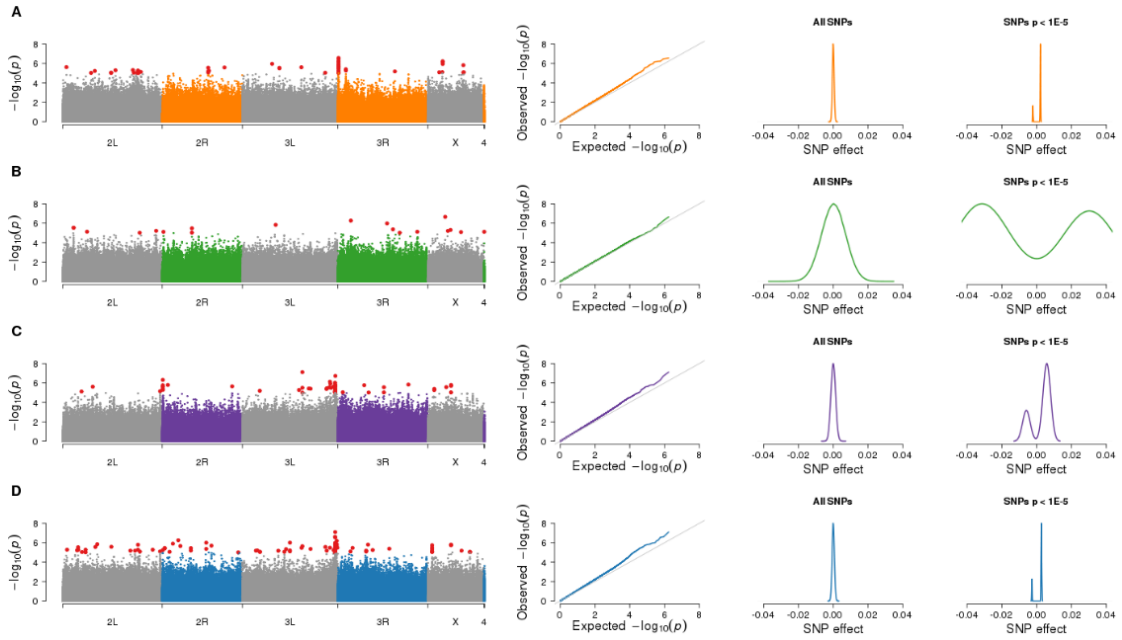
For association mapping and identification of candidate genes, it is not the magnitude of the estimates *per se* that are important, but rather the rank order of the lines. When comparing the rank order of the DGRP lines between the experimental groups and estimates based on randomly sampled groups we found equally high correlations (R^2 = 0.94, 0.86, 0.80, 0.74 for V_{plast}, $V_{v.plast}$, V_{WE}, V_{AE}, respectively). To illustrate the importance of the rank rather than estimate magnitude, we re-ran the GWA analysis on each of the four V_E estimates based on the resampled groups (10,000 x n=3). We identified all the validated candidate genes presented in the ms within the gene list from the resampled estimates. Almost 90% of the validated candidate genes were within top 10% of the candidate genes list for their respective V_E component in the GWA on the resampled data.



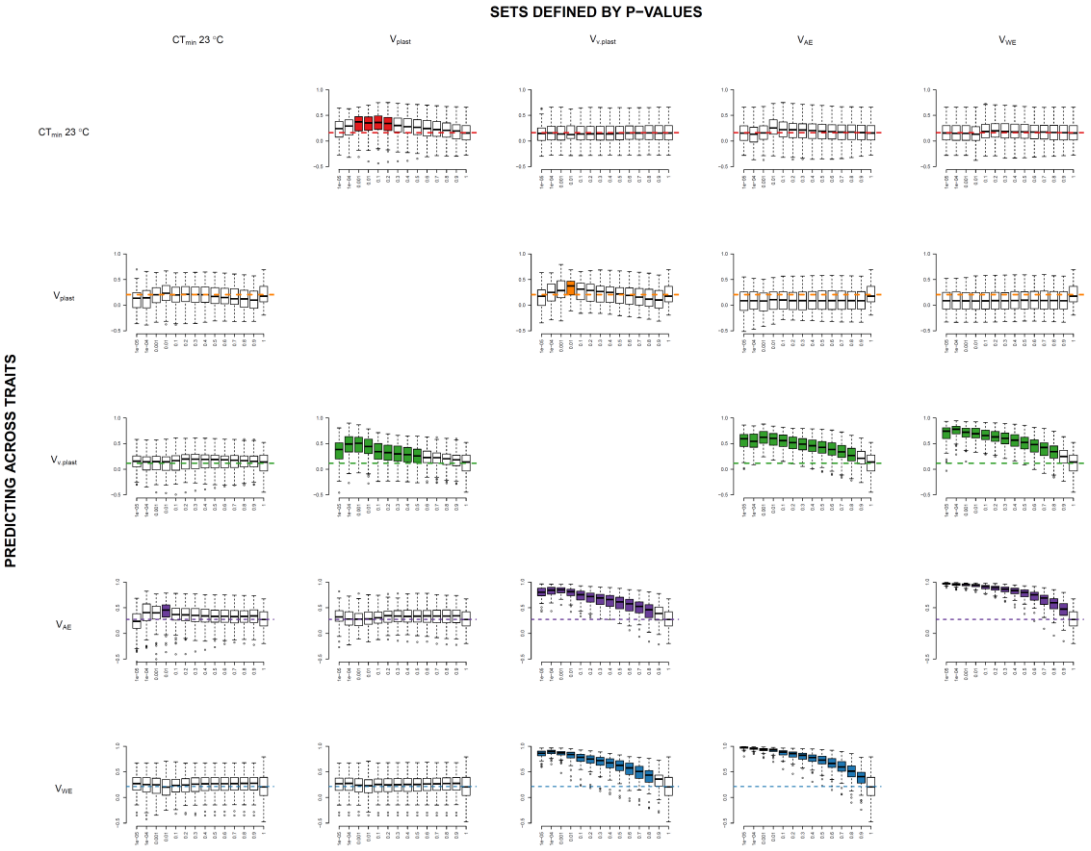
Supplementary Fig. S1. Correlations between V_E measures based on experimental groups, as used for quantitative genetic parameter estimation and association mapping, and V_E measures based on 10,000 resamplings of three equally sized groups (for each resampling the V_E measures were computed as the mean of the three groups, as for the experimental estimates). Correlations are shown for a. slope (as a measure of V_{plast}), b. slope SE (as a measure of $V_{v.plast}$), c. Mean CV ($\overline{CV}$; as a measure of V_{WE}), and d. Delta CV (ΔCV; as a measure of V_{AE}). Each circle represents the mean value of those 10,000 resampled estimates. Coefficients of determination, R^2, is shown as the mean correlation with the 95% confidence interval shown in square brackets. The dashed lines represent the 1:1 relationship.



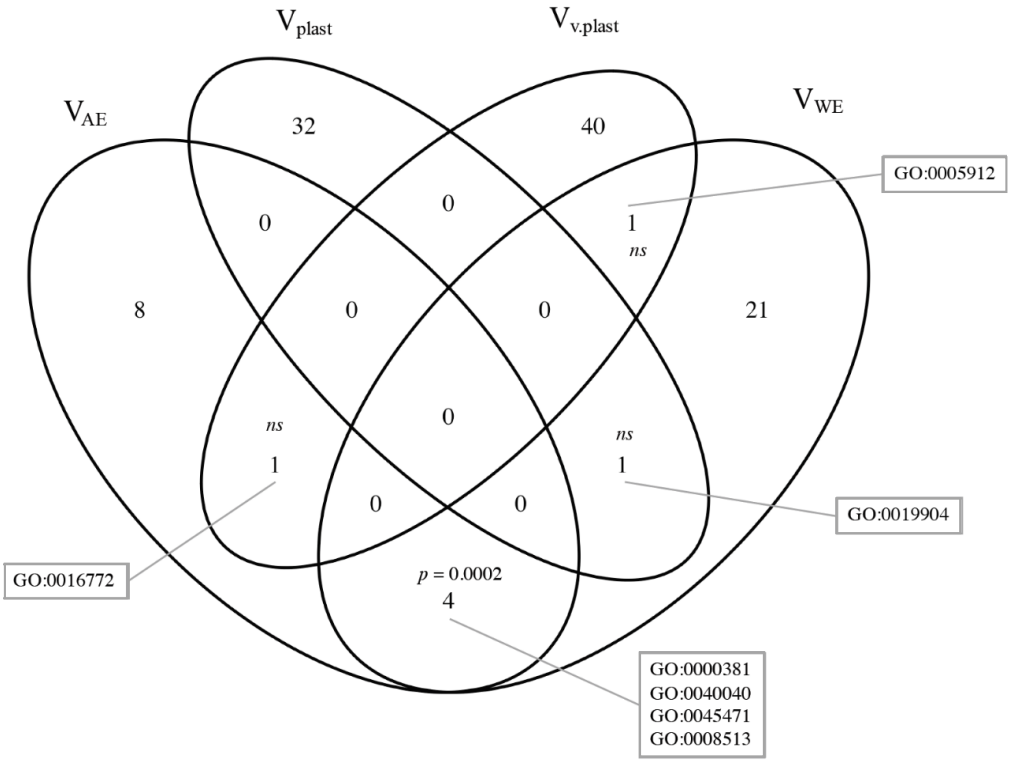
Supplementary Fig. S2. Pairwise correlations between rank order of the DGRP lines for V_E measures based on experimental groups; a. between V_{plast} and $V_{v.plast}$, b. between V_{plast} and V_{WE}, c. between V_{plast} and V_{AE}, d. between $V_{v.plast}$ and V_{WE}, e. between $V_{v.plast}$ and V_{AE}, and f. between V_{WE} and V_{AE}. Coefficients of determinations, R^2, of the correlations are shown above each panel.



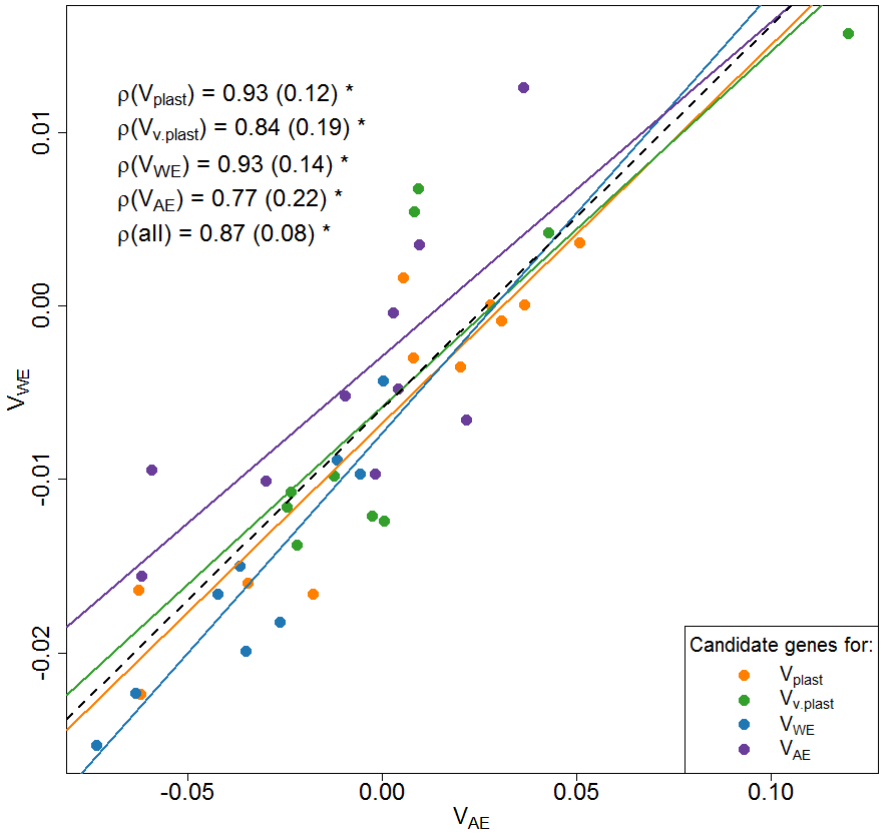
Supplementary Fig. S3. Each panel (A: $V_{plast,}$ B: $V_{v.plast}$,C: V_{AE}, and D: V_{WE}) shows a Manhattan plot with degree of association on the y-axis (i.e., -$\log_{10}$(*p*-value)) and chromosomal SNP position on x-axis, points highlighted in red indicate *p*-value < 1×10^{-5}, a quantile-quantile plot comparing the expected and observed *p*-values, the distribution of all SNP effects, and the distribution of SNP effects of SNP with *p*-value < 1×10^{-5}.



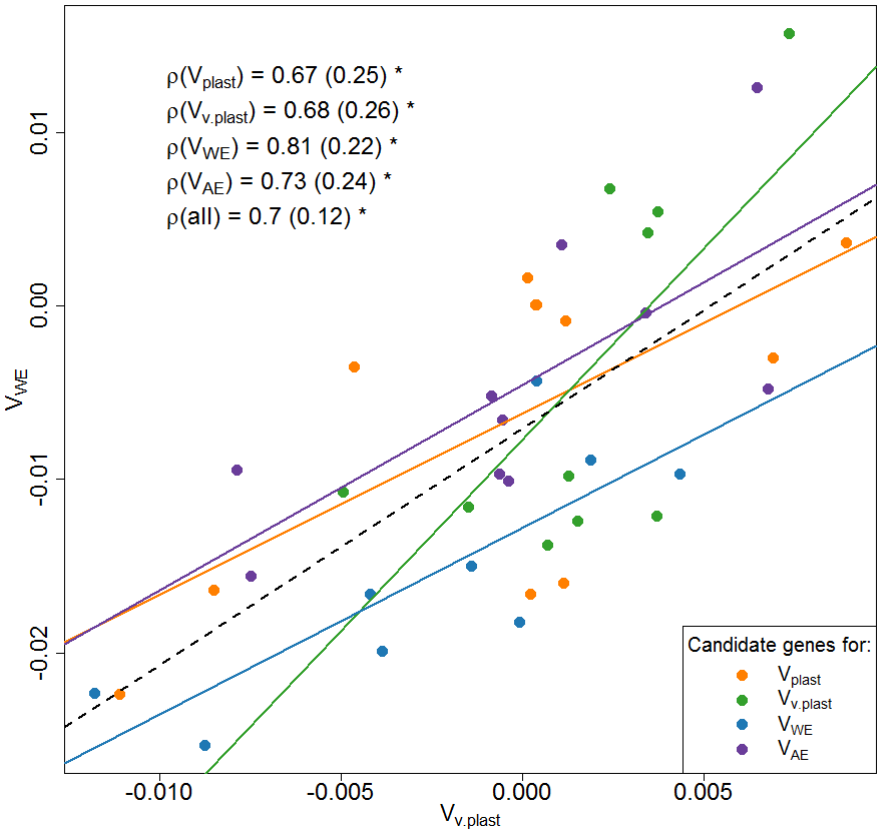
Supplementary Fig. S4. Across trait (i.e., CT_{min} 23 °C., V_{plast} ; $V_{v.plast}$, V_{AE}, V_{WE}) predictions using single marker regression results. The predictions are performed using sets of SNPs defined by single marker *p*-value (from $p < 1 \times 10^{-5}$ to $p < 1$, corresponds to the columns in the figure) to predict the phenotype of another trait (corresponds to the rows in the figure). Color coding indicates whether the set has a significant (adjusted *p*-value < 0.0001) better predictive ability than a GBLUP model within trait (horizontal line).



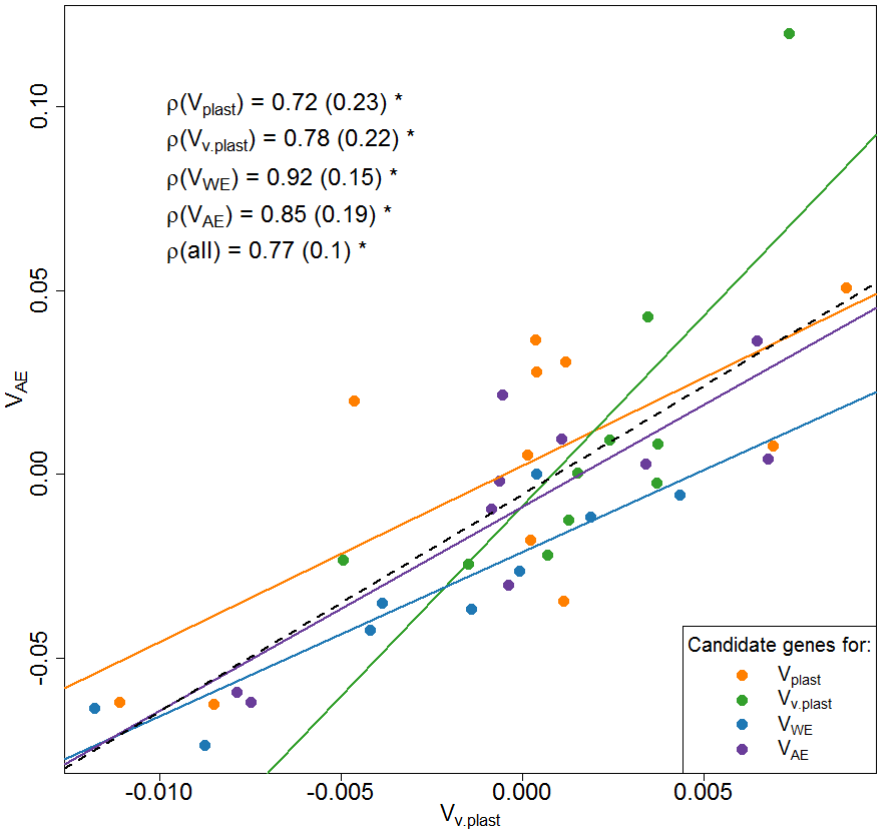
Supplementary Fig. S5. Results of the GO enrichment test analysis of each V_E component. It was tested whether a set of SNPs (in a GO term) contained more SNPs with $p < 1 \times 10^{-5}$ than a randomly sampled set containing the same amount of SNPs. To test the significance level, a count based method was used (Rohde *et al.* 2016; Sørensen *et al.* 2017). The Venn-diagram shows the unique GO terms for each of the four V_E components, overlapping GO terms, and the *p*-value of the overlap (*ns* is non-significant). Only the overlap between V_{AE} and V_{WE} was significantly larger than expected by chance. The GO identification code is only shown for overlapping GO terms.



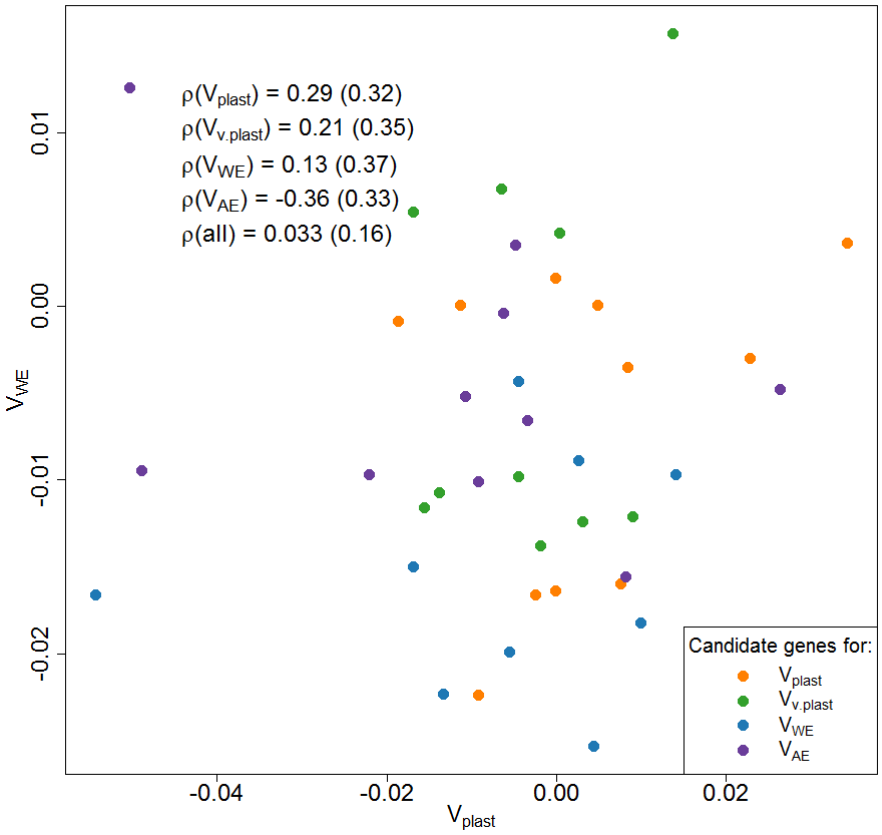
Supplementary Fig. S6. For the UAS-GAL4 lines we obtained the signed differences between VWE and its corresponding control (signed effect sizes) and between V_{AE} and its corresponding control (signed effect sizes) for each of the 40 lines. Colors represent *UAS*-GAL4 lines selected for each of the four components of V_E (top candidate genes selected for each trait). Black dashed line shows the linear regression based on all 40 candidate genes (degree of association is presented as Pearson's correlation coefficient ρ(all)). In addition, regressions were performed on *UAS*-GAL4 lines selected based on top candidate genes for the individual components of V_E (10-11 genes depending on the trait), and the degree of association between them are shown as Pearson's correlation coefficients (ρ values). The SE of the ρ values are given in parenthesis. Significant associations are marked with an asterisks (*). A linear regression line is inserted when the association is significant.



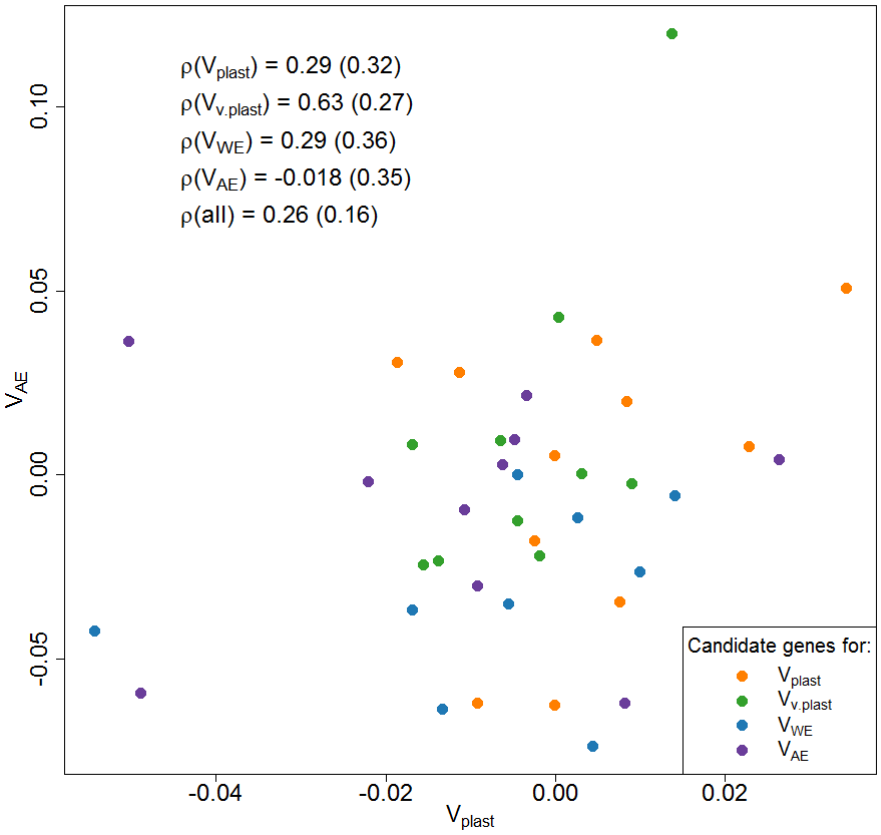
Supplementary Fig. S7. For the *UAS*-GAL4 lines we obtained the signed differences between V_{WE} and its corresponding control (signed effect sizes) and between $V_{v.plast}$ and its corresponding control (signed effect sizes) for each of the 40 lines. Colors represent *UAS*-GAL4 lines selected for each of the four components of V_E (top candidate genes selected for each trait). Black dashed line shows the linear regression based on all 40 candidate genes (degree of association is presented as Pearson's correlation coefficient ρ(all)). In addition, regressions were performed on *UAS*-GAL4 lines selected based on top candidate genes for the individual components of V_E (10-11 genes depending on the trait), and the degree of association between them are shown as Pearson's correlation coefficients (ρ values). The SE of the ρ values are given in parenthesis. Significant associations are marked with an asterisks (*). A linear regression line is inserted when the association is significant.



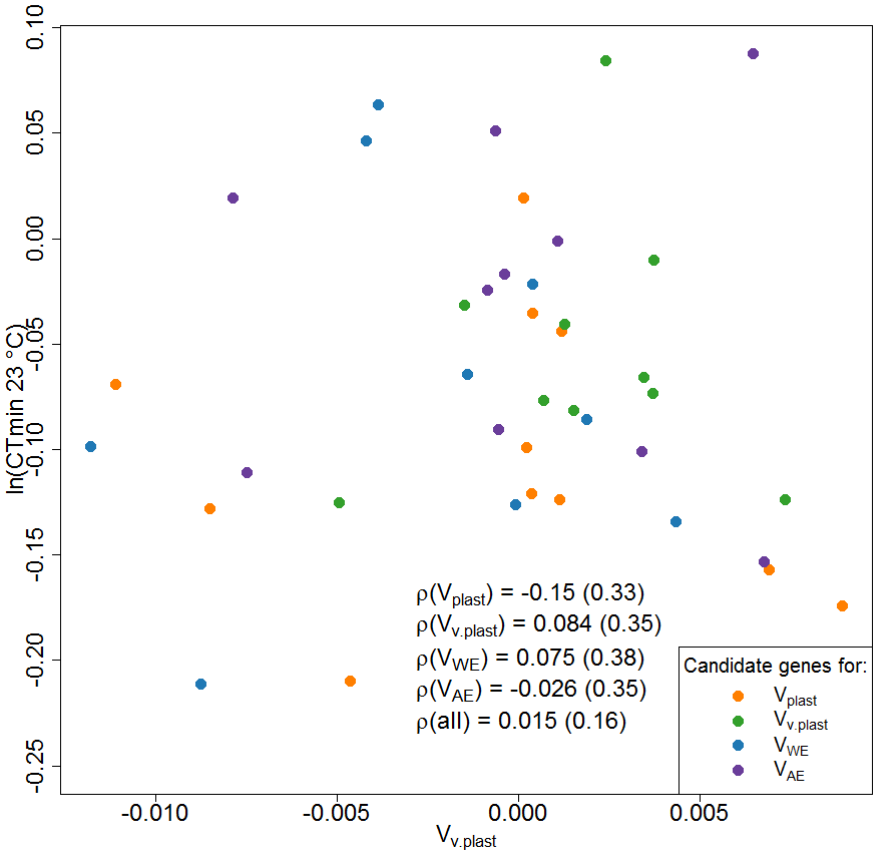
Supplementary Fig. S8. For the *UAS*-GAL4 lines we obtained the signed differences between V_{AE} and its corresponding control (signed effect sizes) and between $V_{v.plast}$ and its corresponding control (signed effect sizes) for each of the 40 lines. Colors represent *UAS*-GAL4 lines selected for each of the four components of V_E (top candidate genes selected for each trait). Black dashed line shows the linear regression based on all 40 candidate genes (degree of association is presented as Pearson's correlation coefficient ρ(all)). In addition, regressions were performed on *UAS*-GAL4 lines selected based on top candidate genes for the individual components of V_E (10-11 genes depending on the trait), and the degree of association between them are shown as Pearson's correlation coefficients (ρ values). The SE of the ρ values are given in parenthesis. Significant associations are marked with an asterisks (*). A linear regression line is inserted when the association is significant.



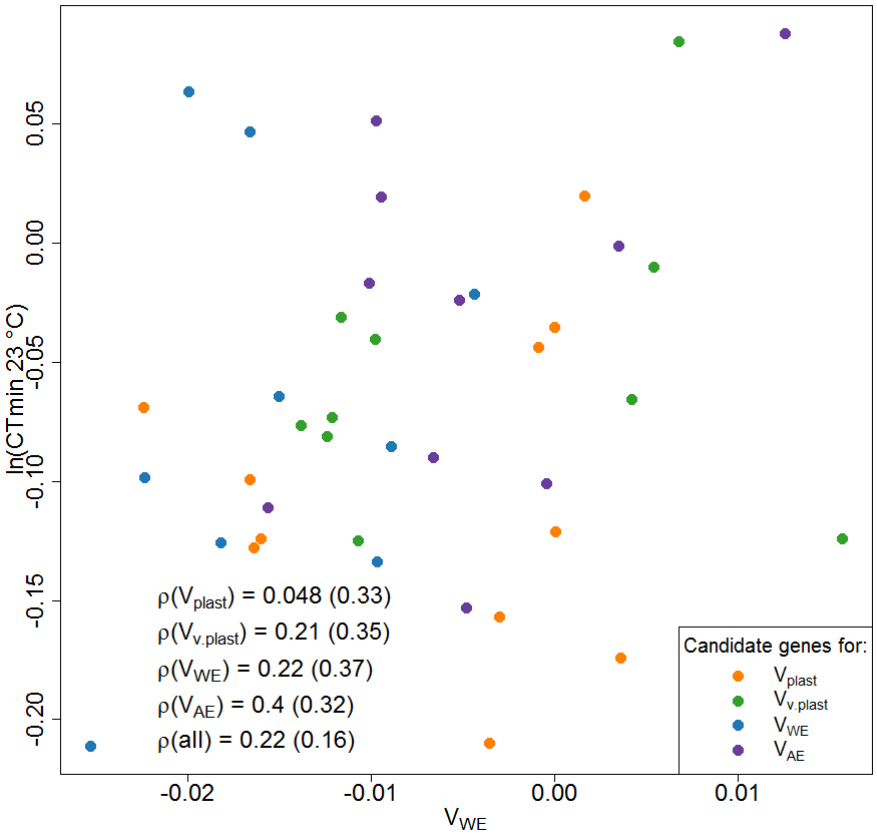
Supplementary Fig. S9. For the *UAS*-GAL4 lines we obtained the signed differences between V_{WE} and its corresponding control (signed effect sizes) and between V_{plast} and its corresponding control (signed effect sizes) for each of the 40 lines. Colors represent *UAS*-GAL4 lines selected for each of the four components of V_E (top candidate genes selected for each trait). Linear regression based on all 40 candidate genes (degree of association is presented as Pearson's correlation coefficient ρ(all)). In addition, regressions wre performed on *UAS*-GAL4 lines selected based on top candidate genes for the individual components of V_E (10-11 genes depending on the trait), and the degree of association between them are shown as Pearson's correlation coefficients (ρ values). The SE of the ρ values are given in parenthesis. None of the regressions presented on the figure are significantly different from zero.



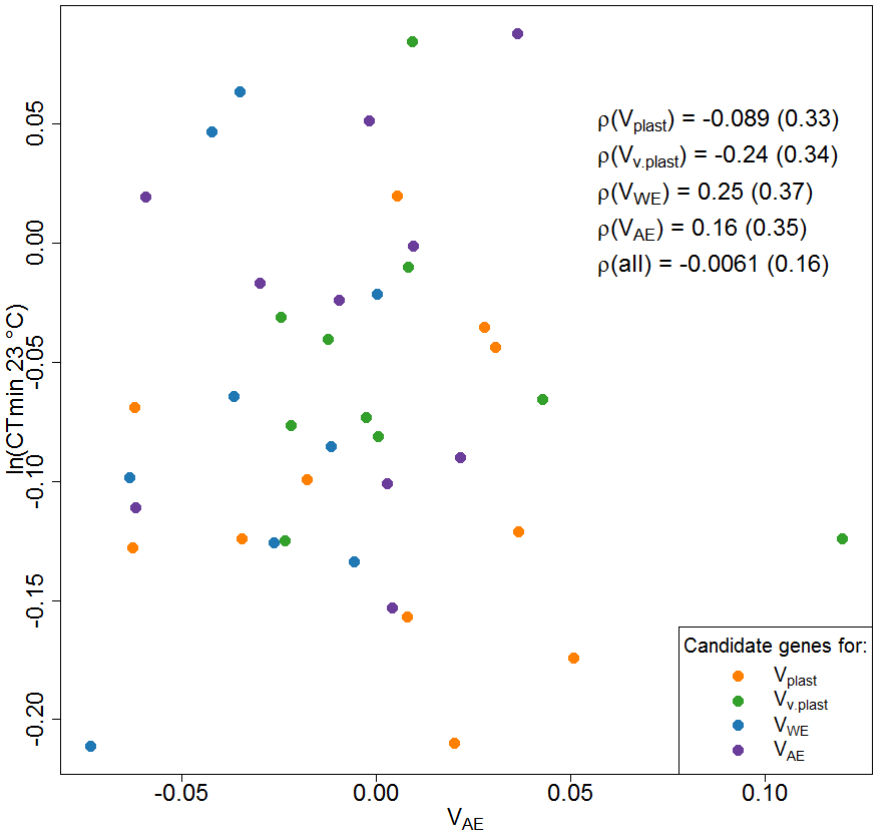
Supplementary Fig. S10. For the *UAS*-GAL4 lines we obtained the signed differences between V_{AE} and its corresponding control (signed effect sizes) and between V_{plast} and its corresponding control (signed effect sizes) for each of the 40 lines. Colors represent *UAS*-GAL4 lines selected for each of the four components of V_E (top candidate genes selected for each trait). Linear regression based on all 40 candidate genes (degree of association is presented as Pearsons' correlation coefficient ρ(all)). In addition, regressions were performed on *UAS*-GAL4 lines selected based on top candidate genes for the individual components of V_E (10-11 genes depending on the trait), and the degree of association between them are shown as Pearson's correlation coefficients (ρ values). The SE of the ρ values are given in parenthesis. None of the regressions presented on the figure are significantly different from zero.



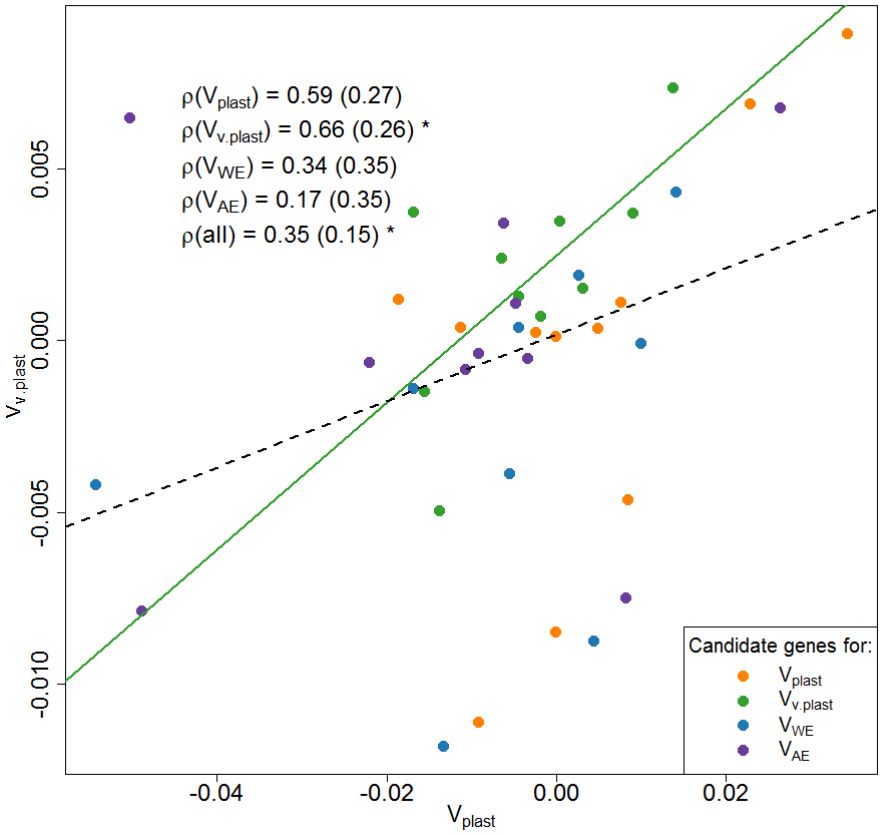
Supplementary Fig. S11. For the *UAS*-GAL4 lines we obtained the signed differences between ln(CT_{min} 23 °C) and its corresponding control (signed effect sizes) and between $V_{v.plast}$ and its corresponding control (signed effect sizes) for each of the 40 lines. Colors represent *UAS*-GAL4 lines selected for each of the four components of V_E (top candidate genes selected for each trait). Linear regression based on all 40 candidate genes (degree of association is presented as Pearson's correlation coefficient ρ(all)). In addition, regressions were performed on *UAS*-GAL4 lines selected based on top candidate genes for the individual components of V_E (10-11 genes depending on the trait), and the degree of association between them are shown as Pearson's correlation coefficients (ρ values). The SE of the ρ values are given in parenthesis. None of the regressions presented on the figure are significantly different from zero.



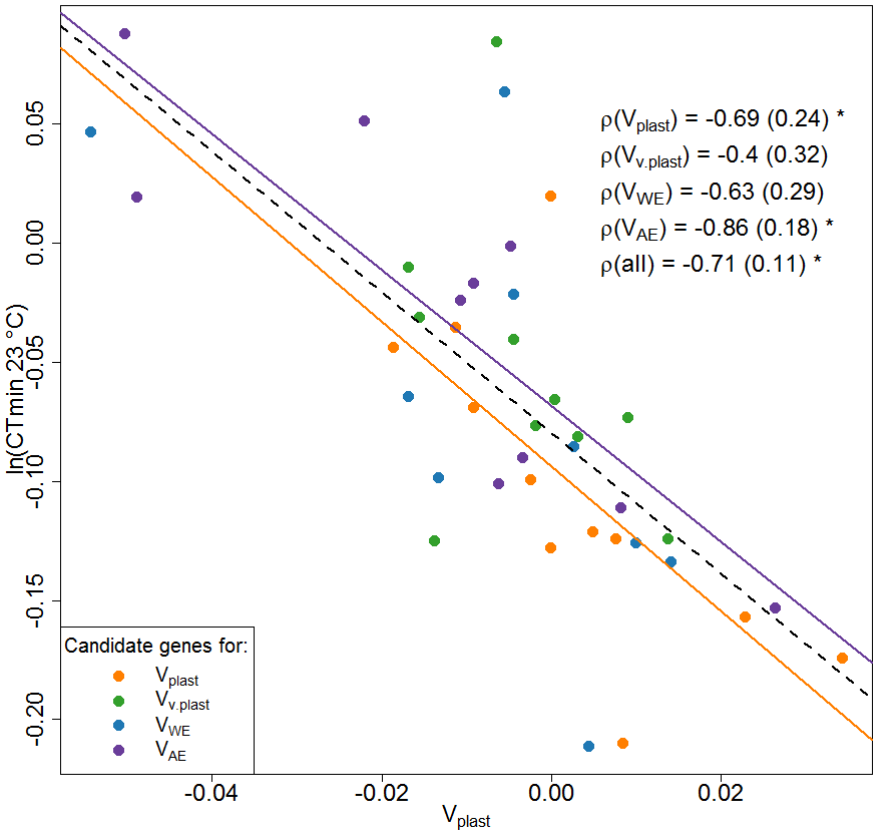
Supplementary Fig. S12. For the *UAS*-GAL4 lines we obtained the signed differences between ln(CT_{min} 23 °C) and its corresponding control (signed effect sizes) and between V_{WE} and its corresponding control (signed effect sizes) for each of the 40 lines. Colors represent *UAS*-GAL4 lines selected for each of the four components of V_E (top candidate genes selected for each trait). Linear regression based on all 40 candidate genes (degree of association is presented as Pearson's correlation coefficient ρ(all)). In addition, regressions wre performed on *UAS*-GAL4 lines selected based on top candidate genes for the individual components of V_E (10-11 genes depending on the trait), and the degree of association between them are shown as Pearson's correlation coefficients (ρ values). The SE of the ρ values are given in parenthesis. None of the regressions presented on the figure are significantly different from zero.



Supplementary Fig. S13. For the *UAS*-GAL4 lines we obtained the signed differences between ln(CT_{min} 23 °C) and its corresponding control (signed effect sizes) and between V_{WE} and its corresponding control (signed effect sizes) for each of the 40 lines. Colors represent *UAS*-GAL4 lines selected for each of the four components of V_E (top candidate genes selected for each trait). Linear regression based on all 40 candidate genes (degree of association is presented as Pearson's correlation coefficient ρ(all)). In addition, regressions wre performed on *UAS*-GAL4 lines selected based on top candidate genes for the individual components of V_E (10-11 genes depending on the trait), and the degree of association between them are shown as Pearson's correlation coefficients (ρ values). The SE of the ρ values are given in parenthesis. None of the regressions presented on the figure are significantly different from zero.



Supplementary Fig. S14. For the *UAS*-GAL4 lines we obtained the signed differences between $V_{v.plast}$ and its corresponding control (signed effect sizes) and between V_{plast} and its corresponding control (signed effect sizes) for each of the 40 lines. Colors represent *UAS*-GAL4 lines selected for each of the four components of V_E (top candidate genes selected for each trait). Black dashed line shows the linear regression based on all 40 candidate genes (degree of association is presented as Pearson's correlation coefficient ρ(all)). In addition, regressions were performed on *UAS*-GAL4 lines selected based on top candidate genes for the individual components of V_E (10-11 genes depending on the trait), and the degree of association between them are shown as Pearson's correlation coefficients (ρ values). The SE of the ρ values are given in parenthesis. Significant associations are marked with an asterisks (*). A linear regression line is inserted when the association is significant.



Supplementary Fig. S15. For the *UAS*-GAL4 lines we obtained the signed differences between ln(CT_{min} 23 °C) and its corresponding control (signed effect sizes) and between V_{plast} and its corresponding control (signed effect sizes) for each of the 40 lines. Colors represent *UAS*-GAL4 lines selected for each of the four components of V_E (top candidate genes selected for each trait). Black dashed line shows the linear regression based on all 40 candidate genes (degree of association is presented as Pearson's correlation coefficient ρ(all)). In addition, regressions were performed on *UAS*-GAL4 lines selected based on top candidate genes for the individual components of V_E (10-11 genes depending on the trait), and the degree of association between them are shown as Pearson's correlation coefficients (ρ values). The SE of the ρ values are given in parenthesis. Significant associations are marked with an asterisks (*). A linear regression line is inserted when the association is significant.

Supplementary Table S1. Line mean per trait. This table is supplied online: goo.gl/Ba1vum. Line mean, standard error (SE) and number of replicates (*n*) per DGRP line per trait. It should be noted that each of the four V_E measures is presented as a mean of the three replicate experimental groups (*n*=3), however each of those groups consists of flies from all environments (i.e. about 16.6 flies on average).

Supplementary Table S2. Phenotypic correlations between V_{plast}, $V_{v.plast}$, V_{WE}, and V_{AE} based on resampled observations. Phenotypic correlations between V_{plast}, $V_{v.plast}$, V_{WE}, and V_{AE} based on the resampled estimates (10,000 x n=3). The numbers presented are the mean of 10,000 Pearson product-moment correlation coefficients, *r* for the resampled data followed by 95% confidence intervals in square brackets. The mean of these correlations were compared to the phenotypic correlations using the experimental groupings (Fig. 2), which are given in parenthesis in the table for reference. A Fisher's z-transformation (*n*=166 for both correlation coefficients) was performed to test whether there were significant differences in the phenotypic correlations, and we found no such differences.

	V_{plast}	**$V_{v.plast}$**	**V_{AE}**	**V_{WE}**
V_{plast}	-	0.33 [0.32;0.34] (0.35)	0.24 [0.23;0.25] (0.20)	0.23 [0.22;0.24] (0.19)
$V_{v.plast}$		-	0.52 [0.51;0.53] (0.54)	0.58 [0.58;0.59] (0.66)
V_{AE}			-	0.84 [0.84;0.85] (0.77)
V_{WE}				-

Supplementary Table S3. The nine DGRP lines used to assess whether differences in CT_{min} between DGRP lines was due to line-specific inbreeding effects. The top four lines had low average CT_{min} (averaged across the five rearing temperatures) in the main experiment and the bottom five had high average CT_{min}. All of these DGRP lines were free of *Wolbachia*. Half diallel crosses were set up using these parental lines; all possible crosses between the selected DGRP lines were performed, resulting in a total of 36 hybrid crosses (approximately equal number of crosses were done with males and females from each line).

Genotype	Mean CT_{min} (°C)
DGRP_332	4.41
DGRP_894	4.75
DGRP_38	4.76
DGRP_301	4.84
DGRP_138	5.79
DGRP_492	5.81
DGRP_559	5.96
DGRP_627	5.98
DGRP_443	5.99

Supplementary Table S4. Top SNPs associated with V_{plast}, $V_{v.plast}$, V_{WE}, and V_{AE}. This table is supplied online: goo.gl/ZrKKM6. Top SNPs associated with each of the four components of environmental variation (V_{plast}, $V_{v.plast}$, V_{WE}, and V_{AE}), i.e. SNPs with a nominal *p*-value < 1×10^{-5}. Each table contains the SNP ID ('*snpname'*), the chromosome on which the SNP is located ('*chr*'), the gene which the SNP is in physical proximity too ('*FB*'), the gene symbol ('*sym*'), the type of sequence ontology the SNP is located in ('*coding*'), the distance in bases from SNP to gene ('*dist.*'), the SNP effect ('*b*'), the t-test statistic for association ('*t*'), the *p*-value of association ('*pt*'), and the minor allele frequency ('*MAF*').

Supplementary Table S5. Results of functional validation of candidate genes for V_{plast}, $V_{v.plast}$, V_{WE}, and V_{AE} using the *UAS*-GAL4 system. Effects of gene expression knockdown of the candidate genes for V_{plast} (11 genes), $V_{v.plast}$ (10 genes), V_{WE} (10 genes), and V_{AE} (10 genes). Values represent the absolute difference between the mean of a given gene assessed in the *UAS*-GAL4 lines and its respective genetic background. Absolute differences are also shown for CT_{min} 23 °C. Asterisks denote significant difference from control as determined by a one-tailed Mann-Whitney *U* test between replicate values of each V_E component compared to the control (n=14): * *p*-value < 0.05, ** *p*-value < 0.01, *** *p*-value < 0.001. To correct for multiple testing *p*-values are Bonferroni corrected. SE represent standard error of the difference of means determined as $\sigma_{M_1-M_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$, where M_1 and M_2 are means, σ_1^2 and σ_2^2 are the variances among replicate values, and n_1 and n_2 are number of replicates of the *UAS*-GAL4 line and its control, respectively.

		CT_{min} 23 °C			V_{WE}			V_{AE}			V_{plast}			$V_{v.plast}$		
Gene	Trait	Abs. diff.	SE		Abs. diff.	SE		Abs. diff.	SE		Abs. diff.	SE		Abs. diff.	SE	
dpr16	V_{plast}	0.174	0.0197	***	0.0036	0.0079		0.0507	0.0279		0.0343	0.0087	**	0.0089	0.002	**
l(1)G0255	V_{plast}	0.157	0.0199	***	0.003	0.0062		0.0078	0.0243		0.0228	0.008	*	0.0069	0.0013	**
Pino	V_{plast}	0.0438	0.0186	*	0.0009	0.0051		0.0307	0.0224		0.0187	0.0098	*	0.0012	0.0017	
NetB	V_{plast}	0.0354	0.0182	**	0	0.0055		0.0278	0.0213		0.0113	0.0097		0.0004	0.0019	
CG7248	V_{plast}	0.0691	0.0245	**	0.0224	0.0079	*	0.0622	0.0298	*	0.0093	0.0075		0.0111	0.0026	**
Magi	V_{plast}	0.21	0.028	***	0.0035	0.0104		0.02	0.0483		0.0084	0.0085		0.0046	0.0029	
CG42747	V_{plast}	0.124	0.0146	***	0.016	0.0048	**	0.0346	0.0194	*	0.0076	0.0085		0.0011	0.0015	
CG12589	V_{plast}	0.121	0.0143	***	0.0001	0.0086		0.0367	0.0323		0.0049	0.0098		0.0004	0.0016	
cno	V_{plast}	0.0993	0.017	***	0.0166	0.005	**	0.0179	0.0172		0.0025	0.0072		0.0002	0.0016	
obst-G	V_{plast}	0.128	0.0223	***	0.0164	0.0077	*	0.0627	0.0318	*	0.0001	0.0089		0.0085	0.0027	**
syd	V_{plast}	0.0194	0.019		0.0016	0.0069		0.0054	0.0246		0.0001	0.0065		0.0001	0.0018	
sick	$V_{v.plast}$	0.124	0.0162	***	0.0157	0.0113		0.12	0.0472	*	0.0137	0.0067		0.0074	0.0018	**
eIF-3p40	$V_{v.plast}$	0.125	0.0173	***	0.0107	0.0049	*	0.0234	0.019		0.0138	0.0072	*	0.0049	0.0014	**
CASK	$V_{v.plast}$	0.0102	0.0249		0.0054	0.005		0.0083	0.0207		0.0169	0.009		0.0037	0.0017	*
CG14739	$V_{v.plast}$	0.0732	0.017	***	0.0121	0.0087		0.0025	0.0258		0.009	0.0101		0.0037	0.0019	
Cyp309a2	$V_{v.plast}$	0.0657	0.0138	***	0.0042	0.0075		0.0428	0.0335		0.0004	0.0075		0.0035	0.0015	*
Cyp309a1	$V_{v.plast}$	0.0842	0.018	***	0.0068	0.0089		0.0092	0.0313		0.0065	0.0106		0.0024	0.0018	
CG4449	$V_{v.plast}$	0.0813	0.0144	***	0.0124	0.0065	*	0.0005	0.0268		0.0031	0.0081		0.0015	0.0019	
CG9121	$V_{v.plast}$	0.0314	0.0198		0.0116	0.0072	*	0.0245	0.0215		0.0156	0.0085		0.0015	0.0015	
EloA	$V_{v.plast}$	0.0407	0.0162	**	0.0098	0.007		0.0124	0.0301		0.0045	0.0088		0.0013	0.0014	
Ir7b	$V_{v.plast}$	0.0767	0.0152	***	0.0138	0.0071		0.0219	0.0216		0.0019	0.0073		0.0007	0.0016	
CG11395	V_{WE}	0.211	0.0249	***	0.0253	0.0079	**	0.0737	0.0307	*	0.0044	0.0078		0.0088	0.0028	*
CS-2	V_{WE}	0.0986	0.0213	***	0.0223	0.0075	*	0.0636	0.0301	*	0.0134	0.0074		0.0118	0.0026	**
milt	V_{WE}	0.0635	0.0165	***	0.0199	0.0056	**	0.0351	0.0202		0.0055	0.0075		0.0039	0.0016	*
nrm	V_{WE}	0.126	0.0144	***	0.0182	0.0053	**	0.0263	0.0243		0.0099	0.0092		0.0001	0.0017	
dpr6	V_{WE}	0.0465	0.0195	*	0.0166	0.008	*	0.0424	0.0225	*	0.0544	0.0098	**	0.0042	0.0024	
luna	V_{WE}	0.0645	0.0185	***	0.015	0.0057	*	0.0368	0.0194	*	0.0169	0.0071	*	0.0014	0.0016	
mub	V_{WE}	0.0877	0.0179	***	0.0126	0.0082		0.0364	0.0195	*	0.0503	0.0087	**	0.0065	0.0018	**
CG7458	V_{WE}	0.134	0.0207	***	0.0097	0.0072		0.0057	0.03		0.0141	0.0078		0.0043	0.0015	*
CG7442	V_{WE}	0.0856	0.016	***	0.0089	0.0064		0.0116	0.0247		0.0026	0.0081		0.0019	0.002	
Mmp2	V_{WE}	0.0215	0.0167		0.0043	0.0078		0.0001	0.0233		0.0045	0.0092		0.0004	0.002	
CG1636	V_{AE}	0.111	0.0226	***	0.0156	0.0073	*	0.0621	0.0306	*	0.0082	0.0097		0.0075	0.0028	**
px	V_{AE}	0.0191	0.0251		0.0094	0.0076		0.0593	0.0296	*	0.0489	0.0084	**	0.0079	0.0029	*
mub	V_{AE}	0.0877	0.0179	***	0.0126	0.0082		0.0364	0.0195	*	0.0503	0.0087	**	0.0065	0.0018	**
CG43340	V_{AE}	0.0168	0.016		0.0101	0.0067		0.0301	0.0211		0.0092	0.0079		0.0004	0.0016	
Pka-R1	V_{AE}	0.0903	0.014	***	0.0066	0.0085		0.0215	0.03		0.0034	0.0086		0.0005	0.0017	
CG15343	V_{AE}	0.0243	0.0177		0.0052	0.006		0.0096	0.0234		0.0107	0.0107		0.0008	0.0017	
Ugt	V_{AE}	0.0012	0.0216		0.0035	0.0071		0.0096	0.0262		0.0049	0.0084		0.0011	0.0014	
CG11534	V_{AE}	0.153	0.0175	***	0.0048	0.0063		0.0041	0.0275		0.0263	0.0077	**	0.0068	0.002	**
Sap47	V_{AE}	0.101	0.0184	***	0.0004	0.0064		0.0028	0.0267		0.0062	0.0075		0.0034	0.0022	
Lsp2	V_{AE}	0.0511	0.0214	*	0.0097	0.006		0.0018	0.0241		0.0221	0.0097	*	0.0006	0.0015	

PAPER IV

TEMPORAL DYNAMICS AND EFFECTS OF GENETIC DISTANCE IN GENETIC RESCUE INVESTIGATED IN A *DROSOPHILA MELANOGASTER* MODEL SYSTEM

In review in *Genetica*

CHARLOTTE JENSEN, MICHAEL ØRSTED &
TORSTEN NYGAARD KRISTENSEN

Temporal dynamics and effects of genetic distance in genetic rescue investigated in a *Drosophila melanogaster* model system

Charlotte Jensen[1,†], Michael Ørsted[1,†], and Torsten Nygaard Kristensen[1,2]

[1] *Department of Chemistry and Bioscience, Aalborg University, Aalborg E, Denmark*

[2] *Department of Bioscience, Aarhus University, Aarhus C, Denmark*

[†] *These authors contributed equally (shared first authorship)*

ABSTRACT

A rapidly increasing number of species experience population fragmentation and suffer from ecological and genetic consequences of small population sizes. Inbreeding and loss of genetic variation can decrease the ability to adapt to altered environmental conditions through evolutionary changes and can cause inbreeding depression. One solution to these genetic problems is the implementation of genetic rescue, which re-establishes gene flow between separated populations. This management strategy has proven beneficial in several conservation projects. In this study, we conducted highly replicated interpopulation crosses between isogenic Drosophila melanogaster lines grouped in two genetic distance groups to study the effect of genetic divergence between populations on the expression of heterosis in two fitness components; starvation resistance and reproductive output. We further investigated the temporal effects of outcrossing by investigating the fitness consequences in both F_1- and the F_3-generations. The results provided clear evidence for the beneficial effects of genetic rescue as high fitness enhancements were observed in hybrid offspring compared to parental lines, especially for reproductive output. However, the level of heterosis declined from the F_1- to the F_3-generation, likely due to loss of heterozygosity, and disruption of co-adapted gene complexes. Generally, genetic distance did not have strong impact on the level of heterosis detected, although there were exceptions to this pattern. Overall, our results show that genetic rescue can have very strong positive fitness consequences for genetically depauperate populations and they thereby provide experimental evidence for the great potential for genetic rescue to be used actively in management of fragmented small populations.

Keywords: Genetic distance, heterosis, genetic rescue, temporal effects, starvation resistance, egg productivity, *Drosophila melanogaster*.

INTRODUCTION

Destruction or deterioration of habitats, climate change, and other abiotic and biotic factors currently cause a rapid increase in the number of threatened and fragmented populations (Baillie *et al.* 2004). Natural gene flow is typically reduced between these fragmented populations and their survival is impaired for numerus stochastic and deterministic reasons. The genetic consequences of declining populations, e.g. due to genetic bottlenecks, include elevated risk of inbreeding depression, erosion of genetic variability and thereby loss of evolutionary potential, less efficient selection and a higher probability of accumulating deleterious alleles (Amos & Balmford 2001).

To reduce the extinction risk in small and genetically deteriorated populations suffering from inbreeding depression and low evolutionary potential, conservation management is typically necessary. Some approaches have already been implemented in the conservation of endangered species including 1) e*x situ* conservation in e.g. zoos, which may generate and amplify source populations of endangered species, 2) exploitation of natural selection's ability to reduce the genetic load created by inbreeding, i.e. purging the recessive detrimental alleles exposed by inbreeding, and 3) re-establishment of gene flow among fragmented populations (translocations). The possible result of translocations is an increase in population fitness owing to immigration of new alleles - a phenomenon termed genetic rescue (Whiteley *et al.* 2015). Genetic rescue can reduce extinction risk in small, inbred, and endangered populations by reducing rates of inbreeding and the level of inbreeding depression and by introducing new genetic variation increasing evolutionary potential (Tallmon *et al.* 2004; Hoffmann *et al.* 2017). This approach has been demonstrated to be successful in numerous cases. For instance, one study showed that the immigration of individuals to a small population of red flour beetles (*Tribolium castaneum*) reduced the risk of extinction and improved the long-term survival by reducing inbreeding depression enabling population growth (Hufbauer *et al.* 2015). Other well-known examples are genetic rescue of the Florida panther (*Puma concolor*) (Pimm *et al.* 2006; Johnson *et al.* 2010) and a Scandinavian population of grey wolves (*Canis lupus*) (Vilà *et al.* 2003).

Despite success stories, concerns have been raised relating to the risk of outbreeding depression caused by introducing new genetic material into a donor population, which can lead to a reduction in population fitness (Tallmon *et al.* 2004; Frankham *et al.* 2011; Hedrick & Garcia-Dorado 2016). Outbreeding depression can be caused by local adaptive differences between immigrants and the local population (extrinsic outbreeding depression) (Allendorf *et al.* 2001; Edmands *et al.* 2003). A second reason for the potential harmful effects of hybridization is due to genetic incompatibilities between two populations. With different co-adapted gene complexes between loci (consisting of tightly linked genes), the risk of disrupting beneficial

interactions between genes increases, causing a reduced fitness of the hybrid offspring population (intrinsic outbreeding depression) (Templeton *et al.* 1986; Allendorf *et al.* 2001). These harmful effects are often not observed before the F_2 or later generations, when recombination breaks up favourable gene combinations (Montalvo & Ellstrand 2001; Edmands *et al.* 2003).

As illustrated, genetic rescue can be an efficient management tool in conservation genetics. However, outbreeding depression is a risk, and little is known about trait specific and long-term consequences of genetic rescue. Further, to gain the most heterosis, i.e. improved performance of hybrid offspring, and thereby increase population fitness, a suitable donor population must be identified in genetic rescue projects. It is therefore relevant to investigate whether e.g. genetic distance between fragmented populations is a good predictor of fitness benefits observed when securing gene flow between these populations. Previous studies investigating the association between heterosis and genetic distance between donor and recipient population have been far from concluding. Some studies have showed a positive correlation (Goddard & Ahmed 1982; Graml & Pirchner 1984; Mohamed & Pirchner 1998; Pandey *et al.* 2015), while others found no association (Geleta *et al.* 2004; Singh & Singh 2004; Teklewold & Becker 2006; Kawamura *et al.* 2016). Some authors have suggested that future studies calculating genetic distance based on a larger number of genetic markers than possible at the time, would provide more accurate measures of genetic distance and thereby better reveal the potential of using genetic distance as a predictor of the amount of heterosis expected in genetic rescue projects (Ehiobu *et al.* 1990; Mohamed & Pirchner 1998). Despite novel possibilities to investigate thousands of markers in model as well as non-model organisms, little progress has been made on using such data to guide management decisions in conservation genetics.

In this study we use *D. melanogaster* as a model organism, to investigate the effect of genetic distance (GD) between donor and recipient populations on levels of heterosis in two fitness related traits; starvation resistance (SR) and reproductive output (RO). Because of the importance of persistent genetic rescue effects, we measured the two traits in both F_1 and F_3. We used flies from the *Drosophila melanogaster* Genetic Reference Panel (DGRP), which consists of a publicly available collection of 192 different completely inbred lines that have all been sequenced (Mackay *et al.* 2012; Huang *et al.* 2014). Thus, the genetic distance estimates in our study are based on 1,725,755 single nucleotide polymorphism (SNP) markers (Edwards *et al.* 2016). The experiments were designed to simulate populations suffering from inbreeding, loss of genetic variation and at risk of extinction. To represent the hypothetically genetically impaired populations in need for genetic management, five DGRP lines with low fitness were selected based on data from previous studies investigating the performance of the DGRP lines in several fitness components. Six DGRP lines were selected to constitute genetically rather

similar populations, and six additional DGRP lines were selected to constitute genetically distant populations, to each of the five threatened populations. Results revealed strong heterosis with highly trait specific long-term benefits and little effect of genetic distance on the amount of expressed heterosis.

MATERIALS AND METHODS

DGRP line selection

To identify the five DGRP lines with lowest fitness representing endangered model populations in need of genetic rescue, 141 inbred DGRP lines (expected F≈0.986 (Falconer & Mackay 1996)) from which phenotypic information could be obtained, were ranked according to their performance in five chosen phenotypes: paraquat resistance (Weber *et al.* 2012), starvation resistance and chill-coma recovery time (CCRT) (Mackay *et al.* 2012), negative geotaxis (Jordan *et al.* 2012), and longevity (Ivanov *et al.* 2015). For all, except CCRT, higher value ranked highest. The five DGRP lines with the lowest average rank were chosen to constitute endangered and genetically impaired populations (subsequently referred to as 'model populations'). For each of these five model populations, an additional 12 better performing lines were identified: six genetically similar (termed 'Short' GD) and six genetically different lines (termed 'Long' GD; see 'Genetic distance measures'), resulting in 60 crosses in total. Due to several of the lines recurring in more than one cross, the 60 crosses were based on 41 inbred DGRP lines (Supplementary Table S1). Some lines and crosses without complete data for both generations and both traits ended up being discarded.

Genetic distance measures

Genotypes were acquired from whole genome sequences using an integrative genotyping procedure (Huang *et al.* 2014). All genomic analyses were based on 1,725,755 SNPs across five chromosome arms (Edwards *et al.* 2016). The genetically similar and distant DGRP lines were chosen on the basis of genetic distance measures obtained from a Genomic Relationship Matrix (G-matrix; Supplementary Table S2) containing the genetic relationship between DGRP lines based on differences in segregating SNPs. The G-matrix is calculated from a W-matrix, which is a scaled genotype matrix, with rows containing DGRP lines and the columns containing SNPs: $G=(W\times W')/m$, where W' is the transposed W-matrix and m is the number of SNP markers. For more details see (VanRaden 2008; Edwards *et al.* 2016). Average GD based on the G-matrix for each model population can be seen in Supplementary Fig. S1)

To obtain a measure of genetic distance comparable to other studies we also computed Nei's pairwise F_{ST} (Nei 1973) between pairs of populations of different GD using the R-packages '*adegenet*' (Jombart & Ahmed 2011) and '*hierfstat*' (Goudet & Jombart 2015). Heterozygosities (<2% segregating sites) were weighted by group sizes. For each of the model populations, we considered the model population DGRP line together with the genetically similar DGRP lines (short GD based on the G-matrix) as one population, and the genetically distant DGRP lines (long GD based on the G-matrix) as another population (Supplementary Fig. S2). For each pairwise comparison, F_{ST} was calculated as the average of five random samplings (25 in total) of 100K SNPs across the entire genome (standard errors within populations were below 0.001). For every sampling of SNPs, these F_{ST} estimates were furthermore compared to F_{ST} estimates based on the same 100K SNPs calculated from 25 randomly sampled sets of 20 DGRP lines (10 in each of two populations, from the total of 205), to test if the F_{ST} values between our selected populations were greater than by chance. To visualise the genetic differences and subsequent clustering of short and long GD populations in dendrograms, we performed a hierarchical cluster analysis (HCA) by Ward's method on the number of loci for which individual DGRP lines differ across 30K randomly sampled SNPs using the R package '*ape*' (Paradis *et al.* 2004). (Supplementary Fig. S3).

Crossing design

The DGRP lines were obtained from the Bloomington *Drosophila* Stock Center (NIH 337 P40OD018537). All lines were maintained at 23°C, 50% relative humidity, and a 12:12 h L:D photoperiod on a standard *Drosophila* medium (16 g/L agar, 30 g/L oatmeal, 40 g/L sugar, 60 g/L dry yeast, 12 mL/L nipagen solution, 1 mL/L acetic acid). Vials were checked every 8 h, and emerged flies were sexed under light CO_2-anaesthesia and males and females kept in separate vials for max 24 h prior to setting up the crosses. To set up crosses, 10 males from one DGRP line and 10 virgin females from another were placed in a vial to generate the F_1-generation (see DGRP genotypes in Supplementary Table S1). Both males and females were 16±8 h old when initiating the crosses. The parental lines were maintained at an equal density with 10 males and 10 females from the same DGRP line. The same number of crosses was set up with males and females stemming from the model populations. One-way ANOVAs showed no effect of whether males or females originated from the weaker model populations ($F_{(1,110)}$=0.472; p=0.49 and $F_{(1,110)}$=0.613, p=0.44, for SR and RO, respectively). To produce the F_3, approximately 100 flies from the RO test (F_2) from each cross were distributed to two 100 mL bottles containing 25 mL medium for 24 hours, and then transferred to a new set of bottles for an additional 24 hours. The emerging flies were tested using the same procedure as for the F_1-generation.

Phenotypic assessments

Starvation resistance of parental lines and crosses was measured as total survival time (hours) in vials containing a 4 mL 2% agar solution. The vials were kept at 23°C, 50% RH, and at a 12:12 h L:D photoperiod. After collection, the flies had a 48-hour period to recover from the CO_2-anaesthesia, before starting the starvation test. Hereafter one individual male was transferred to each of 10 vials. The flies were scored every 8 hours until all flies had died.

The productivity of parental lines and crosses was recorded as the total number of hatched offspring produced by a pair of flies. One male and one virgin female from the F_1 generation were transferred to each of 13 vials containing 7 mL medium, and after 72 h they were transferred to a new vial. Each pair laid eggs on three vials during three 72 h periods for a total of nine days. The vials from each ovipositing period were left for 15 days after the initial ovipositing before being frozen, after which the emerged flies from each vial were counted. For some vials, not all flies survived the entire nine-day period. To account for this, the productivity measure was therefore recalculated to our reproductive output (RO) measure (in flies per day) by dividing the total sum of flies across the three periods with the number of days from which data could be obtained (see Supplementary methods).

Statistical analysis

Data on both traits met assumptions of normality, and we found no evidence of heteroscedacity in either trait (Non-constant Variance Score Test (NCV-test); SR: χ^2 (1,N=114)=3.67; p=0.155), and RO: χ^2 (1,N=114)=0.35, p=0.554). To test for heterosis in both F_1 and F_3, two different measures were calculated. Heterosis was calculated as both mid-parent heterosis (MPH) and best-parent heterosis (BPH). MPH, which is the superiority of the hybrid offspring compared to the mid-parental value (MP), was computed as:

$$\text{MPH \%} = \frac{(\text{F}_\text{i} - \text{MP})}{\text{MP}} \text{x } 100$$

where F_i is the mean value of the hybrid individuals in the i'th generation, produced in each cross population, and MP is the mean value of the two parents in each cross population; $(P_1+P_2)/2$ (Solieman *et al.* 2013).

BPH, which is the superiority of the hybrid offspring compared to the best-parental value (BP), was computed as:

$$\text{BPH \%} = \frac{(\text{F}_i - \text{BP})}{\text{BP}} \text{x } 100$$

where BP is the value of the best performing parent in the particular cross (Gixhari & Sulovari 2010). To test whether the levels of heterosis were significant, i.e. whether F_i offspring value was higher than the mid-parental value of the two parental lines (MPH), we used a one-tailed one-sample Wilcoxon signed rank test with individual

F_i values against the MP. To test whether the offspring outperformed the best parent (BPH) we used a one-tailed two-sample Mann-Whitney *U*-test with individual F_i offspring values against individual best-parent values (BP). To test whether heterosis was significantly different in F_1 and F_3, we used a two-sample paired Wilcoxon signed rank test, as the levels of heterosis in a given cross in the two generations are not independent.

The nature of inheritance was also estimated by the potence ratio (PR) (Solieman et al., 2013; Stuber et al., 1987), which is based on the ratio of dominance to additive parameters (d/a):

$$\mathrm{PR} = \frac{\mathrm{d}}{\mathrm{a}} = \frac{\mathrm{F}_i - \mathrm{MP}}{\frac{1}{2}\mathrm{x}\,(\mathrm{P}_2 - \mathrm{P}_1)}$$

where d is the difference between the hybrid offspring and mid-parent value, and a is half the difference between the two parents (P_2 is the mean of the higher parent and P_1 is the mean of the lower parent). The PR value is then evaluated as follows (Solieman et al., 2013): PR=0 means no dominance (additive); PR between -1 and 1 (except 0) means partial dominance; PR=+1 means complete dominance; PR above 1 or below -1 means overdominance. A positive or negative PR specifies the direction of dominance towards one of the parents (Solieman *et al.* 2013).

In order to test the associations of heterosis and heritability of the traits, we calculated narrow-sense heritabilities (h^2) based on the slope of parent-offspring regressions in both F_1 and F_3 (Falconer & Mackay 1996). We used line mean offspring values that were regressed against mean mid-parental values for both the female and male parental lines.

We investigated the effect of GD on both forms of heterosis, as well as the differences across generations and between traits, we performed an initial ANOVA with effects: GD, generation and trait. Separate ANOVAs for each trait were also performed including the effects: GD, generation and model population. Lastly, we performed ANOVAs within each model population with the effects: GD, generation and donor population to investigate model population specific GD effects, and cross-generational heterosis. We performed sequential AIC based model selection (both backwards and forwards) using *F*-tests to find the most parsimonious model and to obtain *p* values for the retained predictors. In some cases, linear regression t-tests were performed on the reduced models to test the difference of e.g. interactions between predictors. The model reductions of predictors were halted if they were part of a significant interaction. We used Tukey's HSD post-hoc test for multiple comparison testing. All statistical analyses were performed in R (R Core Team 2017).

RESULTS

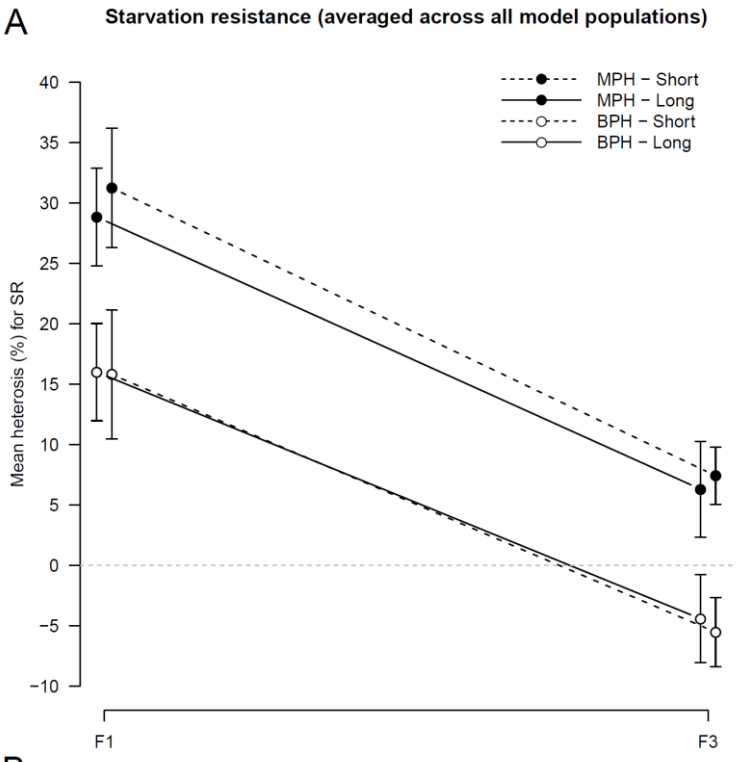
In total, 39 parental lines and 56 hybrid crosses were assessed for starvation resistance (SR) and reproductive output (RO). Mean values for parents, F_1 and F_3 offspring, mid- and, best-parent heterosis, and potence ratio (PR) can be seen in Supplementary Table S1. F_{ST} values between genetically close and more distant donor and recipient (model) populations ranged between 0.100 and 0.134, all of which were significantly different from the 0.057 for the randomly sampled populations ($t_{(26)}$=-21.94; p<0.001; Supplementary Fig. S2).

Highest expression of heterosis for reproductive output

Generally, a high occurrence of heterosis was observed in both traits (Fig. 1, Table 1). For SR, the number of crosses showing significant MPH was 45, while 32 showed significant BPH in F_1. In F_3, these numbers were 22 and 11 for MPH and BPH, respectively. For RO, almost all crosses showed significant MPH and BPH (55 and 50 crosses, respectively) in F_1. In F_3, 46 and 32 crosses exhibited significant MPH and BPH, respectively. Additionally, we observed much higher heterosis in RO compared to SR in both F_1 and F_3 (Table 1); average MPH was 30% in F_1 and 7% in F_3 for SR (Fig. 1A), compared to 211% in F_1 and 132% in F_3 for RO (Fig. 1B). BPH was similarly higher for RO in both generations, in fact for SR, BPH was on average negative in F_3 (Fig. 1A). The h^2 estimates (±SE) for SR were 0.58±0.17 in F_1 and 0.62±0.14 in F_3, while for RO they were 0.70±0.24 in F_1 and 0.69±0.24 in F_3. Thus, we found no significant difference in h^2 estimates between generations ($t_{(108)}$=0.18; p=0.86 for SR, and $t_{(108)}$=-0.04; p=0.97), nor between the two traits ($t_{(216)}$=0.97; p=0.34). We observed significant negative correlations between the mid-parental values (MP) and the amount of heterosis for both traits and both generations, as well as significant negative correlations between lowest parental values (LP) and the amount of heterosis for RO in both generations and for SR in F_1 (Supplementary Table S3).

Trait specific declines in heterosis from F_1 to F_3

A three-way ANOVA with effects: trait, generation, and GD confirmed the significant difference in the amount of MPH and BPH in the two traits, and also that there was a significant effect of generation (Table 2A), with heterosis generally being higher in F_1. Interestingly, we also found a significant interaction between trait and generation for both types of heterosis, suggesting that the decline in heterosis across generations is dependent on the trait. When comparing the slopes for both groups across generations, RO showed a steeper decline than SR in both MPH and BPH ($t_{(220)}$=-2.15; p=0.033, and $t_{(220)}$=-2.13; p=0.034, respectively).


A
Starvation resistance (averaged across all model populations)
MPH − Short
MPH − Long
BPH − Short
BPH − Long
Mean heterosis (%) for SR
40
35
30
25
20
15
10
5
0
−5
−10
F1
F3

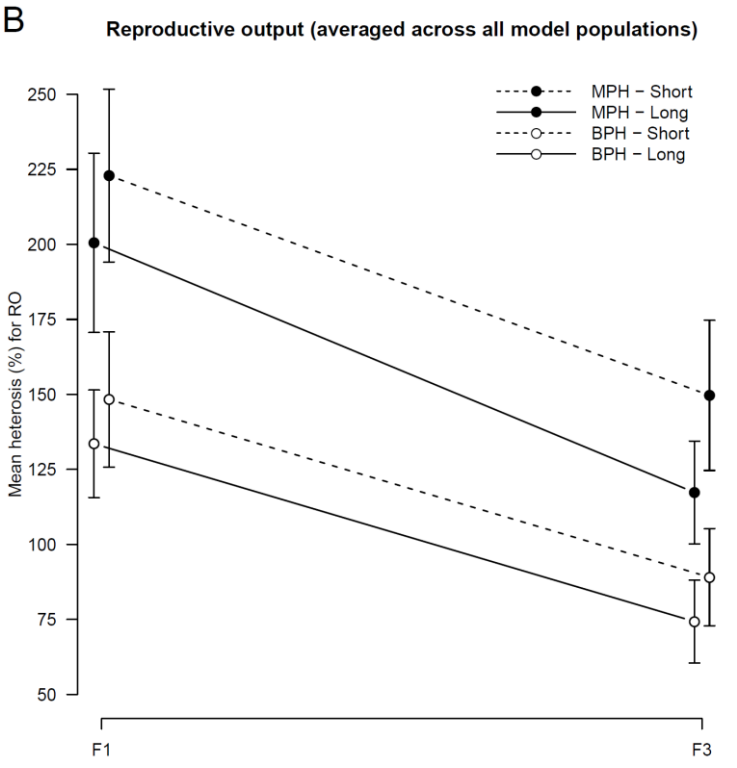

B
Reproductive output (averaged across all model populations)
MPH − Short
MPH − Long
BPH − Short
BPH − Long
Mean heterosis (%) for RO
250
225
200
175
150
125
100
75
50
F1
F3

Fig. 1 Mean mid- and best-parent heterosis for the two genetic distance groups across generations. Mean mid-parent heterosis (MPH) and best-parent heterosis (BPH) for the two genetic distance (GD) groups 'Short' and 'Long', i.e. genetically similar or different. Values are averaged across all model populations in the two generations: F_1 and F_3 for (a) starvation resistance (SR) and (b) reproductive output (RO). MPH is shown in black circles, and BPH in white circles. The cross-generational effect is shown in solid lines for GD group 'Long' and in dashed lines for GD group 'Short'. Error bars represent SE (n=30 for long GD, and n=26 for short GD). Asterisks designate significant difference between F_1 and F_3 (based on two-sample paired Wilcoxon signed rank test, $p<0.05$).

Table 1. Summary of effects of genetic distance across generations. Mid-parental values (MP), hybrid offspring values (Offspring), mid-parent heterosis (MPH), best-parent heterosis (BPH) and potence ratio (PR) in F_1 and F_3 for the two genetic distance groups separately (Short and Long) and combined (Comb.) summarized across all 56 DGRP crosses for starvation resistance (top half) and reproductive output (lower half). Values are given as medians and minimum and maximum (range) values in square brackets, except for the combined column where only medians are shown as the range can be derived from the separate genetic distance groups. Additionally, the total number of crosses and the number of crosses showing significant MPH and BPH are shown. For starvation resistance (SR), MP and offspring values are given in hours, while for reproductive output (RO), MP and offspring values are given in number of flies produced per day.

	Generation	F_1			F_3		
	Genetic distance	Short	Long	Comb.	Short	Long	Comb.
Starvation resistance	MP	48.91 [35.60; 54.80]	43.80 [28.44; 60.40]	45.5	48.91 [35.60; 54.80]	43.80 [28.44; 60.40]	45.5
	Offspring	59.33 [41.60; 85.60]	56.40 [38.86; 74.40]	57.6	50.80 [34.40; 62.40]	46.00 [29.60; 65.60]	48.8
	MPH (%)	26.40 [-8.77; 91.01]	32.77 [-14.67; 70.11]	28.87	8.19 [-14.43; 39.41]	3.00 [-29.91; 51.88]	4.1
	BPH (%)	10.56 [-21.79; 72.55]	17.95 [-20.00; 60.38]	12.85	-9.17 [-32.22; 30.77]	-2.87 [-41.43; 31.67]	-5.69
	PR	1.70 [-5.00; 25.70]	2.63 [-2.20; 33.00]	2.28	0.45 [-4.40; 11.00]	0.59 [-8.00; 14.50]	0.52
	Number of crosses	26	30	56	26	30	56
	Crosses MPH	22	23	45	11	11	22
	Crosses BPH	13	19	32	4	7	11
	Generation	F_1			F_3		
	Genetic distance	Short	Long	Comb.	Short	Long	Comb.
Reproductive output	MP	5.30 [2.23; 12.08]	5.58 [1.98; 11.67]	5.51	5.30 [2.23; 12.08]	5.58 [1.98; 11.67]	5.51
	Offspring	15.85 [5.24; 24.44]	15.94 [6.98; 23.48]	15.91	13.41 [4.19; 20.56]	10.43 [5.15; 20.42]	11.35
	MPH (%)	180.87 [-5.70; 493.97]	163.37 [36.87; 872.37]	171.69	118.62 [-7.13; 467.89]	86.99 [-28.40; 361.63]	98.94
	BPH (%)	130.45 [-16.09; 472.80]	100.41 [25.55; 453.60]	112.9	63.91 [-36.12; 274.60]	57.50 [-45.87; 256.30]	59.58
	PR	6.25 [-0.46; 155.00]	8.22 [1.86; 89.40]	7.5	4.64 [-0.16; 95.90]	4.14 [-0.88; 116.00]	4.42
	Number of crosses	26	30	56	26	30	56
	Crosses MPH	25	30	55	21	25	46
	Crosses BPH	23	27	50	17	15	32

Table 2. Effects of trait, genetic distance, generation and model population. Results of ANOVAs for both traits combined (a), with generation (Gen.), Trait, and genetic distance (GD) for mid-parent heterosis (MPH), and best-parent heterosis (BPH). Because of clear trait specificity, separate ANOVAs for each trait were also performed including the effects: GD, Gen. and model population (ModelPop.). The results of these are shown for (b) starvation resistance (SR), and (c) reproductive output (RO). Sequential AIC based model selection (both backwards and forwards) was performed, using *F*-tests to find the most parsimonious model and to obtain *p*-values for the retained predictors. The model reductions of predictors were halted if they were part of a significant interaction. Sum of squares (SS), degrees of freedom (Df), mean squares (MS), *F*-values, and *p*-values are shown for each retained predictor. Asterisks denote significance level: *** $p<0.001$; ** $p<0.01$; * $p<0.05$; · $p<0.1$. A '-' in a row means the predictor or interaction were retained in the model for MPH or BPH but not in the other.

a (Both traits)			MPH						BPH			
Parameter	SS	Df	MS	*F*	*p*		SS	Df	MS	*F*	*p*	
Gen.	144854	1	144854	15.565	<0.001	***	89774	1	89774	19.704	<0.001	***
Trait	1315068	1	1315068	141.309	<0.001	***	620807	1	620807	136.259	<0.001	***
Gen.*Trait	43024	1	43024	4.623	0.033	*	20632	1	20632	4.528	0.034	*
Residuals	2047389	220	9306				1002336	220	4556			
b (SR)												
Parameter	SS	Df	MS	*F*	*p*		SS	Df	MS	*F*	*p*	
GD			-				12	1	12	0.033	0.856	
Gen.	14995	1	14995	44.719	<0.001	***	12165	1	12165	33.831	<0.001	***
ModelPop.	10147	4	2537	7.565	<0.001	***	8102	4	2025	5.632	<0.001	***
GD*ModelPop.			-				3764	4	941	2.617	0.040	*
Gen.*ModelPop.	3218	4	804	2.399	0.055	·	2716	4	679	1.888	0.119	
Residuals	34202	102	335				34881	97				
c (RO)												
Parameter	SS	Df	MS	*F*	*p*		SS	Df	MS	*F*	*p*	
Gen.	172883	1	172883	11.362	0.001	**	98241	1	98241	14.511	<0.001	***
ModelPop.	386961	4	96740	6.358	<0.001	***	235257	4	58814	8.688	<0.001	***
Residuals	1612861	106	15216				717606	106	6770			

Model population specific effects of genetic distance

Because heterosis was trait specific, the two traits were subsequently analysed in two separate three-way ANOVAs (Table 2B-C), with effects: GD, generation and model population, to investigate the importance of GD when selecting populations for genetic rescue projects, and to examine model population specificity as well. We found no overall significant difference in the magnitude of heterosis between the two groups of GD (Long and Short) for neither MPH or for BPH in the two traits. The five model populations did however display very different levels of both kinds of heterosis (Fig. 2), confirmed by highly significant effects of model population (Table 2B-C). This prompted separate analyses for each model population (Table 3).

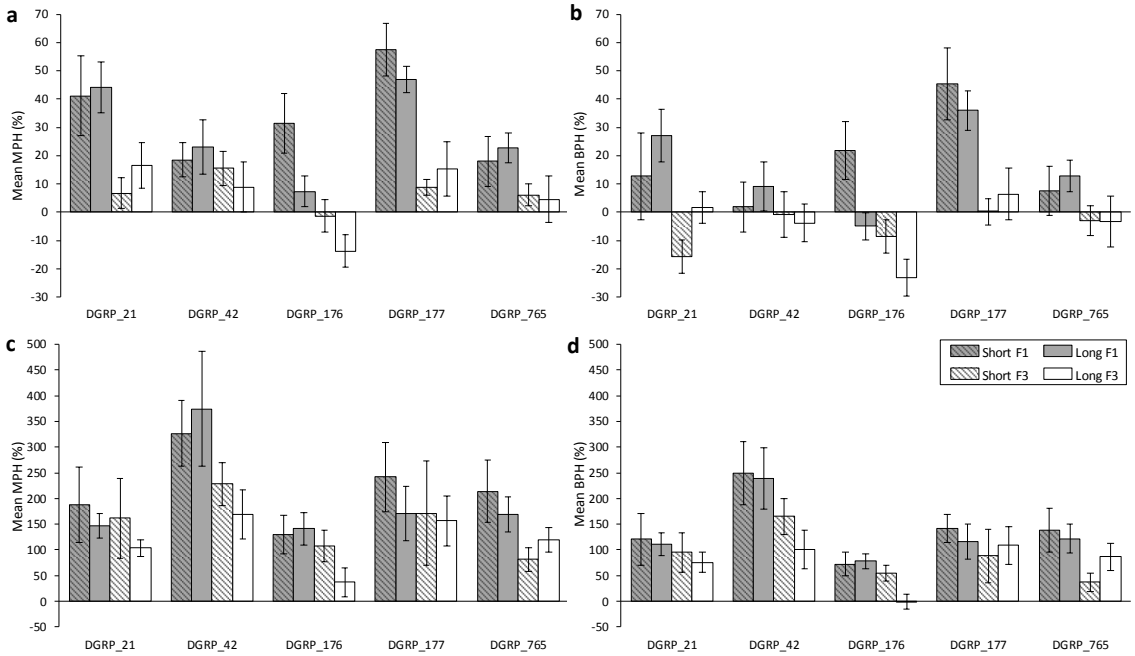


Fig. 2 Population specific mid- and best-parent heterosis for the two genetic distance groups across generations. Population specific mid-parent heterosis (MPH) and best-parent heterosis (BPH) for the two genetic distance (GD) groups 'Short' and 'Long', i.e. genetically similar or different, in the two generations: F_1 and F_3. Each panel shows results for the five DGRP model populations: DGRP_42, DGRP_176, DGRP_765, DGRP_21, and DGRP_177. Values are averaged across the donor populations to each of the model populations for (a) MPH for starvation resistance (SR), (b) BPH for SR, (c) MPH for reproductive output (RO), and (d) BPH for RO. The Short GD group is shown as hatched bars, and the Long GD group is shown as open bars. Data from F_1 is shown in grey bars, and data from F_3 is shown in white bars. Error bars represent SE (n=5-6 donor populations).

Table 3. Model population specific effects of genetic distance, generation and donor population. Results of ANOVAs for model effects: genetic distance (GD), generation (Gen.) and donor population (DonorPop.) for mid-parent heterosis (MPH), and best-parent heterosis (BPH) for (a) starvation resistance (SR) and (b) reproductive output (RO). Separate models were used for each DGRP model population (Model Pop.): DGRP_42, DGRP_176, DGRP_765, DGRP_21, and DGRP_177. Only retained predictors are shown after sequential AIC based model selection (both backwards and forwards) was performed, using *F*-tests to find the most parsimonious model and to obtain *P* values for the retained predictors. The model reductions of predictors were halted if they were part of a significant interaction. Donor population was not included in interactions, because each donor population were part of only one of the two GD groups. Sum of squares (SS), degrees of freedom (Df), mean squares (MS), *F*-, and *P* values are shown for each retained predictor. Asterisks denote significance level: *** $p<0.001$; ** $p<0.01$; * $p<0.05$; · $p<0.1$. A '-' in a row means the predictor or interaction were retained in the model for MPH or BPH but not in the other.

a (SR)		MPH						BPH					
Model Pop.	Source	SS	Df	MS	F	P		SS	Df	MS	F	P	
DGRP_42	GD	6	1	6	0.070	0.797		30	1	30	0.424	0.530	
	Gen.	438	1	438	5.083	0.048	*	368	1	368	5.225	0.045	*
	DonorPop.	6488	10	649	7.530	0.002	**	7081	10	708	10.053	<0.001	***
	GD*Gen.	188	1	188	2.184	0.170		159	1	159	2.254	0.164	
	Residuals	862	10	86				704	10	70			
DGRP_176	GD	1832	1	1832	23.575	<0.001	***	2339	1	2339	40.552	<0.001	***
	Gen.	3857	1	3857	49.648	<0.001	***	3099	1	3099	53.726	<0.001	***
	DonorPop.	4024	9	447	5.755	0.008	**	4214	9	468	8.117	0.002	**
	GD*Gen.	186	1	186	2.395	0.156		202	1	202	3.509	0.094	·
	Residuals	699	9	78				519	9	58			
DGRP_765	Gen.	1353	1	1353	5.195	0.033	*	1067	1	1067	4.718	0.053	·
	DonorPop.			-				4082	11	371	1.641	0.212	
	Residuals	5728	22	260				2488	11	226			
DGRP_21	Gen.	5184	1	5184	16.055	0.002	**	3917	1	3917	14.676	0.003	**
	DonorPop.	6094	10	609	1.887	0.166		7674	10	767	2.875	0.055	·
	Residuals	3229	10	323				2669	10	267			
DGRP_177	Gen.	7381	1	7381	27.300	<0.001	***	6431	1	6431	17.845	<0.001	***
	Residuals	4866	18	270				6487	18	360			

b (RO)		MPH						BPH					
Model Pop.	Source	SS	Df	MS	F	P		SS	Df	MS	F	P	
DGRP_42	GD	170	1	170	0.016	0.900		8235	1	8235	1.963	0.192	
	Gen.	137952	1	137952	13.357	0.004	**	74218	1	74218	17.687	0.002	**
	DonorPop.	520291	10	52029	5.038	0.009	**	257120	10	25712	6.127	0.004	**
	GD*Gen.	17258	1	17258	1.671	0.225		4272	1	4272	1.018	0.337	
	Residuals	103279	10	10328				41962	10	4196			
DGRP_176	GD	4924	1	4924	3.577	0.091	·	3388	1	3388	4.416	0.065	·
	Gen.	24785	1	24785	18.008	0.002	**	14346	1	14346	18.698	0.002	**
	DonorPop.	89375	9	9931	7.215	0.004	**	20892	9	2321	3.026	0.057	·
	GD*Gen.	9259	1	9259	6.727	0.029	*	5148	1	5148	6.710	0.029	*
	Residuals	12387	9	1376				6905	9	767			
DGRP_765	GD	47	1	47	0.008	0.929		1742	1	1742	0.655	0.437	
	Gen.	49354	1	49354	8.970	0.013	*	27781	1	27781	10.448	0.009	**
	DonorPop.	125162	10	12516	2.275	0.105		83505	10	8351	3.141	0.043	*
	GD*Gen.	10421	1	10421	1.894	0.199		6345	1	6345	2.386	0.153	
	Residuals	55018	10	5502				26590	10	2659			
DGRP_21	Gen.	6887	1	6887	1.186	0.302		5332	1	5332	1.462	0.254	
	DonorPop.	210008	10	21001	3.618	0.027	*	69654	10	6965	1.910	0.161	
	Residuals	58045	10	5804				36470	10	3647			
DGRP_177	Gen.	6652	1	6652	1.406	0.266		3421	1	3421	1.044	0.334	
	DonorPop.	301906	9	33545	7.093	0.004	**	89016	9	9891	3.017	0.058	·
	Residuals	42566	9	4730				29502	9	3278			

In the majority of the model populations, there was a strong effect of generation for both SR and RO (Fig. 2, Table 3), but in a few cases there were no effect of generation (e.g. DGRP_765 in Fig. 2A-B and DGRP_21 and DGRP_177 in Fig. 2C-D), suggesting that some populations sustained a temporally constant level of heterosis. Additionally, we saw significant effects of which donor populations were used to cross with the weaker model populations, especially for RO (Table 3B). This is also suggested by the relatively large standard errors in heterosis measures across model populations with a strong donor population effect (e.g. for DGRP_42 and DGRP_177 in Fig. 2C and Table 3).

In one of the model populations (DGRP_176), a significant proportion of the total variation for both kinds of heterosis in SR was explained by GD (Table 3). Post-hoc Tukey's HSD tests showed that donor populations more closely related (GD=short) to the model population showed higher MPH and BPH than populations more distantly related in both F_1 and F_3 ($p<0.001$) (Figure 2A-B). In the same population, we observed a significant interaction between GD and generation for RO (Table 3), with long GD donor populations showing a significantly greater decline in both MPH and BPH from F_1 to F_3 (Fig. 2C-D) ($t_{(9)}=2.60$; $p=0.029$ and $t_{(9)}=2.59$; $p=0.029$, respectively).

DISCUSSION

The main objectives of this study were to investigate how genetic distance (GD) between populations can affect the magnitude of heterosis and how heterosis varies across generations. For that purpose we used DGRP lines. By using the GDRP resource we can base our estimates of genetic distance on ~1.7 million markers and thereby obtain more accurate estimates of GD than have been done in previous investigations of the effects of genetic distance on heterosis (Goddard & Ahmed 1982; Graml & Pirchner 1984; Ehiobu *et al.* 1990; Mohamed & Pirchner 1998; Geleta *et al.* 2004; Singh & Singh 2004; Teklewold & Becker 2006; Pandey *et al.* 2015; Kawamura *et al.* 2016). Generally, our results reveal that genetic distance have little and inconsistent effect on the level of heterosis and that heterosis is highly trait specific and decrease significantly from generation F_1 to F_3.

Theory predicts that higher genetic divergence and larger deviations in allele frequencies between two populations increase the probability of heterozygosity in F_1. An increased expression of heterosis may therefore be expected between more genetically divergent populations (Mäki-Tanila, 2007; Pandey et al., 2015). On the contrary offspring from highly divergent parents may suffer from outbreeding depression e.g. caused by adaptive differentiation or fixation of chromosomal variants. Predicting when genetic rescue is expected to cause positive fitness effect (heterosis) or outbreeding depression has been attempted and some general guidelines

have been proposed (Lynch 1991; Frankham *et al.* 2011). However, empirical studies investigating the association between GD and the expression of heterosis are far from concluding. Some studies have found a positive correlation between GD and heterosis (Goddard & Ahmed 1982; Graml & Pirchner 1984; Biton *et al.* 2012; Pandey *et al.* 2015), while others have come to different conclusions (Geleta *et al.* 2004; Singh & Singh 2004; Teklewold & Becker 2006; Kawamura *et al.* 2016).

In our study we investigated crosses between populations with F_{ST} values ranging between 0.10-0.13, which is defined as 'moderate' genetic differentiation (Hartl & Clark 2007). Thus within this range of genetic differentiation no general impact of genetic distance were observed. Although we cannot provide general recommendations for populations more or less differentiated than the ones we have investigated we do provide evidence that within this range outbreeding depression is unlikely and genetic distance is not a good predictor of the outcome of genetic rescue (Table 2, Fig. 1). We propose this information is of relevance for endangered domestic and wild populations where F_{ST} estimates comparable to ours are often observed between populations in the need of rescue (Merilä & Crnokrak 2001; Leinonen *et al.* 2008; Stronen *et al.* 2017).

Despite the lack of a general effect of GD on heterosis across populations, we did see an effect within some of the model populations (Fig. 2, Table 3), where offspring from populations with short GD expressed significantly higher MPH and BPH for SR than offspring from long GD populations. This suggests that the effect of GD in a genetic rescue project is likely to be highly population specific and again illustrate that it is difficult to predict the magnitude of heterosis based solely on GD between receiver and donor populations. This is further complicated by interactions between GD and other predictors. In DGRP_176, we observed a significant interaction between GD and generation for RO (Table 3), with long GD donor populations showing a significantly greater decline in both MPH and BPH from F_1 to F_3 (Fig. 2C-D). This suggests that, although GD may not have an effect on the initial level of heterosis, GD can affect the rate at which heterosis is lost in some populations.

In genetic rescue projects long-term consequences are important to evaluate. Based on previous studies, we expected heterosis to be detected in both the F_1- and F_3-generation, but at a higher level in F_1 (Fenster & Galloway 2000; Willi *et al.* 2007; Bijlsma *et al.* 2010; Whiteley *et al.* 2015; Frankham 2016). Significant and high levels of heterosis were generally observed in both generations (Table 1), but with a significant decline in the level of heterosis from the F_1 to the F_3 generation (Fig. 1, Fig. 2, Table 2). This is in accordance with findings in a wide range of organisms including partridge peas (Fenster & Galloway 2000), copepods (Edmands 1999), and song sparrows (Marr *et al.* 2002). This fitness decline might be due to reduced heterozygosity in generations succeeding F_1, whereby the potential for heterosis decreases (Marr et al., 2002; Tallmon et al., 2004). Another reason for the temporal

decline in heterosis could be recombination uncoupling the loci contributing to heterosis by epistatic interactions, disrupting adaptive gene combinations and decreasing the expression of heterosis in the generations following F_1 (Tallmon et al., 2004; Bijlsma et al., 2010). These data clearly illustrate that fitness benefits associated with heterosis might be transient.

Inbreeding depression and heterosis have been suggested by many studies to be trait specific, with life history traits being more affected by inbreeding depression compared to traits less closely associated with fitness (Mäki-Tanila, 2007; Derose & Roff, 2017). Congruent with this, we observed a significantly higher expression of heterosis in RO, which is likely more closely linked to fitness, than SR (Table 1, Fig. 1). Selection is expected to deplete additive genetic variance faster for traits closely associated with fitness suggesting that relatively more additive variance would be segregating for SR compared to RO; i.e. remaining variance for RO will be mainly non-additive (Crnokrak & Roff, 1995). This type of genetic variance is expressed as either dominance, overdominance or epistatic interactions between loci, which are the mechanisms believed to cause heterosis (Solieman et al., 2013). This is confirmed by the calculated PR values, which estimate the ratio between dominance and additive genetic variance (d/a). The PR values was significantly higher for RO than for SR in both F_1 and F_3, (Table 1), and considerably above 1 indicating that the nature of inheritance is mainly governed by overdominance. Interestingly, our results showed a significant interaction between generation and trait (Table 2), which indicated that, despite RO expressing higher heterosis initially, the level of heterosis decreased more rapidly in RO compared to SR from F_1 to F_3. This suggests a trait specific trade-off between early heterosis gains and long-term persistence of heterosis.

We found high h^2 estimates (~0.60 for SR and ~0.70 for RO). Others have found similar estimates for SR (Mackay *et al.* 2012), while the heritability for reproduction varies more in literature depending in part on the metric used for fecundity (Sgro & Hoffmann 1998; Fernández *et al.* 2003; Long *et al.* 2009). Thus, the difference in heterosis between the two traits seemed to be unrelated to heritability, as these were not different between the two traits suggesting that heritability is a poor predictor of trait specific heterosis, consistent with other findings (e.g. Flint-Garcia et al. 2009).

The five model populations investigated in this study displayed very different levels of heterosis (Table 3, Fig. 2), suggesting that heterosis is strongly dependent on the specific population to which donor individuals are translocated. Furthermore, there were population specific interactions, e.g. in a few model populations there were no effect of generation (Fig. 2, Table 3), suggesting that some populations sustained a temporally constant level of heterosis. Additionally, we saw significant effects of which donor population were used to cross with the weaker model populations. Interestingly, we detected significant negative correlations between both mid- and lowest parental values and the level of heterosis (Supplementary Table S3), i.e. poorly

performing parental lines showed higher hybrid vigour. This supports the theory, and the conclusions from other empirical studies, that the potential for heterosis is proportional to the level of inbreeding depression (Bijlsma *et al.* 2010). Furthermore, it suggests that fitness estimates of parental populations might be reliable predictors of heterosis. This is also suggested by other authors, who propose to use other measures than GD to predict heterosis, e.g. parental phenotypic distance (Teklewold & Becker 2006).

In conclusion, our results support genetic rescue as an effective management tool in conservation of threatened populations. We find that consequences of genetic recue are highly population specific and that genetic distance (within the range of genetic distances investigated) is not a good predictor of observed heterosis. As expected heterosis typically decrease from F_1 to F_3 but interestingly this decrease is trait specific and less pronounced for SR compared to RO. The best predictor of heterosis seemed to be mid- and lowest parental values, which correlated with heterosis; i.e. lower fitness in the crossed population led to higher heterosis. However, using such populations in genetic rescue also increase the risk of introducing deleterious alleles into recipient populations.

ACKNOWLEDGEMENTS

We are grateful for laboratory assistance from Helle Blendstrup, Susan M. Hansen, Louise B. Rasmussen and Neda Nasiri Moghadam and to Palle Duun Rohde for advice regarding calculations of genetic distances based on SNP data. This work was supported by the Danish Natural Science Research Council through a Sapere aude stipend to TNK (DFF–4002-00036).

REFERENCES

Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001) The problems with hybrids: setting conservation guidelines. *Trends in Ecology and Evolution*, **16**, 613–622.

Amos W, Balmford A (2001) When does conservation genetics matter? *Heredity*, **87**, 257–265.

Baillie J, Hilton-Taylor C, Stuart SN (2004) *IUCN red list of threatened species: a global species assessment*. IUCN, Gland, Switzerland and Cambridge, UK.

Bijlsma R, Westerhof MDD, Roekx LP, Pen I (2010) Dynamics of genetic rescue in inbred *Drosophila melanogaster* populations. *Conservation Genetics*, **11**, 449–462.

Biton I, Shevtsov S, Ostersetzer O *et al.* (2012) Genetic relationships and hybrid vigour in olive (*Olea europaea* L.) by microsatellites (J Léon, Ed,). *Plant Breeding*, **131**, 767–774.

Crnokrak P, Roff DA (1995) Dominance variance: associations with selection and fitness. *Heredity*, **75**, 530–540.

DeRose MA, Roff DA (1999) A comparison of inbreeding depression in life-history and morphological traits in animals. *Evolution*, **53**, 1288–1292.

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

Edmands S, Timmerman CC, Timmermant CC (2003) Modeling factors affecting the severity of outbreeding depression. *Conservation Biology*, **17**, 883–892.

Edwards SM, Sørensen IF, Sarup P, Mackay TFC, Sørensen P (2016) Genomic prediction for quantitative traits is improved by mapping variants to gene ontology categories in *Drosophila melanogaster*. *Genetics*, **203**, 1871–1883.

Ehiobu NG, Goddard ME, Taylor JF (1990) Prediction of heterosis in crosses between inbred lines of *Drosophila melanogaster*. *TAG: Theoretical and Applied Genetics*, **80**, 321–325.

Falconer DS, Mackay TFC (1996) *Introduction to Quantitative Genetics*. Longman Publishing Group, Harlow, UK.

Fenster CB, Galloway LF (2000) Inbreeding and outbreeding depression in natural populations of *Chamaecrista fasciculata* (Fabaceae). *Conservation Biology*, **14**, 1406–1412.

Fernández J, Rodríguez-Ramilo ST, Pérez-Figueroa A, López-Fanjul C, Caballero A (2003) Lack of nonadditive genetic effects on early fecundity in *Drosophila melanogaster*. *Evolution*, **57**, 558–565.

Flint-Garcia SA, Buckler ES, Tiffin P, Ersoz E, Springer NM (2009) Heterosis is prevalent for multiple traits in diverse maize germplasm. *PLOS ONE*, **4**.

Frankham R (2016) Genetic rescue benefits persist to at least the F3 generation, based on a meta-analysis. *Biological Conservation*, **195**, 33–36.

Frankham R, Ballou JD, Eldridge MDB *et al.* (2011) Predicting the probability of outbreeding depression. *Conservation Biology*, **25**, 465–475.

Geleta LF, Labuschagne MT, Viljoen CD (2004) Relationship between heterosis and genetic distance based on morphological traits and AFLP markers in pepper. *Plant Breeding*, **123**, 467–473.

Gixhari B, Sulovari H (2010) Nature of inheritance and heterosis estimated on some morphological quantitative characters that influence the tobacco yield. *Scientific Study & Research - Biology*, **18**, 46–50.

Goddard ME, Ahmed AM (1982) The use of the genetic distance between cattle breeds to predict the heterosis in crosses. In: *Proceedings of the 2nd World Congress on Genetics applied to Livestock Production, Madrid (Vol. 8)*, pp. 377–382.

Goudet J, Jombart T (2015) hierfstat: estimation and tests of hierarchical F-Statistics. R package version 0.04-22. https://cran.r-project.org/package=hierfstat.

Graml R, Pirchner F (1984) Relation of genetic distance between cattle breeds and heterosis of resulting crosses. *Animal Blood Groups and Biochemical Genetics*, **15**, 173–180.

Hartl DL, Clark AG (2007) *Principles of Population Genetics*. Sinauer Associates, Sunderland, UK.

Hedrick PW, Garcia-Dorado A (2016) Understanding inbreeding depression, purging, and genetic rescue. *Trends in Ecology and Evolution*, **31**, 940–952.

Hoffmann AA, Sgrò CM, Kristensen TN (2017) Revisiting adaptive potential, population size, and conservation. *Trends in Ecology and Evolution*, **32**, 506–517.

Huang W, Massouras A, Inoue Y *et al.* (2014) Natural variation in genome architecture among 205 *Drosophila melanogaster* Genetic Reference Panel lines. *Genome Research*, **24**, 1193–1208.

Hufbauer RA, Szűcs M, Kasyon E *et al.* (2015) Three types of rescue can avert extinction in a changing environment. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 10557–10562.

Ivanov DK, Escott-Price V, Ziehm M *et al.* (2015) Longevity GWAS using the *Drosophila* Genetic Reference Panel. *Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, **70**, 1470–1478.

Johnson WE, Onorato DP, Roelke ME *et al.* (2010) Genetic restoration of the Florida panther. *Science*, **329**, 1641–1645.

Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, **27**, 3070–3071.

Jordan KW, Craver KL, Magwire MM *et al.* (2012) Genome-wide association for sensitivity to chronic oxidative stress in *Drosophila melanogaster* (P Csermely, Ed,). *PLOS ONE*, **7**, e38722.

Kaufman BP, Demerec M (1942) Utilization of sperm by the female *Drosophila melanogaster*. *American Naturalist*, **76**, 445–469.

Kawamura K, Kawanabe T, Shimizu M *et al.* (2016) Genetic distance of inbred lines of Chinese cabbage and its relationship to heterosis. *Plant Gene*, **5**, 1–7.

Leinonen T, O'Hara RB, Cano JM, Merilä J (2008) Comparative studies of quantitative trait and neutral marker diverence: a meta-analysis. *Journal of Evolutionary Biology*, **21**, 1–17.

Long TAF, Miller PM, Stewart AD, Rice WR (2009) Estimating the heritability of female lifetime fecundity in a locally adapted *Drosophila melanogaster* population. *Journal of Evolutionary Biology*, **22**, 637–643.

Lynch M (1991) The genetic interpretation of inbreeding depression and outbreeding depression. *Evolution*, **45**, 622–629.

Mackay TFC, Richards S, Stone EA *et al.* (2012) The *Drosophila melanogaster* Genetic Reference Panel. *Nature*, **482**, 173–178.

Marr AB, Keller LF, Arcese P (2002) Heterosis and outbreeding depression in descendants of natural immigrants to an inbred population of song sparrows (*Melospiza melodia*). *Evolution*, **56**, 131–142.

Merilä J, Crnokrak P (2001) Comparison of genetic differentaition at marker loci and quantitative traits. *Journal of Evolutionary Biology*, **14**, 892–903.

Mohamed SA, Pirchner F (1998) Relationships between genetic distance and dominance and epistatic effects in line-crosses of mice. *Journal of Animal Breeding and Genetics*, **115**, 313–321.

Montalvo AM, Ellstrand NC (2001) Nonlocal transplantation and outbreeding depression in the subshrub *Lotus scoparius* (Fabaceae). *American Journal of Botany*, **88**, 258–269.

Mäki-Tanila A (2007) An overview on quantitative and genomic tools for utilising dominance genetic variation in improving animal production. *Agricultural and Food Science*, **16**, 188–198.

Nei M (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America*, **70**, 3321–3323.

Pandey P, Pandey VR, Kumar A *et al.* (2015) Relationship between heterosis and genetic diversity in Indian pigeonpea [*Cajanus cajan* (L.) Millspaugh] accessions using multivariate cluster analysis and heterotic grouping. *Australian Journal of Crop Science*, **9**, 494–503.

Paradis E, Claude J, Strimmer K (2004) APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, **20**, 289–290.

Pimm SL, Dollar L, Bass OL (2006) The genetic rescue of the Florida panther. *Animal Conservation*, **9**, 115–122.

R Core Team (2017) R: A language and environment for statistical computing. Version 3.4.0. R Foundation for Statistical Computing, Vienna, Austria. http://www.r-project.org/.

Sgro CM, Hoffmann AA (1998) Heritable variation for fecundity in field-collected *Drosophila melanogaster* and their offspring reared under different environmental temperatures. *Evolution*, **52**, 134–143.

Singh S., Singh M (2004) Multivariate analysis in relation to genetic improvement in *Cuphea procumbens*. *Journal of Genetics and Breeding*, **58**, 105–112.

Solieman THI, El-Gabry MAH, Abido AI (2013) Heterosis, potence ratio and correlation of some important characters in tomato (*Solanum lycopersicum* L.). *Scientia Horticulturae*, **150**, 25–30.

Stronen A V., Salmela E, Baldursdottir BK *et al.* (2017) Genetic rescue of an endangered domestic animal through outcrossing with closely related breeds: a case study of the Norwegian Lundehund. *PLOS ONE*, **12**, e0177429.

Stuber CW, Edwards MD, Wendel JF (1987) Molecular-marker-facilitated investigations of quantitative trait loci in maize. II. Factors influencing yield and its component traits. *Crop Science*, **27**, 639–648.

Tallmon DA, Luikart G, Waples R (2004) The alluring simplicity and complex reality of genetic rescue. *Trends in Ecology & Evolution*, **19**, 489–496.

Teklewold A, Becker HC (2006) Comparison of phenotypic and molecular distances to predict heterosis and F1 performance in Ethiopian mustard (*Brassica carinata* A. Braun). *Theoretical and Applied Genetics*, **112**, 752–759.

Templeton AR, Hemmer H, Mace G *et al.* (1986) Local adaptation, coadaptation, and population boundaries. *Zoo Biology*, **5**, 115–125.

VanRaden PM (2008) Efficient methods to compute genomic predictions. *Journal of Dairy Science*, **91**, 4414–4423.

Vilà C, Sundqvist A-KA, Flagstad Ø *et al.* (2003) Rescue of a severely bottlenecked wolf (*Canis lupus*) population by a single immigrant. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 91–97.

Weber AL, Khan GF, Magwire MM *et al.* (2012) Genome-wide association analysis of oxidative stress resistance in *Drosophila melanogaster* (T Flatt, Ed,). *PLoS One*, **7**, e34745.

Whiteley AR, Fitzpatrick SW, Funk WC, Tallmon DA (2015) Genetic rescue to the rescue. *Trends in Ecology & Evolution*, **30**, 42–49.

Willi Y, van Kleunen M, Dietrich S, Fischer M (2007) Genetic rescue persists beyond first-generation outbreeding in small populations of a rare plant. *Proceedings of the Royal Society B: Biological Sciences*, **274**, 2357–2364.

SUPPLEMENTARY INFORMATION FOR PAPER IV

Text S1: Supplementary methods: Reproductive output calculations

Table S1: Mean phenotypic values for starvation resistance and reproductive output averaged for each cross in F_1 and F_3.

Table S2: Genomic Relationship Matrix of the selected DGRP lines. (supplied online: https://goo.gl/t6yPJs)

Table S3: Correlations between mid- or lowest-parental values, and mid-parent heterosis and/or best-parent heterosis.

Figure S1: Genetic distances based on the Genomic Relationship Matrix of the two genetic distance groups for each DGRP model population.

Figure S2: Nei's pairwise F_{ST} between pairs of populations of different genetic distances crossed to the five model populations.

Figure S3: Dendrograms with clustering of DGRP lines used in the study as either genetically similar or distant to each of the five DGRP model populations.

Reproductive output calculations

The productivity of the flies for both the parental DGRP lines and the hybrid offspring were recorded as the total number of hatched offspring produced by a pair of flies consisting of one male and one virgin female over a nine-day period on three separate vials (vials 1-3) each for 72 hours. Thirteen such replicate vials were set up per DGRP line and per hybrid cross, however for some vials not all flies survived the entire nine-day period. To account for this, the productivity measure was therefore recalculated to a reproductive output measure by dividing the total sum of flies across the three periods with the number of days from which data is obtained. Meaning, if e.g. no data was obtained from the third vial (if either the female or both flies were dead prior to transferring to vial III, see regulations below) the total number of flies was divided by six and if both flies survived until being discarded after three periods of 72 hours, the total number of flies was divided by nine. To avoid discarding too much data, we defined some regulations that were based on the fact that a single insemination can provide the female fly with enough sperm to lay fertile eggs for approximately eight days (Kaufman & Demerec 1942):

- If the male was dead before transfer to vial 2, and if vial 1 was without larvae, the replicate was discarded. The male had to be alive for at least 72 hours. If vial 1 had live larvae, all replicate vials were used, as the female was certainly fertilised.
- If the female was dead before transfer to vial 2, the replicate was discarded, as it was impossible to determine for how long the female had been alive on vial 1.
- If the female was dead before transfer to vial 3, only data from vial 1 was used, as it was impossible to determine for how long the female had been alive on vial 2. This was regardless of whether larvae were observed in vial 1.
- If the female was dead before being discarded from vial 3, data from vial 1 and 2 was used.

Supplementary Table S1. Mean phenotypic values for starvation resistance (in hours) and reproductive output (in eggs laid per day) averaged for each cross (ID) in F_1 and F_3. Male parent (♂) and female parent (♀) DGRP genotypes are given for each cross. Gen. indicates in which generation the performance is measured. Note that the parental values were only measured in the first generation. GD designates the genetic distance group to which the given cross belongs (S=short, L=long). M and F designates the male and female parent phenotypic values, respectively, and MP is the average of the two parents. Offspring (O) is the average performance of the F_1 or F_3 offspring. MPH is the mean mid-parent heterosis (in %) and BPH the mean best-parent heterosis (in %). Significant MPH and BPH are designated with asterisks: *** $p<0.001$; ** $p<0.01$; * $p<0.05$; · $p<0.1$, based on one-sided one sample Wilcoxon signed rank tests against the MP value for MPH or one-sided Mann-Whitney U test against the best parent values for BPH. PR shows the mean potence ratio.

Table S1

					Starvation resistance							Reproductive output						
ID	♂	♀	GD	gen	M	F	MP	O	MPH	BPH	PR	M	F	MP	O	MPH	BPH	PR
1	42	357	S	F1	36.4	66.4	51.4	55.2	7.35 *	-16.87	0.25	3.47	0.99	2.23	13.26	493.97 **	281.80 ***	8.89
1	42	357	S	F3	36.4	66.4	51.4	49.6	-3.54	-25.30	-0.12	3.47	0.99	2.23	10.71	379.85 ***	208.40 ***	6.83
2	517	42	S	F1	49.6	36.4	43.0	54.4	26.45 **	9.68	1.73	6.50	3.47	4.99	13.92	179.24 **	114.20 ***	5.90
2	517	42	S	F3	49.6	36.4	43.0	50.4	17.15 **	1.61	1.12	6.50	3.47	4.99	11.32	127.13 **	74.23 **	4.19
3	42	786	S	F1	36.4	62.4	49.4	48.8	-1.26	-21.79	-0.05	3.47	9.50	6.49	14.36	121.38 ***	51.15 *	2.61
3	42	786	S	F3	36.4	62.4	49.4	59.2	19.78 *	-5.13	0.75	3.47	9.50	6.49	14.26	119.84 **	50.10	2.58
4	177	42	S	F1	40.8	36.4	38.6	48.8	26.35 **	19.61 **	4.67	5.39	3.47	4.43	21.21	378.79 **	293.60 ***	17.50
4	177	42	S	F3	40.8	36.4	38.6	44.8	16.00 *	9.80 *	2.84	5.39	3.47	4.43	16.45	271.35 ***	205.30 ***	12.50
5	42	855	S	F1	36.4	61.6	49.0	55.2	12.60 **	-10.39	0.49	3.47	3.34	3.41	13.35	291.85 ***	284.50 ***	153.00
5	42	855	S	F3	36.4	61.6	49.0	51.2	4.44	-16.88	0.17	3.47	3.34	3.41	9.66	183.47 **	178.20 **	95.90
6	491	42	S	F1	41.6	36.4	39.0	54.4	39.41 **	30.77 ***	5.97	3.26	3.47	3.37	19.89	490.92 ***	472.80 ***	155.00
6	491	42	S	F3	41.6	36.4	39.0	54.4	39.41 **	30.77 **	5.97	3.26	3.47	3.37	13.01	286.50 ***	274.60 ***	90.60
7	42	802	L	F1	36.4	36.8	36.6	40.0	9.22	8.70	19.00	3.47	9.53	6.50	18.68	187.31 **	96.01 ***	4.02
7	42	802	L	F3	36.4	36.8	36.6	39.2	7.04	6.52	14.50	3.47	9.53	6.50	8.98	38.20	-5.72	0.82
8	161	42	L	F1	38.4	36.4	37.4	50.4	34.68 **	31.25 **	13.30	4.09	3.47	3.78	16.44	334.54 ***	301.60 ***	40.80
8	161	42	L	F3	38.4	36.4	37.4	38.4	2.61	4.1E-15	1.00	4.09	3.47	3.78	14.58	285.50 **	256.30 ***	34.80
9	42	358	L	F1	36.4	42.4	39.4	53.0	34.44 **	25.00 **	4.56	3.47	4.87	4.17	8.85	112.10 *	81.65	6.69
9	42	358	L	F3	36.4	42.4	39.4	40.8	3.50	-3.77	0.46	3.47	4.87	4.17	7.76	85.95	59.26	5.13
10	350	42	L	F1	64.8	36.4	50.6	56.0	10.62 *	-13.58	0.38	1.69	3.47	2.58	15.00	480.65 ***	332.00 ***	14.00
10	350	42	L	F3	64.8	36.4	50.6	50.4	-0.44	-22.22	-0.02	1.69	3.47	2.58	6.69	159.06 **	92.74 *	4.62
11	42	837	L	F1	36.4	48.0	42.2	38.9	-7.97	-19.05	-0.58	3.47	0.48	1.98	19.22	872.37 **	453.60 ***	11.50
11	42	837	L	F3	36.4	48.0	42.2	37.6	-10.95	-21.67	-0.80	3.47	0.48	1.98	8.69	339.34 *	150.10	4.49
12	409	42	L	F1	20.4	36.4	28.4	44.8	57.50 **	22.93 **	2.04	6.81	3.47	5.14	18.46	259.03 ***	171.00 ***	7.98
12	409	42	L	F3	20.4	36.4	28.4	43.2	51.88 **	18.54 *	1.84	6.81	3.47	5.14	10.61	106.36 *	55.77	3.28
13	176	75	S	F1	56.0	53.6	54.8	85.6	56.20 **	52.86 **	25.70	9.51	5.48	7.50	16.36	118.26 **	71.97 ***	4.39
13	176	75	S	F3	56.0	53.6	54.8	62.4	13.87 *	11.43	6.33	9.51	5.48	7.50	13.82	84.37 *	45.27	3.13
14	761	176	S	F1	52.0	56.0	54.0	54.4	0.74	-2.86	0.20	14.65	9.51	12.08	18.01	49.06 **	22.93 **	2.31
14	761	176	S	F3	52.0	56.0	54.0	48.0	-11.11	-14.29	-3.00	14.65	9.51	12.08	14.58	20.71	-0.45	0.97
15	176	142	S	F1	56.0	38.7	47.3	58.7	23.94	4.76	1.31	9.51	1.40	5.46	13.55	148.28 **	42.43 *	2.00
15	176	142	S	F3	56.0	38.7	47.3	44.8	-5.35	-20.00	-0.29	9.51	1.40	5.46	15.70	187.68 **	65.03 *	2.53
16	703	176	S	F1	42.4	56.0	49.2	76.0	54.47 **	35.71 **	3.94	9.13	9.51	9.32	16.06	72.25 **	68.76 ***	34.90
16	703	176	S	F3	42.4	56.0	49.2	54.4	10.57	-2.86	0.77	9.13	9.51	9.32	16.66	78.72 ***	75.09 **	38.00
17	176	379	S	F1	56.0	52.4	54.2	66.4	22.46 **	18.57 *	6.85	9.51	3.91	6.71	24.44	264.18 ***	156.80 ***	6.32
17	176	379	S	F3	56.0	52.4	54.2	46.4	-14.43	-17.14	-4.40	9.51	3.91	6.71	18.09	169.64 **	90.17 **	4.06
19	176	93	L	F1	56.0	37.6	46.8	48.0	2.56	-14.29	0.13	9.51	7.94	8.73	11.94	36.87 *	25.55	4.09
19	176	93	L	F3	56.0	37.6	46.8	32.8	-29.92	-41.43	-1.52	9.51	7.94	8.73	8.69	-0.44	-8.67	-0.05
20	386	176	L	F1	32.8	56.0	44.4	53.6	20.72 **	-4.29	0.79	13.83	9.51	11.67	23.05	97.49 **	66.68 ***	5.27
20	386	176	L	F3	32.8	56.0	44.4	37.6	-15.32	-32.86	-0.59	13.83	9.51	11.67	14.50	24.24 *	4.85	1.31
21	176	894	L	F1	56.0	49.6	52.8	60.8	15.15 **	8.57	2.50	9.51	2.08	5.80	21.18	265.19 **	122.60 ***	4.14
21	176	894	L	F3	56.0	49.6	52.8	51.2	-3.03	-8.57	-0.50	9.51	2.08	5.80	15.21	162.34 **	59.89	2.53
22	358	176	L	F1	42.4	56.0	49.2	51.2	4.07	-8.57	0.29	4.87	9.51	7.19	19.45	170.48 ***	104.50 ***	5.28
22	358	176	L	F3	42.4	56.0	49.2	39.2	-20.33	-30.00	-1.47	4.87	9.51	7.19	5.15	-28.40	-45.87	-0.88
23	176	350	L	F1	56.0	64.8	60.4	70.4	16.56 **	8.64 *	2.27	9.51	1.69	5.60	14.71	162.47 **	54.61 *	2.33
23	176	350	L	F3	56.0	64.8	60.4	65.6	8.61	1.24	1.18	9.51	1.69	5.60	9.40	67.73 **	-1.20	0.97
24	359	176	L	F1	64.0	56.0	60.0	51.2	-14.67	-20.00	-2.20	11.74	9.51	10.63	23.04	116.86 **	96.32 ***	11.20
24	359	176	L	F3	64.0	56.0	60.0	46.4	-22.67	-27.50	-3.40	11.74	9.51	10.63	10.24	-3.62	-12.75	-0.35
25	765	774	S	F1	46.4	60.0	53.2	64.8	21.81 **	8.00 *	1.71	4.73	16.30	10.52	22.39	112.95 ***	37.38 **	2.05
25	765	774	S	F3	46.4	60.0	53.2	55.2	3.76	-8.00	0.29	4.73	16.30	10.52	19.92	89.44 **	22.22	1.63

Table S1 continued

					Starvation resistance							Reproductive output						
ID	♂	♀	GD	gen	M	F	MP	O	MPH	BPH	PR	M	F	MP	O	MPH	BPH	PR
26	531	765	S	F1	44.8	46.4	45.6	41.6	-8.77	-10.34	-5.00	5.56	4.73	5.14	14.04	172.91 **	152.70 ***	21.60
26	531	765	S	F3	44.8	46.4	45.6	54.4	19.30 **	17.24 **	11.00	5.56	4.73	5.14	8.33	61.93	49.92	7.73
27	765	379	S	F1	46.4	52.4	49.4	73.6	48.92 **	40.34 ***	8.00	4.73	3.91	4.32	15.65	262.29 **	230.70 ***	27.40
27	765	379	S	F3	46.4	52.4	49.4	57.6	16.55	9.83	2.71	4.73	3.91	4.32	6.98	61.69 **	47.58 *	6.45
28	786	765	S	F1	62.4	46.4	54.4	60.0	10.29 **	-3.85	0.70	9.50	4.73	7.12	20.10	182.51 ***	111.60 ***	5.45
28	786	765	S	F3	62.4	46.4	54.4	55.2	1.47	-11.54	0.10	9.50	4.73	7.12	18.79	164.01 ***	97.75 ***	4.89
29	765	142	S	F1	46.4	38.7	42.5	57.6	35.42 **	24.14 **	3.90	4.73	1.40	3.07	17.93	484.66 **	278.90 ***	8.93
29	765	142	S	F3	46.4	38.7	42.5	41.6	-2.19	-10.34	-0.24	4.73	1.40	3.07	6.67	117.40 *	40.90	2.16
30	859	765	S	F1	63.2	46.4	54.8	55.2	0.73 *	-12.66	0.05	12.59	4.73	8.66	14.40	66.24 **	14.36	1.46
30	859	765	S	F3	63.2	46.4	54.8	53.6	-2.19	-15.19	-0.14	12.59	4.73	8.66	8.04	-7.13	-36.12	-0.16
31	765	359	L	F1	46.4	64.0	55.2	74.4	34.78 **	16.25 **	2.18	4.73	11.74	8.23	21.76	164.27 **	85.40 ***	3.86
31	765	359	L	F3	46.4	64.0	55.2	45.6	-17.39	-28.75	-1.09	4.73	11.74	8.23	16.38	98.92 **	39.55	2.33
32	313	765	L	F1	52.8	46.4	49.6	59.2	19.36 **	12.12 **	3.00	6.42	4.73	5.58	15.76	182.59 ***	145.40 ***	12.00
32	313	765	L	F3	52.8	46.4	49.6	65.6	32.26 **	24.24 *	5.00	6.42	4.73	5.58	14.05	151.94 **	118.80 **	10.00
33	765	712	L	F1	46.4	48.0	47.2	62.4	32.20 **	30.00 ***	19.00	4.73	6.44	5.58	10.48	87.67 **	62.82 **	5.74
33	765	712	L	F3	46.4	48.0	47.2	48.8	3.39 *	1.67	2.00	4.73	6.44	5.58	10.25	83.53 **	59.23 **	5.47
34	358	765	L	F1	42.4	46.4	44.4	48.8	9.91	5.17	2.20	4.87	4.73	4.80	7.62	58.80 **	56.54 **	40.70
34	358	765	L	F3	42.4	46.4	44.4	52.0	17.12 **	12.07 *	3.80	4.87	4.73	4.80	8.37	74.29 **	71.80 *	51.40
35	765	350	L	F1	46.4	64.8	55.6	58.7	5.52 *	-9.47	0.33	4.73	1.69	3.21	11.44	255.93 **	141.70 **	5.42
35	765	350	L	F3	46.4	64.8	55.6	44.8	-19.42	-30.86	-1.17	4.73	1.69	3.21	5.98	86.17 *	26.42	1.82
36	161	765	L	F1	38.4	46.4	42.4	56.8	33.96 **	22.41 ***	3.60	4.09	4.73	4.41	16.13	265.55 ***	240.90 ***	36.70
36	161	765	L	F3	38.4	46.4	42.4	47.2	11.32 *	1.72	1.20	4.09	4.73	4.41	14.38	225.84 **	203.80 ***	31.20
37	21	358	S	F1	31.2	42.4	36.8	45.7	24.22 **	7.82	1.59	6.25	4.87	5.56	5.24	-5.70	-16.09	-0.46
37	21	358	S	F3	31.2	42.4	36.8	34.4	-6.52	-18.87	-0.43	6.25	4.87	5.56	8.45	52.07 *	35.31 *	4.20
39	21	790	S	F1	31.2	40.0	35.6	68.0	91.01 **	70.00 **	7.36	6.25	3.44	4.85	20.34	319.77 **	225.60 ***	11.10
39	21	790	S	F3	31.2	40.0	35.6	41.6	16.85 *	4.00	1.36	6.25	3.44	4.85	7.68	58.55	22.98	2.02
40	375	21	S	F1	72.0	31.2	51.6	57.3	11.11 *	-20.37	0.28	5.88	6.25	6.06	18.98	213.04 ***	203.80 ***	70.30
40	375	21	S	F3	72.0	31.2	51.6	48.8	-5.43	-32.22	-0.14	5.88	6.25	6.06	14.05	131.71 **	124.90 ***	43.50
41	21	367	S	F1	31.2	64.0	47.6	60.8	27.73 *	-5.00	0.81	6.25	9.87	8.06	11.70	45.19 *	18.55	2.01
41	21	367	S	F3	31.2	64.0	47.6	52.0	9.24 *	-18.75	0.27	6.25	9.87	8.06	16.03	98.96 **	62.45 **	4.40
42	357	21	S	F1	66.4	31.2	48.8	74.0	51.64 *	11.45	1.43	0.99	6.25	3.62	16.92	367.36 **	170.80 ***	5.06
42	357	21	S	F3	66.4	31.2	48.8	58.4	19.67	-12.05	0.55	0.99	6.25	3.62	20.56	467.89 ***	229.00 ***	6.44
43	21	386	L	F1	31.2	32.8	32.0	41.6	30.00 **	26.83 ***	12.00	6.25	13.83	10.04	20.41	103.32 ***	47.59 ***	2.74
43	21	386	L	F3	31.2	32.8	32.0	29.6	-7.50	-9.76	-3.00	6.25	13.83	10.04	20.42	103.38 **	47.64 *	2.74
44	161	21	L	F1	38.4	31.2	34.8	59.2	70.12 **	54.17 ***	6.78	4.09	6.25	5.17	10.13	95.92 **	62.13 ***	4.60
44	161	21	L	F3	38.4	31.2	34.8	40.8	17.24 *	6.25	1.67	4.09	6.25	5.17	8.93	72.75 *	42.95	3.49
45	21	712	L	F1	31.2	48.0	39.6	62.9	58.73 **	30.95 ***	2.77	6.25	6.44	6.34	14.75	132.54 **	129.10 ***	89.40
45	21	712	L	F3	31.2	48.0	39.6	48.8	23.23 **	1.67	1.10	6.25	6.44	6.34	17.22	171.56 **	167.60 ***	116.00
46	26	21	L	F1	44.8	31.2	38.0	53.6	41.05 **	19.64 **	2.29	9.32	6.25	7.79	23.48	201.55 ***	151.80 ***	10.20
46	26	21	L	F3	44.8	31.2	38.0	56.8	49.47 **	26.79 ***	2.76	9.32	6.25	7.79	17.22	121.20 **	84.71 ***	6.13
47	21	595	L	F1	31.2	51.2	41.2	44.8	8.74	-12.50	0.36	6.25	4.42	5.33	18.06	238.54 **	189.00 ***	13.90
47	21	595	L	F3	31.2	51.2	41.2	48.0	16.51	-6.25	0.68	6.25	4.42	5.33	8.87	66.27 **	41.93 *	3.86
48	93	21	L	F1	37.6	31.2	34.4	53.7	56.15 **	42.86 ***	6.04	7.94	6.25	7.09	14.89	109.88 ***	87.51 ***	9.21
48	93	21	L	F3	37.6	31.2	34.4	34.4	0.00	-8.51	0.00	7.94	6.25	7.09	13.32	87.82 *	67.81	7.36
49	177	142	S	F1	40.8	38.7	39.7	62.4	57.05 **	52.94 ***	21.30	5.39	1.40	3.40	10.19	200.25 **	89.18 *	3.41
49	177	142	S	F3	40.8	38.7	39.7	40.8	2.69	7.1E-15	1.00	5.39	1.40	3.40	4.19	23.46	-22.21	0.40

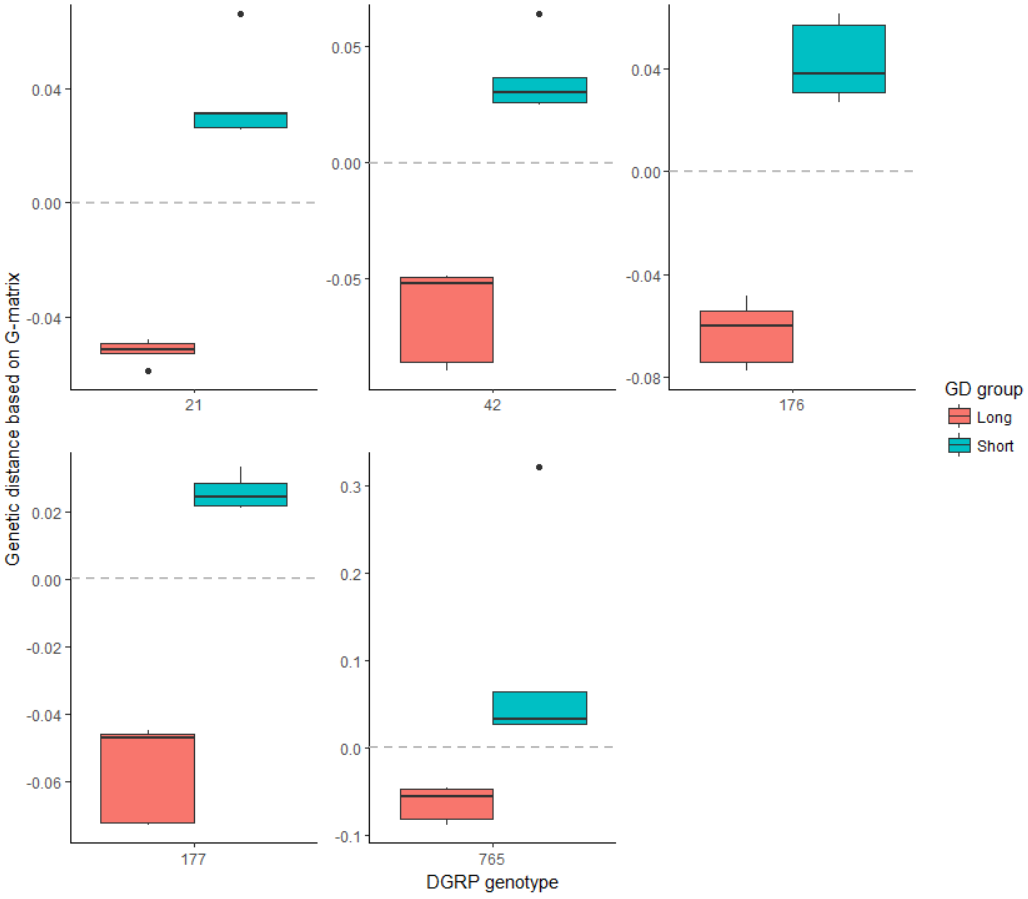
Table S1 continued

					Starvation resistance							Reproductive output						
ID	♂	♀	GD	gen	M	F	MP	O	MPH		BPH		PR	M	F	MP	O	MPH	BPH	PR
51	177	229	S	F1	40.8	37.6	39.2	70.4	79.59	**	72.55	***	19.50	5.39	5.93	5.66	14.50	156.23 ***	144.60 ***	32.90
51	177	229	S	F3	40.8	37.6	39.2	45.6	16.33	**	11.76	*	4.00	5.39	5.93	5.66	9.65	70.51 *	62.78 *	14.90
52	855	177	S	F1	61.6	40.8	51.2	68.8	34.38	**	11.69	*	1.69	3.34	5.39	4.37	11.66	167.04 **	116.30 **	7.12
52	855	177	S	F3	61.6	40.8	51.2	54.9	7.14	*	-10.95		0.35	3.34	5.39	4.37	9.80	124.57 **	81.92 *	5.31
54	887	177	S	F1	50.0	40.8	45.4	72.0	58.59	**	44.00	**	5.78	0.89	5.39	3.14	17.05	443.23 ***	216.40 ***	6.18
54	887	177	S	F3	50.0	40.8	45.4	49.6	9.25	*	-0.80		0.91	0.89	5.39	3.14	17.81	467.43 **	230.50 **	6.52
55	177	712	L	F1	40.8	48.0	44.4	62.4	40.54	**	30.00	***	5.00	5.39	6.44	5.91	10.34	74.92 ***	60.69 **	8.46
55	177	712	L	F3	40.8	48.0	44.4	63.2	42.34	**	31.67	***	5.22	5.39	6.44	5.91	19.66	232.56 ***	205.50 ***	26.30
56	894	177	L	F1	49.6	40.8	45.2	66.4	46.90	**	33.87	***	4.82	2.08	5.39	3.74	19.29	416.33 ***	258.00 ***	9.41
56	894	177	L	F3	49.6	40.8	45.2	59.2	30.97	**	19.35	**	3.18	2.08	5.39	3.74	17.25	361.63 **	220.00 **	8.17
57	177	28	L	F1	40.8	45.6	43.2	57.6	33.33	*	26.32	**	6.00	5.39	4.16	4.78	14.30	199.36 **	165.30 **	15.50
57	177	28	L	F3	40.8	45.6	43.2	57.6	33.33	**	26.32	**	6.00	5.39	4.16	4.78	7.49	56.88 *	39.03 *	4.43
58	358	177	L	F1	42.4	40.8	41.6	68.0	63.46	**	60.38	***	33.00	4.87	5.39	5.13	9.48	84.84 **	75.95 *	16.80
58	358	177	L	F3	42.4	40.8	41.6	35.2	-15.39		-16.98		-8.00	4.87	5.39	5.13	12.12	136.19 **	124.80 **	26.90
59	177	350	L	F1	40.8	64.8	52.8	73.6	39.39	**	13.58	**	1.73	5.39	1.69	3.54	6.98	97.12 **	29.55	1.86
59	177	350	L	F3	40.8	64.8	52.8	52.0	-1.52		-19.75		-0.07	5.39	1.69	3.54	6.43	81.44 **	19.24	1.56
60	93	177	L	F1	37.6	40.8	39.2	62.0	58.16	*	51.96	***	14.30	7.94	5.39	6.66	16.72	150.85 **	110.50 **	7.88
60	93	177	L	F3	37.6	40.8	39.2	40.0	2.04		-1.96		0.50	7.94	5.39	6.66	11.38	70.75 **	43.32 *	3.70

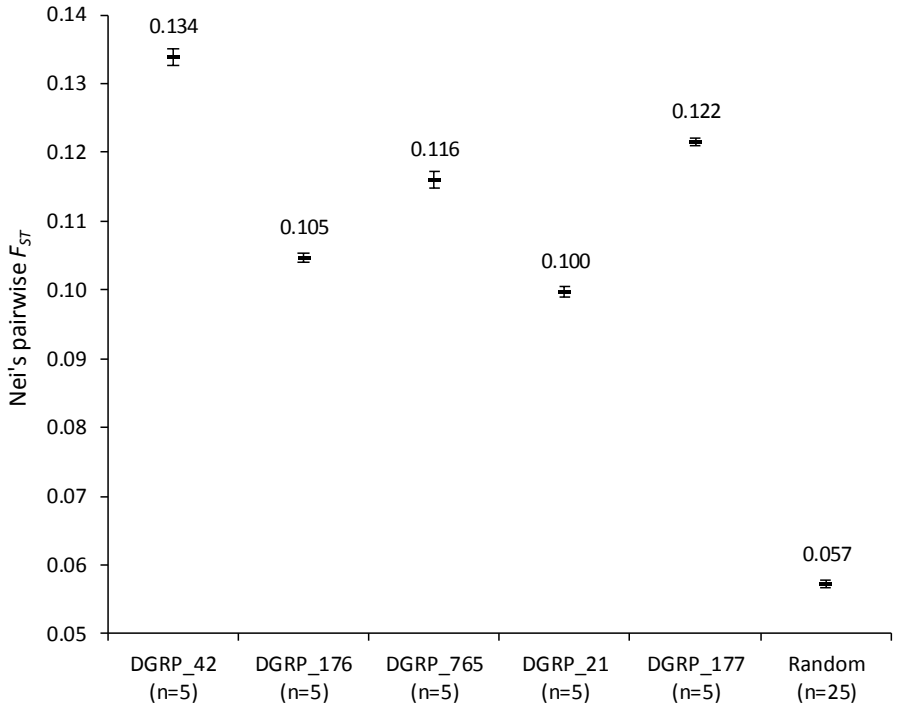
Supplementary Table S2. This table is supplied online: https://goo.gl/t6yPJs. Genomic Relationship Matrix (G-matrix) of the selected DGRP lines. G-matrix based on SNPs (single nucleotide polymorphisms) in the DGRP lines identified based on the phenotypic ranking. The values are measures on how genetically uniform two DGRP lines are. A high value indicates high similarity in SNPs and hence high genetic relatedness, which is referred to as short genetic distance (GD). Oppositely, a low value indicates a low number of similar SNPs and hence a low genetic similarity, which is referred to as long GD. The diagonal (yellow) is the inbreeding coefficient, F+1 (scaling can result in these being above 2). The G-matrix is calculated from a W-matrix, which is a scaled genotype matrix, with rows containing DGRP lines and the columns containing SNPs (see 'Genetic distance measures'). The G-matrix was kindly provided by Palle Duun Rohde.

Supplementary Table S3. Correlations between mid-parental values (MP), mid-parent heterosis (MPH) and best-parent heterosis (BPH), as well as between lowest parental value (LP), MPH and BPH. Correlations are calculated separately for each trait and each generation. Values are Pearson's correlation coefficients, *R*. Asterisks denote significance of correlations: *** $p<0.001$; ** $p<0.01$; * $p<0.05$; · $p<0.1$. MP and LP are used as proxies for the level of inbreeding depression in the parental line(s), thus negative correlations indicate that poorly performing parental lines show higher heterosis.

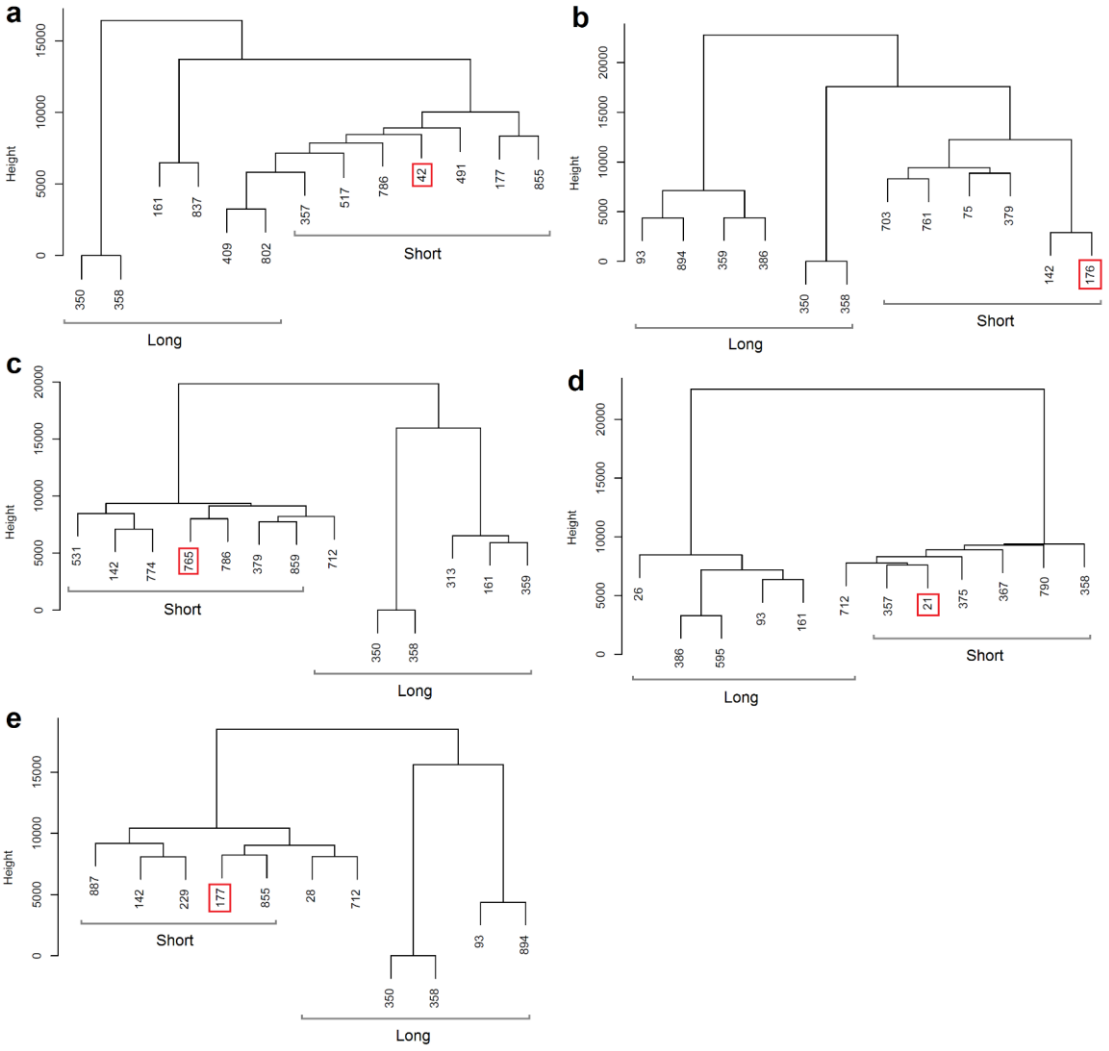
		Mid-parental value (MP)		Lowest parental value (LP)	
Trait	**Generation**	**MPH**	**BPH**	**MPH**	**BPH**
Starvation resistance	F_1	-0.496 ***	-0.484 ***	-0.333 *	-0.103
	F_3	-0.397 **	-0.411 **	-0.320 *	-0.062
Reproductive output	F_1	-0.639 ***	-0.623 ***	-0.633 ***	-0.457 ***
	F_3	-0.540 ***	-0.502 ***	-0.538 ***	-0.320 *



Supplementary Fig. S1. Boxplots showing average genetic distance based on the Genomic Relationship Matrix (G-matrix; Supplementary Table S2) of the two genetic distance (GD) groups (Long=red, Short=blue) for each of the five DGRP model populations; DGRP_21, DGRP_42, DGRP_176, DGRP_177, and DGRP_765. The values in the G-matrix are measures on how genetically uniform two DGRP lines are. A high value indicates high similarity in SNPs and hence high genetic relatedness (Short GD). Oppositely, a low value indicates a low number of similar SNPs and hence a low genetic similarity (Long GD). Dashed horizontal lines are shown at 0 for reference.



Supplementary Fig. S2. Nei's pairwise F_{ST} between pairs of populations of different genetic distances (GD). For each of the model populations (DGRP_42, DGRP_176, DGRP_765, DGRP_21, and DGRP_177), we considered the model population DGRP line together with the genetically similar DGRP lines (short GD based on the G-matrix) as one population, and the genetically distant DGRP lines (long GD based on the G-matrix) as another population. For each pairwise comparison, F_{ST} was calculated as the average of five random samplings of 100K SNPs across the entire genome. For each of the five samplings of SNPs, these F_{ST} estimates were compared to F_{ST} estimates based on the same 100K SNPs calculated from five randomly sampled sets of 20 DGRP lines (10 in each of two populations, from the total of 205 DGRP lines) for a total of 25 samplings. F_{ST} values between our selected populations were significantly greater than expected by chance ($t_{(26)}$=-21.94; p<0.001).



Supplementary Fig. S3. Dendrograms showing clustering of DGRP lines used in the study as either genetically similar or distant to each of the five DGRP model populations (marked with a red outline); (a) DGRP_42, (b) DGRP_176, (c) DGRP_765, (d) DGRP_21, and (e) DGRP_177. These dendrograms are based on hierarchical cluster analysis (HCA) of the number of loci for which individual DGRP lines differ among 30K randomly sampled SNPs (represented by the height of the vertical bars), and visualise the genetic differences and subsequent clustering into Short and Long genetic distance (GD) groups. It should be noted that the lines, that are distantly related (in terms of number of different loci = Long GD) to a particular model population, are not necessarily closely related to each other.

ADDITIONAL RESULTS

Consequences of population bottlenecks on adaptive genetic variation revealed in a highly replicated experimental evolution study

In preparation

Michael Ørsted[1,2], Ary Anthony Hoffmann[1,2], Elsa Sverrisdóttir[1], Kåre Lehmann Nielsen[1], and Torsten Nygaard Kristensen[1,3]

[1] *Department of Chemistry and Bioscience, Aalborg University, Aalborg E, Denmark.*

[2] *Bio21 Institute, University of Melbourne, Parkville, Victoria, Australia.*

[3] *Department of Bioscience, Aarhus University, Aarhus C, Denmark.*

INTRODUCTION

Recent reviews suggests that associations between the N_e, genetic variation and evolutionary potential is more complex than previously assumed (Wood *et al.* 2016; Hoffmann *et al.* 2017). Many researchers have studied these relationships, however the results are somewhat ambiguous. The majority of the studies reviewed by Wood *et al.* (2016) investigate high heritability traits such as morphology traits, which often have unclear connections to fitness. Thus, such studies might not correctly reflect genetic variation important for fitness in natural populations, where low heritabilities are common (Carlson & Seamons 2008; Hansen *et al.* 2011). In addition, many experimental studies on the effects of inbreeding on genetic variation and evolvability, employ unrealistically high levels of inbreeding (Pemberton *et al.* 2017), compared to what is found in most natural populations of both plants and animals (e.g. Crnokrak and Roff 1999; Newman and Tallmon 2001; Bowling et al. 2003; Walling et al. 2011; Huisman et al. 2016). Also, the majority of studies on the effects of inbreeding on selection and/or adaptation tend to employ only one level of inbreeding compared to outbred controls, and are therefore unable to quantify the full extent of the relationship between a range of inbreeding levels (and thus a greater range of genetic variation) and adaptive responses. Lastly, we argue that because many quantitative trait

estimates are inherently noisy (Hansen *et al.* 2011) especially for traits that are highly responsive to environmental variability (Hoffmann *et al.* 2017), large sample sizes and highly replicated inbreeding designs are needed to yield reliable estimates of the adaptive capability, especially for low heritability traits (Hoffmann *et al.* 2016).

In light of these shortcomings of many experimental studies, we setup a large-scale empirical study, where we investigated the effects of varying and ecologically relevant levels of inbreeding on the adaptive potential of *Drosophila melanogaster*. Specifically, we set up ca. 40 lines of each of three different levels of inbreeding as well as 10 outbred control lines and measured their baseline response to stressful media reduced in nutrition and increased in acidity. Each line was then reared on this media for 10 successive generations, during which we measured reproductive output and body size, to examine how the varying levels of inbreeding relates to the potential to adapt to stressful environments. We quantified viability before and after rearing on stressful medium to assess the response on survival, and to identify potential trade-offs between costs and benefits of adaptation. Lastly, we also obtained a molecular estimate of genetic variation of each line in the generation before we started the experiment, enabling quantification of the relationship between evolutionary response and genetic diversity across inbreeding levels.

MATERIALS AND METHODS

Fly stock and maintenance

The *D. melanogaster* population used in this study originated from flies caught at Oakridge winery in the Yarra Valley, Victoria, Australia (37°41'15"S 145°27'27"E) in April 2016. A total of 232 wild caught inseminated females each contributed with an equal number of offspring (five males and five females) to the establishment of a mass bred population. This population was maintained at a minimum size of 1000 individuals at 19 °C in a 12:12 L:D photoperiod. To control density, 200 parental flies laid eggs for 4-5 days in 175 mL bottles with 50 mL standard *Drosophila* sucrose-yeast-agar media. Nipagen (10 mL/L) and acetic acid (1 mL/L) was added to the media to control fungal growth. The flies were maintained for 4 generations prior to establishing the inbred lines. At the beginning of the inbreeding procedure the flies were moved to 25 °C and a 12:12 L:D photoperiod and maintained as such for the remainder of the experiments.

Inbreeding procedure

We set up lines of each of three different levels of inbreeding (hereafter referred to as low (L), medium (M) and high (H)) by controlling the number of breeding flies (N=4) in successive generations of bottlenecks (Fig. 1). To set up inbred lines from the mass bred population, virgin flies were sorted less than 8 hours after emergence under light

CO_2 anaesthesia and for each line two males and two females were transferred to a 27 mL vial with 10 mL food. After three days, the flies were tipped to another vial and discarded after another three days. To set up lines of low, medium, and high levels of inbreeding, this procedure of sorting two male and two female virgin flies was followed for a total of 2, 3 and 5 succeeding generations, respectively (Fig. 1). The lines with different levels of inbreeding were set up asynchronously, so that they reached the desired inbreeding level at the same time, after which they were flushed to a population size of minimum 200 individuals maintained in bottles. We assume there was no inbreeding in the founding population and that the effective population size (N_e) was equal to census size (N), thus, the estimated coefficient of inbreeding (F) at a given generation (t) with 2 breeding pairs, i.e. $N_e = 4$, was computed as (Crow & Kimura 1970; Kristensen *et al.* 2005):

$$F_t = \frac{F_{t-1} + (1 - 2F_{t-1} + F_{t-2})}{2N_e}$$

The estimated F of the low, intermediate and high inbreeding lines were 0.125, 0.219 and 0.381, respectively. The inbred lines went from mass bred (~1000 individuals) through a total of 2, 3 and 5 generations of bottlenecks of 4 individuals to a population size flushed to 200 individuals (Fig. 1). The effective population sizes of the three inbred populations were estimated as the harmonic mean of the fluctuating population sizes over t generations (Frankham *et al.* 2013):

$$N_e = t \cdot \left(\sum_{i=1}^{t} \frac{1}{N_i}\right)^{-1}$$

The estimated N_e of the low, intermediate and highly inbred lines were 5.6, 6.6 and 7.9, respectively. Some lines were lost due to extinction or the death of one or more of the four flies during breeding. Therefore, more lines than needed were set up to ensure that enough reached the expected level of inbreeding. The total number of lines after the inbreeding procedure was approximately 40 lines per inbreeding level plus 10 outbred lines (hereafter referred to as outbreds 1-10; OB1-OB10) totalling ca. 130 lines at the beginning of the experiment (for exact numbers see Table 1).

Baseline characterization of stress response

Preliminary range finding tests of both yeast and acid concentrations revealed large differences in responses to the treatments both between and within inbreeding levels. Therefore, we characterized the response of each of the 130 lines to the varying stress levels in order to start the experiment at a line specific stress level that yielded an approximately similar response in all lines (Fig. 1). We set up all lines including the outbred on four different stressful low-nutrition-low-pH media consisting only of 9.5 g/L yeast, 16 g/L agar and 1.0, 2.5, 5.0 or 10.0 mL/L acetic acid, plus a benign control of standard *Drosophila* sucrose-yeast-agar media. Five replicate vials with 10 mL of the respective media were set up for each treatment. In this baseline characterization, we measured egg-to-adult viability by allowing approximately 20 flies (4-5 days old) to lay eggs in a vial containing a plastic spoon with 1.5 mL standard media. After 12 h, 15 eggs were picked from the spoon and transferred to the vials of each treatment, while carefully avoiding transferring media to the low nutrition vials. In total 48,750 eggs were distributed to 3,250 vials. The acetic acid concentration yielding the survival closest to 50 % egg-to-adult viability was selected as the acid concentration used in the experimental evolution study. The results of the baseline characterization of egg-to-adult viability can be seen in Supplementary Table S1.

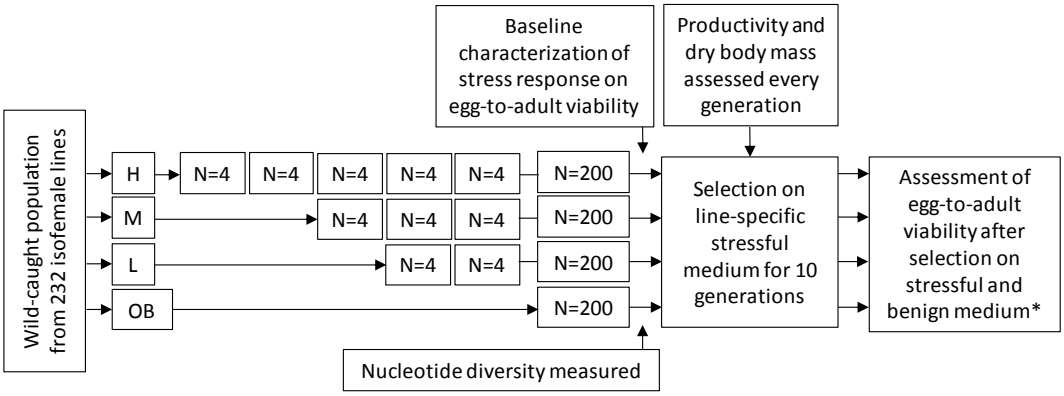


Fig. 1. Experimental procedure from setting up inbreeding regimes, to measuring initial stress response and starting nucleotide diversity, followed by 10 generations of exposure to stressful medium, and lastly assessment of egg-to-adult viability. The stressful medium was line-specific, i.e. the acetic acid concentration yielding the survival closest to 50 % egg-to-adult viability in the initial baseline characterization of the stress response was selected as the acid concentration used in the experimental evolution study. *Egg-to-adult viability was assessed after the 10 generations on the stressful medium on which the specific line had been reared, and a benign medium. This viability was compared to the egg-to-adult viability from the baseline response, to identify adaptive responses for this trait. Productivity and dry body mass was assessed every generation. See text for details on each step in the procedure.

Main evolutionary experiment

Based on the baseline characterization of egg-to-adult viability, lines were exposed to two different stress levels from the beginning of the experiment, and for each line this level of stress was maintained throughout the experiment. A total of 123 inbred lines (43 low, 40 medium, and 40 high; Table 1) were started on stressful media containing 9.5 g/L yeast, 16 g/L agar and either 1 mL/L (85 lines) or 2.5 mL/L (38 lines) acetic acid (see Supplementary Table S1). The 10 outbred control lines were started on the 2.5 mL/L acetic acid media. Approximately 200 adult flies (5 days old) were transferred to 175 mL bottles containing 50 mL of the respective stressful media, and laid eggs for 48 h, then tipped to a new identical bottle and laid eggs for another 48 h before being stored in absolute ethanol and counted. Flies that had died on the medium in either of the two bottles were also counted. When the first flies emerged from a given line, adult flies were collected over the following days and transferred to a 175 mL bottle with 50 mL standard media sprinkled with dry yeast, to recover before again being exposed to the stressful medium. This was done to stimulate egg-production and to reduce maternal carry-over effects of the low-nutrition media, i.e. a cumulative reduction of egg-production throughout the generations. When approximately 200 emerged flies from a line had been collected and all flies had had a minimum of 5 days recovery, they were transferred to a new bottle with stressful media similar to the previous generation, and the egg-laying procedure was repeated. This was carried out over 10 generations (Fig. 1).

Phenotypes assessed

Productivity

All the flies that emerged from the bottles were stored in ethanol and counted to provide an estimate of total number of flies produced by each line in each generation. This included the collected flies that contributed to the next generation after the egg-laying periods, and all the flies that emerged after enough flies had been collected. All flies were considered emerged from a bottle when no flies had emerged from a given bottle for 10 consecutive days (because of the poor nutritional quality flies often emerged over a long period). We computed a total productivity measure (adult flies produced per female per day) to account for slight deviations in egg-laying time and in number of females. We assumed a 1:1 sex ratio, thus the total number of egg-laying females for a new generation was half of the total number of flies (approximately 200 total flies = 100 females per generation).

Dry body mass

From each line and each generation, the dry body mass (hereafter referred to simply as body mass) of 15 males were measured by drying the flies at 60 °C for 24 h (for exact numbers see Supplementary Table S2). To prevent re-absorption of humidity,

the samples were transferred to a desiccator with silica gel after drying, and from there flies were transferred and measured individually on a Quintix35-1S laboratory scale with a resolution of 0.01 mg (Sartorius, Göttingen, Germany). In total 15,343 males were individually assessed for body mass.

Egg-to-adult viability after 10th generation

To assess the evolutionary response on number of surviving adults on the stressful medium, we assessed egg-to-adult viability after 10 generations. This followed the same procedure as the assessment of egg-to-adult viability in the initial baseline characterization of stress response, with the exception that viability was determined only on the stressful medium on which the specific line had been exposed to, and a benign medium. In addition, the numbers of replicate vials were increased to 10 per line per environment. For this assessment of egg-to-adult viability, 28,880 eggs were distributed to 1920 vials. For each line, the results were compared to the initial response as determined before starting the experiment for that line's respective stressful medium.

Assessment of genetic diversity by GBS

DNA extraction

From each line, a sample of 15 males (~15 mg wet weight) was homogenized in a tube with three sterile 2 mm glass beads by subjecting it to 2x6 s cycles at 6500 rpm using a Precellys mechanical homogeniser (Bertin Technologies, Montigny le Bretonneux, France). DNA was extracted with DNeasy® Blood & Tissue Kit (QIAGEN, Hilden, Germany) following a specialized protocol for insect tissues according to manufacturer's instructions. Concentration and purity of extracted DNA was assessed on a 1 % agarose gel and on a NanoDrop 1000 spectrophotometer (Thermo Scientific, Waltham, MA, USA).

Preparation of genotyping-by-sequencing libraries

5' and 3' barcoding adapters were designed as described in Sverrisdóttir *et al.* (2017). Adapters were designed to contain a 3 bp overhang complementary to the overhang generated by *Ape*KI (CWG). 5' adapters also contained eight different internal 4 to 8 bp barcode sequences, as described in (Elshire *et al.* 2011), while 3' adapters contained 12 different 6 bp barcode sequences compatible with standard Illumina sequencing multiplexing, enabling a 96 multiplexing system. Adapter were designed so that the *Ape*KI recognition site did not occur in any adapter sequence and was not regenerated after ligation to genomic DNA.

DNA samples were digested with *Ape*KI (NEB) and ligated to adapters according to the 96 Plex GBS protocol developed by Elshire *et al.* (2011) with minor revisions. Sets of 66 differently barcoded samples were combined in two pools and purified

using Agencourt AMPure XP PCR purification system (Beckman Coulter, Indianapolis, IN, USA). Restriction fragments from each library were amplified in 50 µL volumes containing 4 µL pooled DNA fragments using Phusion High-Fidelity PCR kit (Thermo Scientific). Primer design and temperature cycling was performed according to the protocol developed by Elshire *et al.* (2011). Libraries were purified as before and diluted to 2 nM as determined by Qubit (Thermo Scientific). Single-read sequencing (200 bp) was performed on a rapid run flow cell on a HiSeq 2500 (Illumina, San Diego, CA, USA).

Nucleotide diversity

Sequenced reads were demultiplexed using fastq-multx (Aronesty 2013) sorting the data into separate files, removing the barcode, and discarding reads that did not perfectly match any of the barcodes. All samples of barcoded GBS tags were pooled and sequenced. To ensure equal chance of detecting variants across all samples, 500000 reads were sampled from each sample and mapped to the reference genome of *Drosophila* r6.14 using the CLC Workbench v9.5.2 using default parameter, no masking of repetitive regions and a length fraction of 0.5 and a similarity fraction of 0.8. Non-specific matches were ignored and thus not included in the analysis. Following mapping, variants were called using the Low Frequency Variant Detection module using a required significance of 1%, min coverage of 20, max coverage of 200000, minimum count of 1 and a min frequency of 5%. Again non-specific matches were ignored. The variant table for each sample were exported. Using a custom Bash script, the variants mapping to autosomes 2, 3 and 4 were used to calculated nucleotide diversity (π) for each variant loci:

$$\pi = 2p(1-p) * 1/130$$

As our measure of genetic variation, a relative population measure of nucleotide diversity (π(apparent)) for each sample was estimated by summing π over all variant loci for each sample. Note that the term 1/130 is used here because after trimming average read length of all samples were very close to 130 bp, thus assuming a single polymorphism in each reads makes the diversity constitute 1/130 of each read. This is strictly speaking not necessarily true, but violation of this assumption does not impact the use of π(apparent) as a relative measure between the samples, only the absolute value of π. From here on in, π will refer to π(apparent).

Statistical analysis

Body mass data conformed to assumptions of normality and was analysed with a linear model. Productivity data were analysed with a generalised linear effect model (GLM) with a Poisson distribution. Egg-to-adult viability data were analysed with a

GLM with a binomial response and a logit link function. We detected overdispersion in this model and corrected for this using a quasi-generalised linear model. In the analysis of the difference in viability from before and after the experiment, the SE of the difference was calculated from the variance sum law: $\sqrt{\frac{\sigma^2_{F10}}{n_{F10}} + \frac{\sigma^2_{F0}}{n_{F0}}}$, where σ^2 are the variances and n are the sample sizes of viability measures from F10 and F0, respectively. In analyses of productivity and egg-to-adult viability across generations general linear mixed models (GLMMs) were employed with line ID included as a random effect, as measures of a given line across generations are not independent. Body mass was analysed across generations with repeated measures (RM) ANOVAs. Tukey HSD post hoc test was used for RM ANOVAs, while for GLMs and GLMMs, post hoc multiple comparisons was performed with the R-package 'multcomp' (Hothorn *et al.* 2008). The p-values were corrected for multiple testing using Bonferroni correction. Nucleotide diversity was compared across e.g. generations using simple Welch's t-tests. As a measure of evolutionary response, we used the regression coefficients (slope) of the linear models of the response across generations, and slopes of the different inbreeding levels were compared using interactions in the models. To assess the linearity of the response, we compared the slope across all 10 generations with the slopes calculated across generations 1-3, 4-6, and 7-10, respectively. A range of measures was compared to nucleotide diversity using Pearson's product-moment correlation for body mass data and Spearman's rank correlation for productivity data. The correlated adaptive response between body mass and productivity was tested in the same manner using the slopes. We calculated between line coefficient of variation (CV=SD/mean) for each generation for each inbreeding group as a measure of divergence within and between groups. All statistical analyses were performed in R (R Core Team 2017) (v. 3.4.0), and mixed models were performed using the R-package 'lme4' (Bates *et al.* 2015).

RESULTS

Extinctions were associated with low genetic variation

The first step was to assess if the lines of the different expected inbreeding levels had lower genetic variation. The nucleotide diversity of all the inbred lines varied more than for the outbred control lines. The nucleotide diversity of the H and M inbred lines were significantly lower than the outbred and L inbred lines, which could not be mutually distinguished (Fig. 2). Although there was an overall declining trend with increased inbreeding, results show that inbreeding does not necessarily result in reduced nucleotide diversity, and that for this parameter some inbred lines perform just as well or even better than some of the outbred lines. We started with 123 inbred lines and 10 control outbred lines (Table 1). While all control lines persisted, 37 of the inbred lines went extinct during the experimental evolution procedure, with more lines going extinct in the medium and highest inbreeding groups (17, 15, and 5 for H, M, and L lines, respectively), and the majority of these going extinct in the first three generations (Fig. 3). The nucleotide diversity of lines, that went extinct were overall lower than the lines that did not go extinct (Fig. 4), with an indication that the earlier the extinction, the lower the diversity, however, this could not be statistically verified. Interestingly, we also observed a few notable exceptions where lines with relatively high diversity went extinct in the first and second generation, as well as the converse where low diversity lines did not go extinct, suggesting that some highly inbred lines performed just as well as less inbred or even outbred lines in terms of persistency.

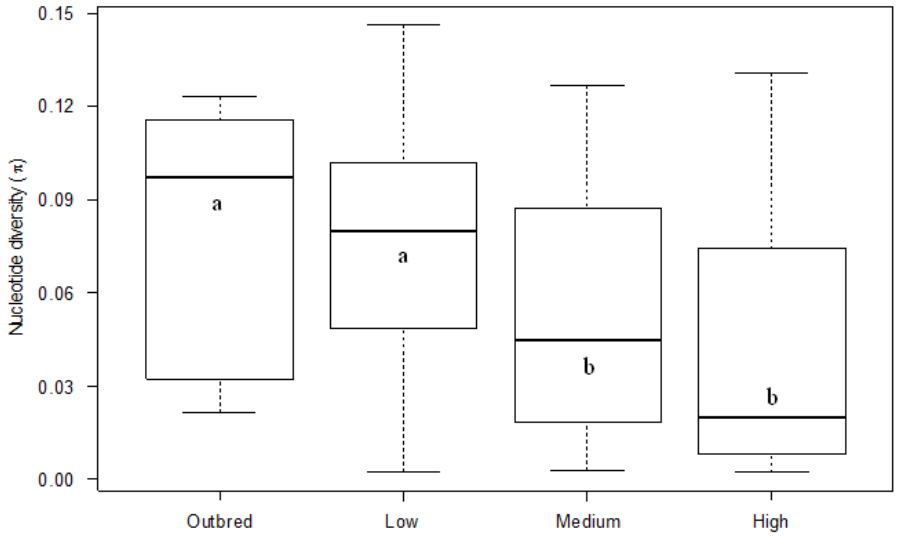


Fig. 2. Boxplots showing nucleotide diversity (π) of the experimental lines of the three inbreeding levels (Low, Medium, and High), and the control (Outbred) lines. Letters denote significant differences (as determined by Welch's t-test; $P < 0.05$).

Table 1. Number of lines for each generation in the current study for each level of inbreeding (Low; L, Medium; M, and High; H) and the outbred control lines (OB). Bottom row shows total number of lines per generation.

	Generation									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
OB	10	10	10	10	10	10	10	10	10	10
L	43	42	41	39	39	38	38	38	38	38
M	40	39	30	29	28	27	26	26	25	25
H	40	32	27	25	25	24	23	23	23	23
Total	133	123	108	103	102	99	97	97	96	96

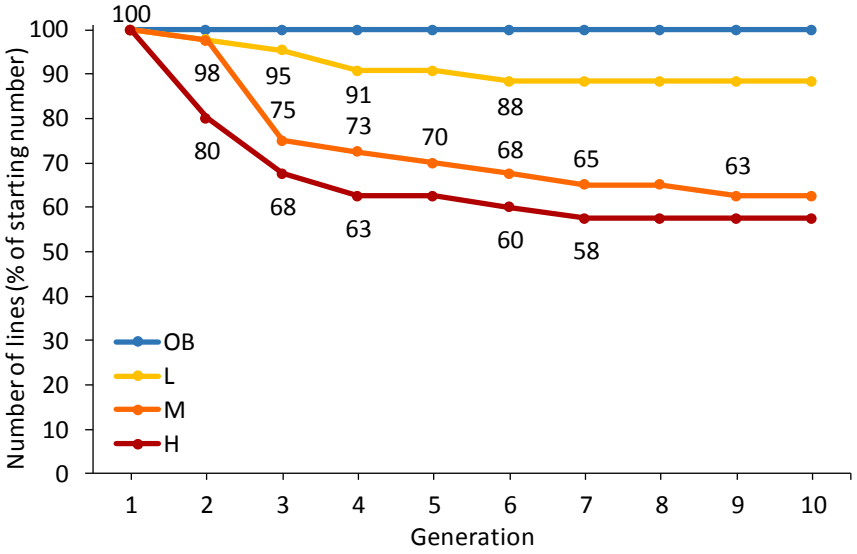


Fig. 3. Number of experimental lines lost during the experiment of the three inbreeding levels (Low; L (yellow), Medium; M (orange), and High; H (red)), and the outbred lines (OB (blue)). Values are expressed as percentage of the number of starting lines within the respective groups.

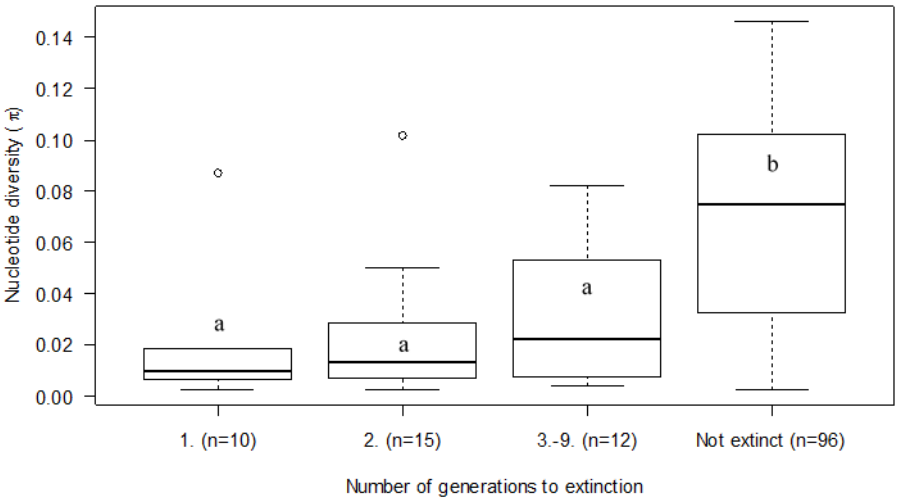


Fig. 4. Boxplot showing nucleotide diversity of lines that went extinct in either the first generation, the second generation, or in the third to the ninth generation, as well as for lines that did not go extinct. Dots represent outliers. Letters denote significant differences (as determined by Welch's t-test; $P < 0.05$). Number of lines is shown for each group (n).

Inbreeding effects on the evolutionary response were trait specific

For an overview of the effects of inbreeding level on the responses across generations we performed separate analyses for the two traits, and treated each line as repeated measures (see 'Materials and methods: Statistical analysis'). Overall, for body mass, there was a significant effect of inbreeding as well as significant effect of the interaction between generation and inbreeding, suggesting differential slopes of the inbreeding levels (Table 2). Tukey's HSD revealed non-significant effects of intercepts for all levels of inbreeding, but highly significant effects of interactions for all levels, indicating that outbred and inbred lines all initially had approximately similar body mass, but that the different inbreeding levels responded differently to the stressful medium. The L lines had a significantly lower slope than the outbred lines, while the M and H lines had an even lower response. This is similarly suggested by the line means, where we observed an adaptive response that was dependent on inbreeding level (Fig. 5A). Starting body mass of the different inbreeding levels (generation 1) could not be distinguished from each other or from the outbred lines, however after 6 generations, the outbred and L lines had increased in body mass, enough to be statistically different from the M and H lines (Fig. 5A). After 10 generations, there was a significant difference in body mass between the outbred control lines and the L lines. The body mass of M and H lines was on average constant throughout the experiment, i.e. they did respond evolutionary to exposure to stressful

conditions. The response in body mass of all lines across generations can be seen for each inbreeding level in Fig. 6.

For productivity, the effects of generation and inbreeding level was significant, however there was no effect of the interaction between the two (Table 2). The intercept for the outbred lines were significantly different from the inbred lines. The intercepts of L and M lines were also different from that of H lines. For all lines, we found no evidence of an interaction. This is also seen for line means for productivity (Fig. 5B), where there was greater difference in the initial measure of productivity of the inbred and outbred lines as compared to for body mass (generation 1; Fig. 5). This could suggest a higher degree of inbreeding depression for productivity than for body mass, which is congruent with the expectation, that inbreeding depression is highest in life-history traits that are closely related to fitness (productivity). Body mass is most likely also important for fitness, however the connection is less clear. In any case, the degree of inbreeding depression in body mass seem negligible. The response in productivity of all lines across generations can be seen for each inbreeding level in Fig. 6.

Table 2. Results of linear models with effects of generation (Gen), and inbreeding level and their interaction (top half) for dry body mass (left side) and productivity (right side). Degrees of freedom (Df), sum of squares (SS), mean squares (MS), F-, and P-values are shown for each source of variation. Symbols denote significance level: *** P < 0.001, ** P < 0.01, * P < 0.05, · P < 0.1. The lower half shows the summary of these models with the outbred lines as the reference parameter, showing estimates, standard error (SE), t- and P-values. All estimates are expressed relative to these lines including interactions i.e. a negative interaction between generation and any of the inbreeding levels (L, M, and H) equals a lower slope. The P-values refer to whether the intercept or gen:inbreeding interaction of the inbred lines is significantly different from that of the outbred controls. To test if the intercepts and interactions were mutually different (P < 0.05), Tukey's HSD post hoc multiple comparisons was performed. Letters denote significant differences in intercepts (uppercase), and slopes/interactions (lowercase). Note that Tukey's HSD for productivity is a modified non-parametric version. All P-values were corrected for multiple testing using Bonferroni correction.

	Dry body mass						Productivity					
	Df	SS	MS	F	P value		Df	SS	MS	F	P value	
Gen	1	0.007	0.007	3.616	0.058	·	1	228.400	228.430	27.599	< 0.001	***
Inbreeding	3	0.362	0.121	63.035	< 0.001	***	3	1781.200	593.720	71.733	< 0.001	***
Gen:Inbreeding	3	0.077	0.026	13.385	< 0.001	***	3	19.600	6.540	0.791	0.499	
Residuals	952	1.822	0.002				952	7879.500	8.280			

	Tukey Grp.	Estimate	SE	t	P value		Tukey Grp.	Estimate	SE	t	P value	
Intercept	A	0.187	0.009	19.822	< 0.001	***	A	6.322	0.621	10.173	< 0.001	***
Gen	a	0.007	0.002	4.701	< 0.001	***	a	0.116	0.100	1.162	0.245	
L	A	-0.001	0.011	-0.108	0.914		B	-1.892	0.699	-2.708	0.007	**
M	A	0.001	0.011	0.093	0.926		B	-1.809	0.735	-2.460	0.014	*
H	A	-0.011	0.011	-0.938	0.348		C	-4.112	0.744	-5.524	< 0.001	***
Gen:L	b	-0.004	0.002	-2.478	0.013	*	a	0.096	0.113	0.850	0.395	
Gen:M	c	-0.009	0.002	-5.025	< 0.001	***	a	0.078	0.119	0.662	0.508	
Gen:H	c	-0.009	0.002	-5.005	< 0.001	***	a	-0.021	0.120	-0.171	0.864	

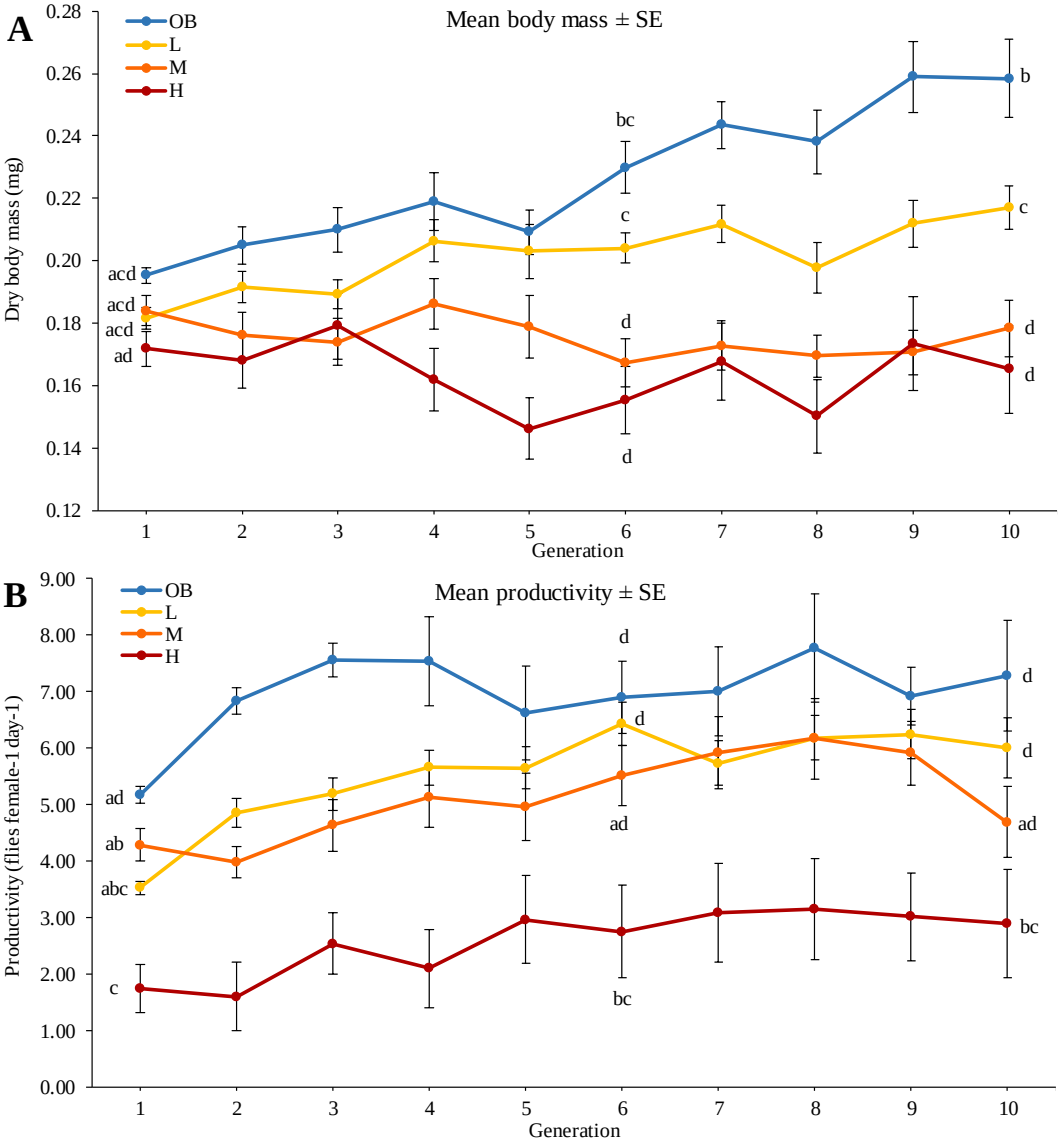


Fig. 5. Response in (A) body mass, and (B) productivity across generations 1-10 of the three inbreeding levels (Low; L (yellow), Medium; M (orange), and High; H (red)), and the outbred lines (OB (blue)). Error bars represent SE. The number of lines at each generation can be seen in Table 1. Letters denote significance groups at selected generations 1, 6, and 10, as based on Tukey's HSD post hoc multiple comparisons test ($P < 0.05$). Note that Tukey's HSD for productivity is a modified non-parametric version. All P-values were corrected for multiple testing using Bonferroni correction.

Only the productivity of L lines was significantly different in generations 1 and 10, the rest were not significantly differentiated. Altogether, these results suggest that the inbreeding effects are trait specific: for body mass, the starting point is the same for all lines, but the evolutionary changes are greatly dependent on inbreeding levels, whereas for productivity inbreeding effects have a large impact on starting point, but a less pronounced impact on the evolutionary response. The response of either trait was independent on which acid level the lines were exposed to (ANOVA; $F_{1,21}$=0.678; P=0.422, for body mass, and $F_{1,94}$=0.717; P=0.401 for productivity; see Supplementary Table S1 for different acid levels).

Line divergence across generations were related to level of inbreeding

From the standard errors of the line means (Fig. 5), it is clear that the lines are becoming increasingly diverged. An overview of all lines plotted across generations similarly reveal this divergence as well as the high variability within some lines across generations (Fig. 6). An increasing degree of variation across generations could be attributed to a decreasing number of lines because of extinctions, however we account for some of that discrepancy by using the SE. In addition, the variation between outbred lines also increases, and none of these lines went extinct during the experiment. To further account for the potential simultaneous increase in means, the increasing variation was expressed as the coefficient of variation (CV; Fig. 7). Interestingly, we also found effects of inbreeding and trait specificity in the CVs across lines. Although statistical comparisons are not performed, it is notable how the CV is generally higher for the inbred lines (at least M and H lines) than for the outbred lines. For body mass, the CVs of all lines are increasing across generations, and for e.g. the H lines, CV is almost doubled after 10 generations of experimental evolution (Fig. 7A). Since the mean is not increasing, this suggest that the SD increases disproportionally with the mean, suggesting that lines are diverging during the study, and that this divergence is dependent on inbreeding level. For productivity, the CVs are more constant across generations, and consistently higher for the most inbred lines compared to L and outbred lines (Fig. 7B).

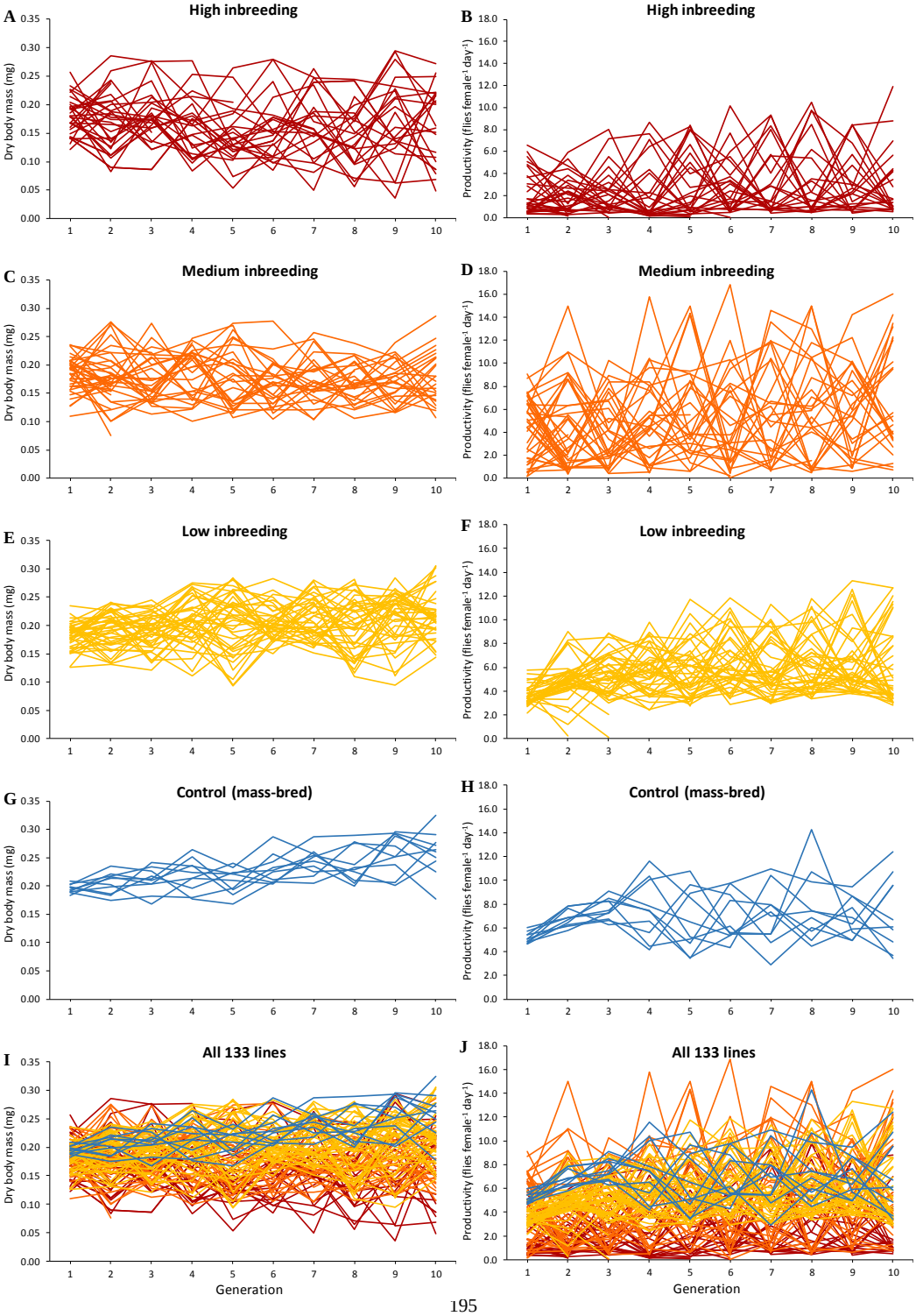

A
High inbreeding
Dry body mass (mg)
B
High inbreeding
Productivity (flies female-1 day-1)
C
Medium inbreeding
Dry body mass (mg)
D
Medium inbreeding
Productivity (flies female-1 day-1)
E
Low inbreeding
Dry body mass (mg)
F
Low inbreeding
Productivity (flies female-1 day-1)
G
Control (mass-bred)
Dry body mass (mg)
H
Control (mass-bred)
Productivity (flies female-1 day-1)
I
All 133 lines
Dry body mass (mg)
Generation
J
All 133 lines
Productivity (flies female-1 day-1)
Generation

Fig. 6. Line plots of all lines across generations 1-10 within each group: High (A-B; in red), Medium (C-D; in orange), Low (E-F; in yellow), and outbred lines (G-H; in blue), and all lines plotted together (I-J) for body mass (left side), and productivity (right side). Y-axes are similar for all plots within a trait for ease of comparison.

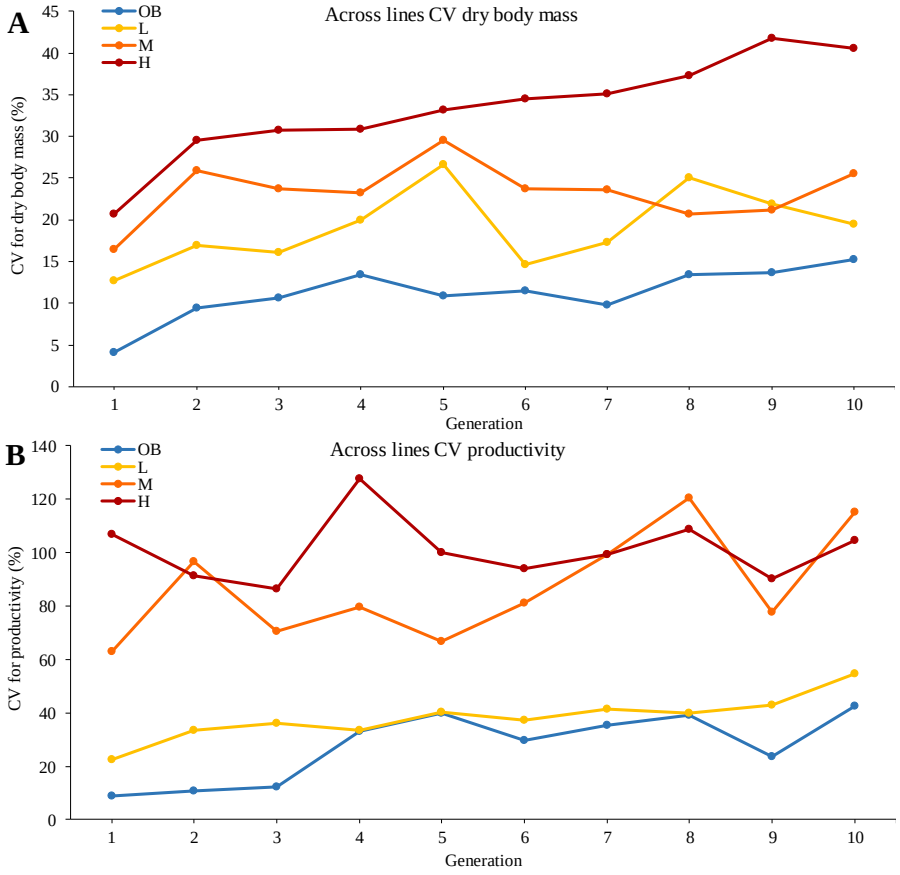


Fig. 7. Coefficient of variation (CV in %) across lines in (A) body mass, and (B) productivity across generations 1-10 of the three inbreeding levels (Low; L (yellow), Medium; M (orange), and High; H (red)), and the outbred lines (OB (blue)).

Rate of adaptation was mostly constant

For the majority of analyses, we employed the slope of the linear regressions as a measure of evolution, which seems appropriate. However, the evolutionary response to the environmental conditions might not be constant across generation, so to test the rate of adaptation and the linearity of the evolutionary response, we compared the slope on all generations with subsets of generations F1-F3, F4-F6, and F7-F10 (Table 3). Firstly this confirmed that the slope of the response in body mass was statistically higher for OB and L lines, while H lines did not respond. The slope of the body mass of M lines was in fact negative. For body mass the slopes were constant across all subset of generations. For productivity, the slopes across the first three generations was significantly steeper than the overall slope for OB and L lines (Table 3), whereas the slopes across generations F4-F6 and F7-F10 was not. This indicate a non-linear response, suggestive of rapid early adaptation followed by plateauing responses. Although not directly measured, this could perhaps point to a fast initial depletion of V_A followed by the responses reaching an adaptation limit. It could turn out to be difficult to fully disentangle plastic effects from adaptive evolution, as plasticity itself might evolve.

Table 3. Results of comparisons of slope for all generations (all) versus slopes for subsets of generations F1-F3, F4-F6, and F7-F10 for the three inbreeding levels (H, M, and L) and the outbred lines (OB) for body mass (top) and productivity (bottom). Asterisks denote slopes that are significantly different from 0: *** P < 0.001, ** P < 0.01, * P < 0.05. An 'a' denotes slopes from subsets of generations within an inbreeding levels, that are significantly different from the slope from all generations within the same level. All P-values were corrected for multiple testing using Bonferroni correction.

Trait	Inbreeding level	Slope (all)	Slope (F1-F3)	Slope (F4-F6)	Slope (F7-F10)
Body mass	H	-0.0020	-0.0028	0.0001	0.0028
	M	-0.0019 *	-0.0080	-0.0072	0.0017
	L	0.0029 ***	0.0041	-0.0019	0.0030
	OB	0.0072 ***	0.0072	0.0055	0.0066
Productivity	H	0.0959	-0.0103	0.3497	-0.0754
	M	0.1949 *	-0.3610	0.0603	0.2060
	L	0.2121 ***	0.9307 ***,a	0.4192	0.0848
	OB	0.1164	1.1932 ***,a	-0.3232	-0.0046

Evolutionary responses of viability suggested trade-offs

Egg-to-adult viability was assessed before and after the experiment on the stressful medium on which the specific line had been reared on throughout, and a benign medium, to identify evolutionary changes in this trait (Fig. 8). All lines except the most inbred (H) exhibited a significantly increased viability on the stressful medium after in generation 10 compared to the baseline. Conversely, all inbred lines performed significantly worse in terms of viability on the benign medium, and only the viability of outbred lines were not significantly different on the benign medium in generation 10. These results suggest a trade-off, which seems related to level of inbreeding to some extent. It also suggests that there has been some genetic change, i.e. adaptive evolution on the stressful medium. Had the change been solely a result of phenotypic plasticity in the ability to tolerate different food sources with varying levels of stress, we would not have expected a trade-off, except maybe in the cost of maintaining a higher plasticity.

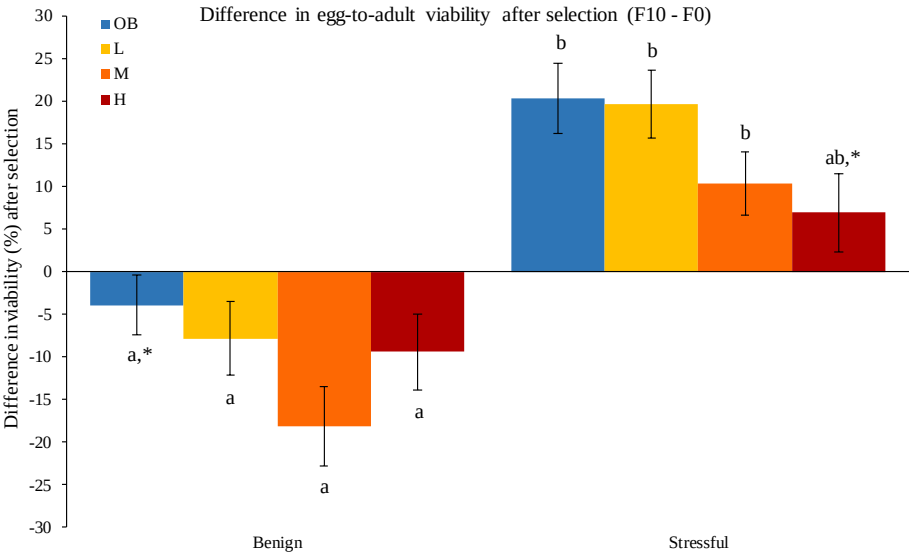


Fig. 8. Differences in egg-to-adult viability (in %) between before and after the evolutionary experiment for the three inbreeding levels (Low; L (yellow), Medium; M (orange), and High; H (red)), and the outbred lines (OB (blue)). Values are expressed as after the experiment (F10) compared to before (F0), i.e. a negative difference means that the viability is lower after the conclusion of the experiment. Values are expressed as the mean of the difference for each line, rather than the difference in means across all lines, to correctly reflect the between line variation. Error bars represent the SE of this difference, which is calculated from the variance sum as described in the methods section. Asterisks denote differences that are not significantly different from 0 ($P < 0.05$). Letters denote significance groupings across all inbreeding levels and across types of medium. All P-values were corrected for multiple testing using Bonferroni correction.

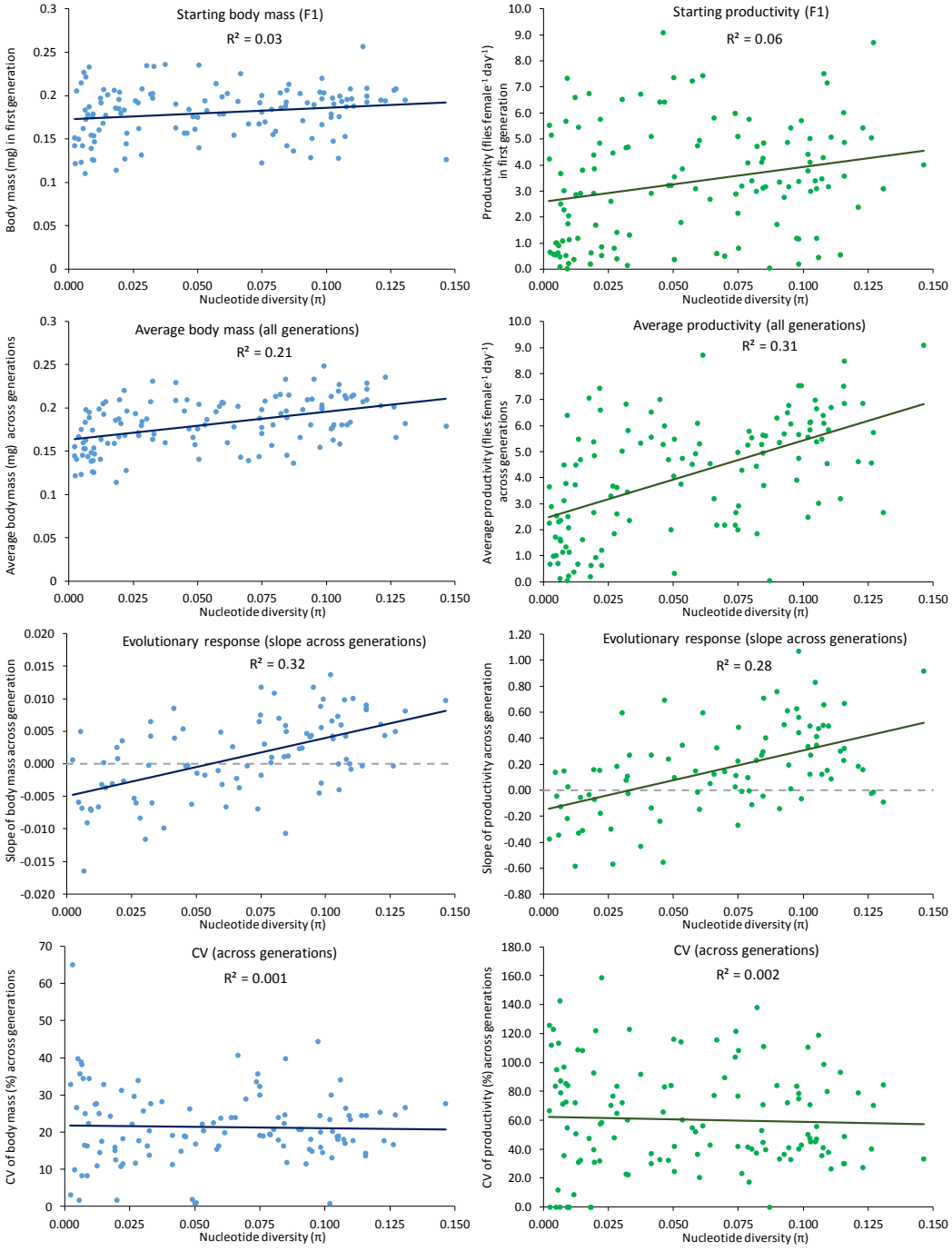

Starting body mass (F1)
R² = 0.03
Body mass (mg) in first generation
Nucleotide diversity (π)
Starting productivity (F1)
R² = 0.06
Productivity (flies female-1 day-1) in first generation
Nucleotide diversity (π)
Average body mass (all generations)
R² = 0.21
Average bodymass (mg) across generations
Nucleotide diversity (π)
Average productivity (all generations)
R² = 0.31
Average productivity (flies female-1 day-1) across generations
Nucleotide diversity (π)
Evolutionary response (slope across generations)
R² = 0.32
Slope of body mass across generation
Nucleotide diversity (π)
Evolutionary response (slope across generations)
R² = 0.28
Slope of productivity across generation
Nucleotide diversity (π)
CV (across generations)
R² = 0.001
CV of body mass (%) across generations
Nucleotide diversity (π)
CV (across generations)
R² = 0.002
CV of productivity (%) across generations
Nucleotide diversity (π)

Fig. 9. Correlations between nucleotide diversity (π) and starting values in the first generation (A-B), average values across all generations (C-D), slope of evolutionary response (E-F), and CV across generations (G-H) for body mass (left side in blue) and productivity (right side in green). For all regressions, R^2 values are shown. For body mass, Pearson's product-moment correlations are used, and for productivity, Spearman's rank correlations are used. Asterisks denote significant correlations: *** P < 0.001, ** P < 0.01, * P < 0.05. All these correlations are based on nucleotide diversity estimates from 128 lines (for 5 lines, the quality of DNA extracts did not allow estimation of π), except for slope where we only considered slope for lines that did not go extinct, to ensure unreliable slope estimates across e.g. 2 generations; in total this yielded slope estimates for 92 lines.

Nucleotide diversity was a good predictor of evolutionary response

As suggested by the nucleotide diversity of the different inbreeding groups, as well as for the outbred groups (Fig. 2), the genetic diversity across these groups is continuously declining. This is what we would expect, as not all lines are inbred to the same degree and in the same regions of the genome. Therefore, it might be more appropriate to treat the loss of genetic variation as a result of inbreeding as a continuous parameter and disregard the expected inbreeding coefficient, F. This allowed us to correlate responses for all lines to nucleotide diversity. We observed weak but significant positive correlations between π and starting values for both body mass and productivity (Fig. 9A-B). The correlations between π and body mass averaged across all generations were stronger and similarly positive (Fig. 9C-D). Both of these measures can be regarded as proxies for inbreeding depression, however the latter is of course also affected by evolution during the 10 generations. A high average across generations can occur both in lines with a high starting values but a low slope, or in lines with low starting values, but a steeper slope, i.e. more adaptive potential, and *vice versa* for a low average across generations. Which of the two contributes to the average cannot be distinguished from this measure. Therefore, we also correlated π with slope across all generations similarly to the measures in Tables 2 and 3, however here presented for all lines regardless of expected F. We found a positive correlation between π and slope for both body mass and productivity (Fig. 9E-F), which indicated that π was a good predictor of evolutionary response. However, there were exceptions reflected by the rather low R^2 values: 0.32 and 0.28 for body mass and productivity, respectively. We correlated the slope for the body mass and productivity for each line (n=92), to explore whether there was a correlated response in the two traits (Fig. 10). We found no evidence of such a correlated response, suggesting that the adaptations in the two traits are somewhat independent. Lastly, we found no correlation between π and CV within line across generations for body mass (Fig. 9G) and we found a very weak, albeit significant, correlation for productivity (Fig. 9H). These CV measures within line across generations are different from the across lines CV presented in Fig. 7, and thus does not yield any information of

variation within line within generations. For productivity, we only have one measurement per generation and thus cannot get this information, but for body mass, where we have ~15 measurements per line per generation, further analyses will hopefully elucidate the relationship between within line variation and genetic diversity.

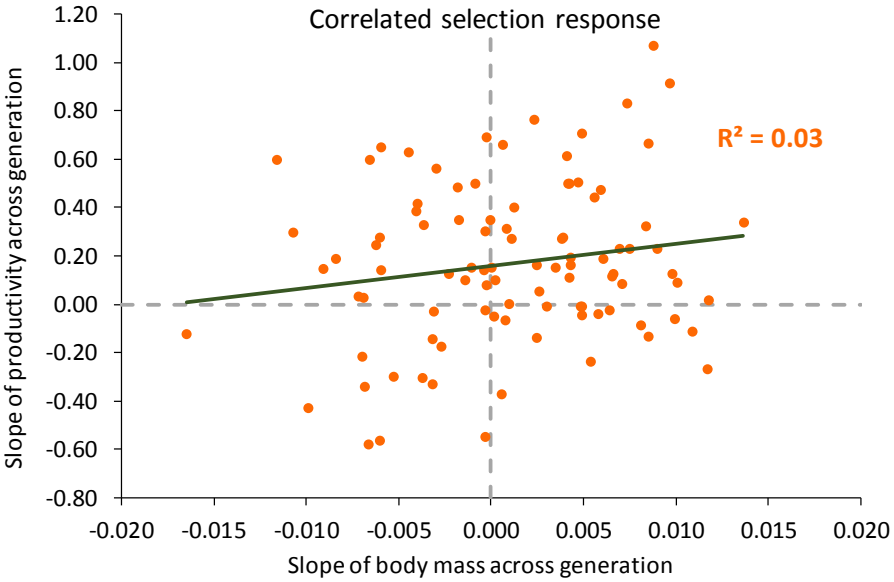


Fig. 10. Correlation between slopes of body mass and productivity as measures of evolutionary responses in the respective traits. The R^2 value is shown. This correlation was not significant (Pearson's r=0.168; t(90)=1.657; P=0.101). We only considered slope for lines that did not go extinct, to ensure unreliable slope estimates across e.g. 2 generations; in total this yielded slope estimates for 92 lines.

ACKNOWLEDGEMENTS

We thank Kelly Richardson for providing the fly stock population used in the experiments, and we thank her and Perran Ross for their help in setting up the inbred lines, and Nancy Endersby, Helle Blendstrup, Susan M. Hansen, and Neda N. Moghadam for laboratory assistance. This research was supported by the Danish Natural Science Research Council through a Sapere aude stipend to TNK (DFF–4002-00036), the Science Industry Endowment Fund and a grant from the Velux Visiting Professor programme to AAH. The authors declare no conflict of interests regarding the publication of this paper.

REFERENCES

Aronesty E (2013) Comparison of sequencing utility programs. *The Open Bioinformatics Journal*, **7**, 1–8.

Bates D, Mächler M, Bolker BM, Walker SC (2015) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, **67**, 1–48.

Bowling AT, Zimmermann W, Ryder O *et al.* (2003) Genetic variation in Przewalski's horses, with special focus on the last wild caught mare. *Cytogenetic and Genome Research*, **102**, 226–234.

Carlson SM, Seamons TR (2008) A review of quantitative genetic components of fitness in salmonids: implications for adaptation to future change. *Evolutionary Applications*, **1**, 222–238.

Crnokrak P, Roff DA (1999) Inbreeding depression in the wild. *Heredity*, **83**, 260–270.

Crow JF, Kimura M (1970) *An introduction to population genetics theory*. Harper and Row, New York, NY.

Elshire RJ, Glaubitz JC, Sun Q *et al.* (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, **6**, e19379.

Frankham R, Ballou JD, Briscoe DA (2013) *Introduction to Conservation Genetics*. Cambridge University Press, Cambridge, United Kingdom.

Hansen TF, Pélabon C, Houle D (2011) Heritability is not evolvability. *Evolutionary Biology*, **38**, 258–277.

Hoffmann AA, Merilä J, Kristensen TN (2016) Heritability and evolvability of fitness and nonfitness traits: lessons from livestock. *Evolution*, **70**, 1770–1779.

Hoffmann AA, Sgrò CM, Kristensen TN (2017) Revisiting adaptive potential, population size, and conservation. *Trends in Ecology and Evolution*, **32**, 506–517.

Hothorn T, Bretz F, Westfall P (2008) Simultaneous inference in general parametric models. *Biometrical Journal*, **50**, 346–363.

Huisman J, Kruuk LEB, Ellis PA, Clutton-Brock T, Pemberton JM (2016) Inbreeding depression across the lifespan in a wild mammal population. *Proceedings of the National Academy of Sciences of the United States of America*, **113**, 3585–3590.

Kristensen TN, Sørensen AC, Sorensen D *et al.* (2005) A test of quantitative genetic theory using *Drosophila* - effects of inbreeding and rate of inbreeding on heritabilities and variance components. *Journal of Evolutionary Biology*, **18**, 763–770.

Newman D, Tallmon DA (2001) Experimental evidence for beneficial fitness effects of gene flow in recently isolated populations. *Conservation Biology*, **15**, 1054–1063.

Pemberton JM, Ellis PE, Pilkington JG, Bérénos C (2017) Inbreeding depression by environment interactions in a free-living mammal population. *Heredity*, **118**, 64–77.

R Core Team (2017) R: A language and environment for statistical computing. Version 3.4.0. R Foundation for Statistical Computing, Vienna, Austria. http://www.r-project.org/.

Sverrisdóttir E, Byrne S, Høegh E *et al.* (2017) Genomic prediction of starch content and chipping quality in tetraploid potato using genotyping‐by‐sequencing. *Theoretical and Applied Genetics*, **130**, 2091–2108.

Walling CA, Nussey DH, Morris A *et al.* (2011) Inbreeding depression in red deer calves. *BMC Evolutionary Biology*, **11**, 1–13.

Wood JLA, Yates MC, Fraser DJ (2016) Are heritability and selection related to population size in nature? Meta-analysis and conservation implications. *Evolutionary Applications*, **9**, 640–657.

Wright LI, Tregenza T, Hosken DJ (2008) Inbreeding, inbreeding depression and extinction. *Conservation Genetics*, **9**, 833–843.

SUPPLEMENTARY INFORMATION FOR ADDITIONAL RESULTS

Table S1: Egg-to-adult viability before and after the experiment.

Table S2: Line mean dry body mass for generations 1-10.
(supplied online: goo.gl/tZUjgx)

Table S3: Line mean productivity for generations 1-10.

Supplementary Table S1. Egg-to-adult viability before and after the experiment. Results of baseline characterization of egg-to-adult viability (%) of 123 lines from the three different inbreeding (Inb.) levels (H, M and L) plus 10 outbred lines (OB1-OB10) from five treatments ranging from a benign standard medium ('Control') to four stressful media consisting of 9.5 g/L yeast, 16 g/L agar and a range of 1.0, 2.5, 5.0 or 10.0 mL/L acetic acid, designated A1, A2, A3, and A4, respectively. The acetic acid concentration yielding the survival closest to 50 % egg-to-adult viability was selected as the acid concentration used in the experimental evolution study, and is given in the 'A(s)' column. The baseline experiment was set up on five replicate vials as described in text. The 'Ext.' column designates lines that went extinct during the experiment. Egg-to-adult viability (%) after the experiment on the stressful medium on which the specific line had been reared on throughout, and a benign medium, is shown in the last two columns (denoted 'Benign' and 'Stress') for the 96 lines that persisted through the experiment.

			Baseline						After	
ID	**Inb.**	**Control**	**A1**	**A2**	**A3**	**A4**	**A(s)**	**Ext.**	**Benign**	**Stress**
1	H	50.00	43.33	41.67	25.00	18.33	A1		29.29	35.88
2	H	33.33	30.00	45.00	23.33	20.00	A2	X		
3	H	56.67	38.33	36.67	30.00	11.11	A1	X		
4	H	26.67	23.33	38.33	21.67	6.67	A2		44.24	38.60
5	H	33.33	18.33	38.89	13.33	8.89	A2		5.93	68.13
6	H	29.58	15.00	26.67	17.78	10.00	A2		32.98	76.53
7	H	21.67	15.00	16.67	11.11	5.00	A2		19.79	47.21
8	H	58.33	25.00	18.33	23.33	10.00	A1		35.01	46.20
9	H	65.00	56.67	36.67	28.33	25.00	A1		25.30	32.65
10	H	48.57	46.67	56.67	38.33	21.67	A1		36.81	49.39
11	H	40.00	25.00	40.00	24.09	35.00	A2		38.54	55.28
12	H	43.33	46.67	25.00	10.00	3.33	A1	X		
13	H	39.17	11.67	23.33	21.67	8.33	A2	X		
14	H	38.33	31.67	60.00	32.62	30.00	A2	X		
15	H	38.33	40.00	53.33	45.00	38.33	A2		36.79	61.64
16	H	40.00	35.00	13.33	10.00	3.33	A1		53.24	12.52
17	H	43.75	45.00	40.00	26.67	8.33	A1	X		
18	H	43.33	58.33	59.00	35.00	15.00	A1		54.50	71.32
19	H	21.67	28.33	27.00	8.94	23.33	A1	X		
20	H	55.00	29.00	30.00	20.00	8.33	A2		10.95	52.64

Table S1 continued

			Baseline						After	
ID	**Inb.**	**Control**	**A1**	**A2**	**A3**	**A4**	**A(s)**	**Ext.**	**Benign**	**Stress**
21	H	73.33	58.33	30.00	35.56	18.33	A1	X		
22	H	52.92	53.33	55.00	35.00	21.67	A1		8.95	68.43
23	H	40.00	45.48	57.00	40.00	18.33	A1		8.33	84.10
24	H	76.67	23.33	66.67	25.00	28.33	A2	X		
25	H	73.33	65.00	51.67	35.00	28.33	A2		61.21	42.50
26	H	35.83	18.33	31.67	26.67	13.33	A2		36.88	7.28
27	H	36.67	25.00	30.48	20.00	6.67	A2	X		
28	H	8.33	31.67	11.67	8.33	6.67	A1	X		
29	H	16.67	16.67	16.67	5.00	13.33	A1	X		
30	H	68.33	60.00	35.00	33.33	18.33	A1		56.68	50.91
31	H	9.52	15.00	10.00	0.83	5.00	A1	X		
32	H	45.00	61.67	31.67	37.00	21.67	A1	X		
33	H	50.00	61.67	63.33	70.00	35.00	A1		18.15	45.60
34	H	53.33	48.33	38.33	43.33	20.00	A1		61.75	50.58
35	H	53.33	53.33	31.67	21.67	11.67	A1		54.49	83.41
36	H	60.00	31.67	25.00	30.00	15.56	A1	X		
37	H	51.67	21.67	28.33	22.22	10.00	A2	X		
38	H	20.00	35.00	21.67	25.00	13.33	A1	X		
39	H	40.00	26.67	16.67	8.33	11.11	A1		31.02	47.66
40	H	30.00	61.67	26.67	15.00	4.44	A1		68.97	32.97
41	M	71.67	36.67	26.67	28.33	20.00	A1	X		
42	M	48.33	26.67	35.00	33.00	20.00	A2		69.98	65.49
43	M	83.33	12.00	15.00	11.67	12.00	A2	X		
44	M	30.00	21.67	15.00	12.00	12.00	A1	X		
45	M	53.33	50.00	30.00	20.00	15.56	A1	X		
46	M	56.67	36.67	46.67	31.67	17.78	A2		25.98	45.01
47	M	34.29	35.00	35.00	28.33	15.00	A1		55.68	73.66
48	M	63.33	30.00	45.00	61.67	31.11	A2		65.28	82.01
49	M	50.00	38.33	85.00	8.33	1.67	A1		14.91	47.10
50	M	38.26	31.11	20.00	10.00	28.89	A1		54.85	46.30
51	M	68.89	65.00	21.67	32.00	20.00	A1	X		
52	M	67.50	57.50	41.67	41.67	30.00	A1		47.25	62.08
53	M	53.33	15.00	23.33	18.00	16.00	A2	X		
54	M	83.33	68.33	23.33	20.00	23.33	A1		12.94	76.85
55	M	63.33	40.00	53.33	43.33	40.00	A2		32.14	76.49
56	M	61.67	45.00	25.00	25.00	18.33	A1		26.53	35.33
57	M	81.67	40.00	33.33	25.00	31.67	A1		55.70	52.91
58	M	55.00	45.00	33.33	31.67	28.33	A1		1.39	32.29

Table S1 continued

			Baseline						After	
ID	**Inb.**	**Control**	**A1**	**A2**	**A3**	**A4**	**A(s)**	**Ext.**	**Benign**	**Stress**
59	M	60.00	38.33	38.00	33.33	31.67	A1		24.94	34.30
60	M	48.33	38.33	41.67	23.33	13.33	A2		57.09	42.46
61	M	2.22	15.00	12.00	10.00	10.00	A1		21.32	84.82
62	M	78.33	45.00	41.67	8.33	2.22	A1		54.30	56.51
63	M	63.33	25.00	26.67	13.33	0.00	A2		34.82	40.42
64	M	65.00	46.67	33.33	31.67	21.67	A1	X		
65	M	43.33	26.67	16.67	16.67	11.67	A1		36.92	43.34
66	M	75.83	75.00	20.00	23.33	18.33	A1	X		
67	M	81.67	53.33	20.00	15.00	16.67	A1		32.36	63.26
68	M	65.00	85.00	14.00	10.00	12.00	A1	X		
69	M	71.67	75.00	41.67	16.67	10.00	A2		83.59	32.02
70	M	65.00	48.33	35.00	20.00	16.67	A1	X		
71	M	78.33	58.33	41.67	31.67	15.00	A1	X		
72	M	38.33	26.67	24.00	12.00	8.00	A1		27.94	45.62
73	M	41.67	63.33	31.67	28.33	15.00	A1	X		
74	M	28.33	28.33	33.33	31.67	26.67	A2		69.51	66.87
75	M	60.00	51.67	30.00	8.33	3.33	A1	X		
76	M	20.00	18.33	18.00	15.00	5.00	A1	X		
77	M	60.00	53.33	35.00	30.00	8.33	A1		44.84	61.10
78	M	85.00	75.00	10.00	16.67	16.67	A1	X		
79	M	33.33	51.67	57.33	18.33	13.33	A1		51.53	70.29
80	M	26.67	15.00	10.00	0.00	1.67	A1		15.33	53.13
81	L	76.83	71.67	50.00	48.00	30.00	A2		63.42	50.94
82	L	38.33	76.67	20.00	15.00	10.00	A1		10.71	79.21
83	L	78.33	76.67	16.67	18.33	16.67	A1		64.82	88.59
84	L	33.33	20.00	16.67	11.67	11.67	A1		25.59	67.92
85	L	78.33	86.67	13.00	3.84	11.67	A1		82.46	45.58
86	L	73.33	61.67	31.67	26.67	8.33	A1		85.03	70.47
87	L	16.67	40.00	28.33	39.00	30.00	A1		41.59	96.10
88	L	41.67	31.67	38.33	20.00	23.33	A2		28.71	74.96
89	L	75.83	30.00	48.33	25.00	13.33	A2	X		
90	L	51.67	61.67	36.67	33.00	33.33	A1		38.40	59.18
91	L	55.00	41.67	28.33	25.00	11.67	A1		35.36	89.18
92	L	63.33	61.67	17.50	18.33	8.33	A1		34.99	19.62
93	L	53.33	17.50	26.67	23.33	23.33	A2		38.40	86.90
94	L	35.00	26.67	25.00	23.33	12.00	A1		37.87	53.03
95	L	57.50	41.67	38.33	36.67	20.00	A1		53.83	48.82
96	L	33.33	38.33	17.50	28.33	30.00	A1		41.38	81.48
97	L	40.00	41.67	15.00	10.00	5.00	A1		67.95	21.13

Table S1 continued

			Baseline						After	
ID	**Inb.**	**Control**	**A1**	**A2**	**A3**	**A4**	**A(s)**	**Ext.**	**Benign**	**Stress**
98	L	58.33	51.67	35.00	31.67	18.33	A1		73.44	72.95
99	L	60.00	43.33	55.00	33.33	43.33	A2		67.17	70.28
100	L	20.00	38.33	46.67	35.00	37.42	A2		65.73	67.41
101	L	41.67	15.00	12.00	12.00	14.67	A1		41.39	18.20
102	L	70.00	51.67	31.67	25.00	10.00	A1		21.00	86.70
103	L	56.67	50.00	55.00	31.67	13.33	A1		98.40	98.33
104	L	50.00	50.00	45.00	33.33	21.67	A1		96.08	74.42
105	L	66.67	38.00	60.00	28.33	38.33	A2		43.72	49.84
106	L	63.33	30.00	21.67	16.67	26.67	A1		82.25	81.33
107	L	53.33	50.00	23.33	16.67	6.67	A1		67.45	89.42
108	L	46.67	43.33	45.00	25.00	13.33	A2		38.87	92.36
109	L	66.67	35.78	36.67	21.67	5.00	A2		63.27	50.45
110	L	44.17	30.00	23.33	18.33	5.00	A1		83.24	39.27
111	L	80.00	53.33	43.33	20.00	21.67	A1		5.95	35.67
112	L	90.00	53.33	45.00	35.00	15.77	A1		54.59	81.37
113	L	56.67	48.33	46.67	35.00	15.00	A1		29.10	96.52
114	L	66.67	46.67	18.33	33.33	11.67	A1		6.35	60.01
115	L	76.67	65.00	26.67	21.67	8.33	A1		29.70	52.20
116	L	43.33	43.33	45.00	23.33	13.33	A2		37.33	48.80
117	L	60.00	38.33	46.67	25.00	4.44	A2	X		
118	L	41.67	28.33	15.00	23.33	4.33	A1		57.40	52.47
119	L	73.33	18.33	46.67	36.67	5.00	A2		39.49	57.59
120	L	43.33	31.57	31.67	23.33	11.67	A2	X		
121	L	38.33	48.33	16.67	35.00	24.44	A1		44.41	82.29
122	L	71.67	63.33	28.33	23.33	23.33	A1	X		
123	L	76.67	51.67	43.33	25.00	16.67	A1	X		
OB1	N	88.33	71.67	61.67	26.67	11.67	A2		69.18	82.95
OB2	N	93.33	76.67	60.00	16.67	6.67	A2		88.65	96.24
OB3	N	86.67	70.00	65.00	31.67	26.67	A2		83.40	78.48
OB4	N	81.67	70.00	66.67	28.33	25.00	A2		97.00	83.85
OB5	N	85.00	73.33	65.00	25.00	26.67	A2		87.34	74.72
OB6	N	81.67	66.67	60.00	26.67	23.33	A2		74.70	97.83
OB7	N	86.67	66.67	53.33	45.00	26.67	A2		60.58	86.11
OB8	N	80.00	73.33	58.33	30.00	26.67	A2		88.00	60.44
OB9	N	81.67	68.33	55.00	31.67	25.00	A2		76.76	85.24
OB10	N	83.33	76.67	58.33	30.00	23.33	A2		83.02	60.97

Supplementary Table S2. Line mean dry body mass for generations 1-10. This table is supplied online: https://goo.gl/tZUjgx. Line mean dry body mass from 123 lines of the three different inbreeding (F) levels (High; H, Medium; M, and Low; L) plus the outbred lines (OB). Mean, n, and SD is shown for generations 1-10.

Supplementary Table S3. Line mean productivity for generations 1-10. Line mean productivity from 123 lines of the three different inbreeding (F) levels (High; H, Medium; M, and Low; L) plus the outbred lines (OB) for generations 1-10. For productivity, we only have one measure per line per generation, so SD cannot be determined.

		Generation									
ID	**F**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	H	0.489	0.774	3.872	0.402	5.871	0.823	2.861	1.576	0.670	4.201
2	H	0.209									
3	H	4.722	0.273	0.469							
4	H	1.185	3.854	1.201	1.819	6.641	0.623	5.731	0.720	8.404	8.806
5	H	3.664	4.466	2.072	0.341	1.363	3.414	0.832	0.977	5.711	0.635
6	H	6.585	4.675	2.570	8.673	4.013	5.542	0.977	0.471	2.690	1.036
7	H	0.445	2.439	0.724	0.308	0.755	6.014	9.318	0.496	8.452	0.982
8	H	1.295	2.333	0.414	0.673	0.904	0.733	0.722	0.496	2.334	6.975
9	H	0.585	1.059	2.067	0.211	1.158	7.658	0.441	0.655	2.174	5.677
10	H	2.885	0.628	0.974	0.146	8.376	0.591	3.067	8.461	0.416	0.856
11	H	2.371	5.376	0.892	7.131	0.642	10.163	4.070	10.465	1.446	3.477
12	H	0.516	0.654	4.607	0.194	0.072					
13	H	0.039									
14	H	0.367	0.258								
15	H	3.804	0.516	5.538	1.750	0.322	0.452	1.464	0.460	0.995	0.720
16	H	0.903	5.916	8.031	0.160	0.699	1.717	0.885	1.486	2.159	0.832
17	H	0.861	0.357								
18	H	4.835	2.072	1.164	0.123	0.225	3.101	0.754	9.787	2.799	11.912
19	H	1.181	0.155								
20	H	3.076	2.689	2.404	0.624	4.567	1.752	8.033	0.708	1.223	1.397
21	H	1.693	0.124								
22	H	0.793	0.375	1.507	4.380	0.514	2.442	1.510	9.720	6.769	0.878
23	H	5.988	0.562	0.909	0.441	0.297	1.460	1.456	6.192	3.172	1.041
24	H	0.186									
25	H	1.298	0.727	0.568	4.062	0.508	0.928	9.339	0.734	0.676	4.467
26	H	1.735	1.473	3.187	0.911	4.998	0.478	5.583	5.388	0.544	0.731
27	H	0.617									
28	H	1.084	1.761	0.718	2.281	0.907	0.008				
29	H	0.021									
30	H	5.082	3.124	1.077	0.426	2.212	3.708	0.708	0.982	0.986	1.613
31	H	1.006									
32	H	0.656									
33	H	0.538	0.880	2.628	0.410	7.938	3.857	8.330	2.287	0.612	4.315

Table S3 continued

		Generation									
ID	**F**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
34	H	0.398	3.410	2.567	6.442	8.261	0.897	0.643	2.347	8.372	2.823
35	H	0.542	2.963	0.445	0.601	1.956	1.233	2.872	0.967	4.790	0.725
36	H	0.093									
37	H	0.577	2.307	0.023							
38	H	0.376	0.332								
39	H	0.994	0.847	2.523	2.227	7.992	5.146	0.964	3.314	0.778	0.525
40	H	5.514	1.284	7.194	7.645	2.762	3.265	0.509	3.551	3.009	1.683
41	M	4.409	0.540								
42	M	4.450	6.311	6.678	0.458	7.880	5.314	2.610	0.960	0.601	1.278
43	M	2.503	0.591								
44	M	1.782	0.827	10.221	8.071	1.363	0.212				
45	M	0.458	5.137	0.433	0.514						
46	M	7.335	11.018	8.564	0.473	8.497	4.565	6.290	0.980	6.454	9.625
47	M	9.080	2.279	8.384	8.357	3.940	2.938	10.425	4.552	1.734	0.952
48	M	1.168	3.413	8.914	7.606	3.338	9.497	11.959	10.312	7.506	2.696
49	M	1.163	5.523	1.881	4.581	0.600	3.274	12.019	4.727	10.167	3.494
50	M	7.161	0.329	7.493	1.091	1.764	7.617	0.614	10.542	3.027	5.685
51	M	4.394	0.917								
52	M	0.192	9.136	0.638	8.059	9.119	0.400	14.578	12.997	5.922	14.221
53	M	7.362	0.724								
54	M	6.719	2.174	7.808	3.679	7.771	16.834	0.814	5.321	1.410	0.749
55	M	4.864	1.593	4.226	10.455	2.399	1.664	13.622	10.522	12.236	3.387
56	M	2.595	8.595	1.888	5.270	2.136	3.060	3.313	0.674	1.182	4.087
57	M	6.408	6.324	6.818	9.622	9.307	10.325	4.459	7.708	3.421	5.415
58	M	0.140	2.628	5.232	4.447	6.548	2.767	0.904	4.376	5.374	2.050
59	M	5.810	0.811	2.226	0.902	0.601	8.073	2.323	4.032	2.257	4.657
60	M	1.719	0.896	0.888	15.764	3.679	12.022	1.793	8.768	7.186	10.143
61	M	5.747	14.984	4.077	1.804	4.959	8.260	6.496	6.265	9.971	3.258
62	M	4.081	8.974	4.150	2.568	3.844	3.763	6.718	6.007	8.495	3.902
63	M	2.270	5.231	0.428	4.352	14.362	2.559	1.993	0.451	9.362	3.853
64	M	0.520	2.109								
65	M	6.515	1.523	1.693	4.410	2.174	1.557	5.473	11.812	10.068	4.926
66	M	5.139	0.601								
67	M	6.421	0.724	2.531	5.373	8.352	0.710	7.544	15.007	0.892	12.162
68	M	5.657	5.316	0.643							
69	M	7.440	9.188	4.816	7.521	15.007	1.359	5.530	5.871	14.227	16.040
70	M	5.685	1.829								
71	M	1.408	3.788								
72	M	6.755	8.613	2.256	10.318	9.070	7.433	10.269	0.508	5.810	9.499
73	M	1.124									
74	M	3.108	8.536	3.407	1.145	3.632	7.146	4.938	4.420	0.852	12.283
75	M	7.225	3.516	0.771	5.446	5.507					
76	M	0.795	1.316	1.916	2.807	3.043	2.448	0.747	1.558		
77	M	7.504	0.964	1.105	4.118	14.262	0.839	1.620	14.966	1.884	13.486
78	M	0.628	0.740								
79	M	8.698	10.927	2.165	5.807	3.721	0.048	1.424	7.068	5.306	12.012
80	M	5.018	5.630	6.047	1.007	2.558	10.178	11.745	0.710	4.992	13.183
81	L	3.337	8.982	5.128	5.816	6.193	3.809	4.716	7.278	4.568	3.651
82	L	3.354	3.284	7.194	3.690	6.966	5.394	5.225	3.378	9.322	8.570

Table S3 continued

		Generation									
ID	**F**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
83	L	2.670	5.537	3.256	3.649	4.367	9.398	4.464	4.829	4.340	2.843
84	L	3.069	4.664	8.861	5.776	4.347	7.374	5.274	7.558	4.484	5.333
85	L	2.981	4.958	3.814	2.456	3.380	3.856	8.374	5.141	4.943	4.441
86	L	2.905	4.763	6.818	4.979	7.366	2.903	3.582	4.372	5.403	3.754
87	L	3.762	4.650	3.581	9.774	4.645	5.100	3.699	3.706	5.001	11.558
88	L	3.396	5.174	5.668	6.331	8.418	10.106	4.543	3.909	3.981	3.659
89	L	3.206	2.657	0.087							
90	L	2.996	4.849	8.818	7.382	4.240	6.192	2.964	11.094	4.132	8.535
91	L	3.095	1.189	4.518	8.948	5.564	3.631	8.999	6.109	3.945	3.030
92	L	4.107	4.978	5.693	3.780	3.351	4.312	3.163	11.800	8.380	3.106
93	L	3.571	4.369	5.036	5.700	11.735	9.360	9.436	4.131	11.578	3.392
94	L	4.266	5.633	4.839	8.780	2.746	10.685	4.446	4.917	6.886	3.058
95	L	5.452	5.207	7.000	8.234	6.073	7.464	3.884	4.173	3.774	3.381
96	L	4.090	8.108	5.908	3.748	5.178	5.242	5.619	4.708	12.577	2.941
97	L	4.945	5.649	7.000	5.498	4.529	5.946	4.750	6.731	4.078	3.722
98	L	3.461	4.236	4.019	5.071	8.807	6.381	8.074	6.451	5.038	3.074
99	L	3.103	4.050	6.333	3.472	5.185	9.926	6.833	6.025	6.966	11.286
100	L	5.755	5.888	4.221	6.837	6.128	5.768	5.609	6.887	6.693	4.066
101	L	3.203	4.247	5.067	2.380	6.708	4.691	2.997	4.829	6.749	5.850
102	L	4.276	5.565	3.531	7.709	6.927	3.341	8.584	5.968	5.804	12.027
103	L	5.041	4.627	3.960	3.056	3.109	9.457	4.132	4.404	3.960	3.802
104	L	3.173	4.209	5.116	4.568	6.207	4.963	5.717	8.782	10.482	5.119
105	L	3.389	5.186	6.165	6.301	3.218	4.721	8.887	10.919	8.156	12.673
106	L	4.001	5.531	7.217	8.611	8.380	11.843	9.401	9.770	13.303	12.685
107	L	3.535	7.971	3.265	4.319	9.744	5.388	3.118	4.357	5.078	7.839
108	L	3.841	2.218	5.613	3.521	3.453	6.162	3.247	3.755	12.140	3.351
109	L	3.152	4.167	6.865	6.392	4.087	5.174	4.639	5.442	5.049	11.082
110	L	4.731	4.564	8.000	5.161	10.516	6.889	6.028	3.332	3.824	7.812
111	L	2.917	4.952	4.600	5.580	5.458	6.295	9.878	3.500	7.408	3.046
112	L	3.150	8.268	8.536	7.647	3.613	6.589	10.017	4.753	3.969	11.021
113	L	3.182	4.781	6.203	4.071	3.506	4.703	2.998	3.852	5.214	4.234
114	L	2.909	4.359	3.760	6.594	6.609	8.322	4.184	8.792	3.755	6.226
115	L	3.568	4.814	3.516	6.323	7.095	11.007	5.084	10.279	4.707	6.939
116	L	2.765	5.794	3.029	4.565	4.693	8.528	4.948	7.199	6.430	8.647
117	L	4.217	0.250								
118	L	2.157	5.324	3.897	6.233	3.806	5.009	4.573	10.000	4.864	3.720
119	L	3.845	4.486	5.642	5.070	8.436	3.459	4.600	3.738	5.555	3.531
120	L	2.846	5.131	3.532	8.147	2.762					
121	L	3.093	4.217	5.281	4.283	2.855	4.779	11.308	8.100	4.803	4.942
122	L	3.010	4.253	2.055							
123	L	2.056									

Table S3 continued

		Generation									
ID	**F**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
123	L	2.056									
OB1	N	4.828	5.784	7.272	11.588	8.520	5.604	5.464	10.685	8.685	5.838
OB2	N	5.708	6.891	7.209	10.023	10.748	5.507	5.460	14.277	5.663	3.686
OB3	N	4.867	6.695	8.499	7.412	4.685	9.773	10.938	9.898	9.460	12.409
OB4	N	4.693	7.656	6.258	6.537	3.522	5.387	7.327	4.472	5.919	6.102
OB5	N	6.000	6.843	7.460	10.378	5.213	4.319	10.404	7.462	6.350	10.706
OB6	N	5.427	6.165	6.613	5.630	9.641	8.807	4.754	6.862	4.969	9.543
OB7	N	5.416	7.818	8.220	4.466	5.085	6.137	2.905	6.000	4.969	9.543
OB8	N	5.072	6.280	6.750	4.152	8.911	9.773	6.957	7.425	6.835	4.844
OB9	N	5.090	7.847	8.239	7.455	3.411	8.300	7.935	5.698	7.685	3.415
OB10	N	4.652	6.418	9.095	7.817	6.511	5.387	7.965	4.911	8.685	6.694

SUMMARY

When organisms are faced with changes in their environment, they are forced to respond, if they are to maintain optimal function. Especially ectotherms must deal with environmental changes in e.g. temperature on a regular basis, and thus their survival and reproductive success depend on their ability to respond on a behavioral, physiological, morphological and/or evolutionary level according to the environmental cues.

At the same time, if populations are small and fragmented, and have limited gene flow, environmental change and environmental stress might interact with intrinsic genetic stress such as inbreeding and genetic drift, which can exacerbate the effects of one or more environmental stresses. Furthermore, inbred populations often have low genetic variation that might constrain evolutionary responses to rapidly changing environments.

This thesis investigates how, and to what extent, insect model species respond to a multitude of different environmental stresses, how the environment interacts with the genetic composition of individuals, and lastly the consequences of low effective population size on the adaptive ability, and how to possible alleviate some of the negative fitness effects of inbreeding and loss of genetic variation by means of genetic rescue.

ISSN (online): 2446-1636
ISBN (online): 978-87-7210-088-3

AALBORG **UNIVERSITY PRESS**