

Micro- and macroevolutionary aspects of body size diversification and thermal adaptation in insects

JOHN WALLER

FACULTY OF SCIENCE | DEPARTMENT OF BIOLOGY | LUND UNIVERSITY



Micro- and macroevolutionary aspects of body size diversification and thermal adaptation in insects

John Waller



LUND
UNIVERSITY

DOCTORAL DISSERTATION

To be defended in the Blue Hall, Ecology Building, Sölvegatan 37,
Lund Sweden on Friday 26th January, 2018 13.00

Faculty opponent

Dr. Locke Rowe

Dean of school of graduate studies and vice-provost of
graduate research and education.

Distinguished professor of Ecology and Evolutionary Biology
University of Toronto, Canada

Organization LUND UNIVERSITY Department of Biology, Evolutionary Ecology, Sölvegatan 37, 223 62, Lund Author(s) John Waller	Document name DOCTORAL DISSERTATION	
	Date of issue	
	Sponsoring organization	
Title and subtitle Micro- and macroevolutionary aspects of body size diversification and thermal adaptation in insects		
<p>Abstract</p> <p>Body size and body temperature are the two most important traits in biology. In this thesis I show how these two variables have shaped the evolution of insects over the last 300 million years. These variables are so important in fact that three rules have been developed to summarize their effects: (1) bigger is better, (2) hotter makes you smaller, and (3) hotter is better.</p> <p>Using primarily dragonflies and damselflies, I explore the implications of these rules in the evolution of insects. Bigger is better (rule 1) states that larger individuals tend to have higher fitness. In insects larger adult body size is expected to produce higher fitness in terms of higher fecundity in females and higher mating success in males, while potentially increasing longevity in both males and females. I show that there are limits to how large one can get, and stabilizing selection body is expected to be the norm in most populations, mediated through conflicting selection between large adults and costs of increased development time.</p> <p>Hotter makes you smaller (rule 2), also known as the temperature-size rule in insects, states that within a population insects that develop in warmer regions will be smaller. The hotter makes you smaller rule is typically observed along a latitudinal gradient within the range of a species. This within species trend suggests strong associations between an organism size and its developmental environment, and suggests a possible resolution to the paradox imposed by the bigger is better rule. While the temperature-size rule is well-established in natural and laboratory populations, I show that other factors such as predation and dispersal are likely determining latitudinal size patterns between species.</p> <p>The final rule (rule 3), hotter is better, is the least well-established "rule" of the three rules. It states that warmer body temperatures will tend to lead higher performance in adult insects. Since dragonflies and damselflies are ectotherms, they are strongly dependent on ambient temperature to achieve optimal body temperatures, I show that hotter is not always better, and individuals with higher heating rates are favoured only in extreme temperature conditions, and otherwise thermal inertia is preferred, since the mean environment will match an individual's optimal body temperature.</p>		
Key words Body Size, Dragonfly, Damselfly, Evolution, Thermoregulation, Macro-Evolution, Calopteryx		
Classification system and/or index terms (if any)		
Supplementary bibliographical information		Language English
ISSN and key title		ISBN 978-91-7753-527-0 (print) ISBN 978-91-7753-528-7 (pdf)
Recipient's notes	Number of pages	Price
	Security classification	

I, the undersigned, being the copyright owner of the abstract of the above-mentioned dissertation, hereby grant to all reference sources permission to publish and disseminate the abstract of the above-mentioned dissertation.

Signature

Date 2017-12-12

Micro- and macroevolutionary aspects of body size diversification and thermal adaptation in insects

John Waller



LUND
UNIVERSITY

Coverphoto by Petro Buyak

Copyright John Waller

Faculty of Science
Department of Biology

ISBN 978-91-7753-527-0 (print)
ISBN 978-91-7753-528-7 (pdf)

Printed in Sweden by Media-Tryck, Lund University
Lund 2017



MADE IN SWEDEN

Media-Tryck is an environmental-
ly certified and ISO 14001 certified
provider of printed material.
Read more about our environmental
work at www.mediatryck.lu.se

Table of Contents

List of Papers.....	6
Author Contributions	7
Abstract	8
Popular Science Summary	9
Populärvetenskaplig Sammanfattning.....	10
General Introduction	11
Thesis Aims.....	12
Background	13
Ecology of dragonflies and damselflies.....	13
Body size, oxygen, and fossils	15
The comparative method	18
Evolution of thermal performance traits.....	21
Materials and Methods	23
Measuring selection in natural populations	23
Field work.....	24
Mark-recapture methods.....	27
Building the odonate phenotypic database	28
Odonate super tree.....	29
Major Results	30
Conclusions	32
Acknowledgments.....	33
References	35

List of Papers

- I.** **Waller**, J., & Svensson, E.I. (2016). The measurement of selection when detection is imperfect: How good are naïve methods?. *Methods in Ecology and Evolution*, 7(5), 538-548.
- II.** **Waller**, J., Willink B., Tschol M., & Svensson E.I. (2018). The Odonate Phenotypic Database: a new open data resource for comparative studies of an old insect order. www.odonatephenotypicdatabase.org. (Manuscript).
- III.** **Waller**, J., & Svensson, E.I. (2017). Body size evolution in an old insect order: No evidence for Cope's Rule in spite of fitness benefits of large size. *Evolution*.
- IV.** **Waller**, J. (2018). Is the blunderbuss a misleading visual metaphor for stasis and punctuated evolution? (Manuscript).
- V.** **Waller**, J., & Svensson, E.I. (2017). Biotic and abiotic drivers of global body size distributions and macroevolutionary diversification in an old insect order. (Manuscript).
- VI.** **Waller**, J., Kell, A., Ballesta, M., Giraud, A., Abbott, J., & Svensson, E. (2017). Limited genetic variation for male mating success reveals low evolutionary potential for thermal plasticity in *Drosophila melanogaster*. *bioRxiv*, 166801.
- VII.** **Waller**, J. & Svensson, E.I. (2018). Sexual and natural selection on thermal plasticity and thermal canalization in insects at high latitudes. (Manuscript).

Papers I & III are reprinted with permission of the publisher.

Other interesting papers published during thesis but not included:

1. Svensson, E. I., & **Waller**, J. T. (2013). Ecology and sexual selection: evolution of wing pigmentation in calopterygid damselflies in relation to latitude, sexual dimorphism, and speciation. *The American Naturalist*, 182(5), E174-E195.
2. Gosden, T. P., **Waller**, J.T., & Svensson, E. I. (2015). Asymmetric isolating barriers between different microclimatic environments caused by low immigrant survival. *Proceedings of the Royal Society of London B: Biological Sciences*, 282(1802), 20142459.
3. Svensson, E. I., Nordén, A., **Waller**, J. T., & Runemark, A. (2016). Linking intra-and interspecific assortative mating: Consequences for asymmetric sexual isolation. *Evolution*, 70(6), 1165-1179.

Author Contributions

- I. **Waller** and **Svensson** conceived the study. **Svensson** advised direction of study and assisted in framing manuscript in a broader context. **Waller** performed simulations and data analysis and writing of the manuscript.
- II. **Waller, Willink** and **Svensson** conceived, planned, and designed the database. **Tschol** contributed a significant amount of data to the database. **Waller, Willink, Tschol**, and **Svensson** contributed to the manuscript. **Waller** published the website.
- III. **Waller** and **Svensson** conceived and planned the study. **Waller** and **Svensson** wrote the manuscript. **Waller** conducted the data analysis and preparation.
- IV. **Waller** conceived the commentary, wrote the manuscript, and conducted the data analysis and simulations.
- V. **Waller** and **Svensson** conceived and designed the study. **Waller** and **Svensson** wrote and contributed to the manuscript. **Waller** prepared the data and conducted the data analysis.
- VI. **Waller, Svensson** and **Abbott** conceived and designed the study. **Kell, Ballesta**, and **Giraud** conducted experiments. **Waller, Svensson, Kell, Ballesta**, and **Abbott** contributed to the manuscript. **Waller, Kell**, and **Ballesta** prepared the data and conducted data analysis.
- VII. **Waller** and **Svensson** conceived and designed the study. **Waller** conducted the field experiments. **Waller** prepared and analyzed the data. **Waller** and **Svensson** wrote the manuscript.

Abstract

Body size and body temperature are the two most important traits in biology. In this thesis I show how these two variables have shaped the evolution of insects over the last 300 million years. These variables are so important in fact that three rules have been developed to summarize their effects: (1) bigger is better, (2) hotter makes you smaller, and (3) hotter is better.

Using primarily dragonflies and damselflies, I explore the implications of these rules in the evolution of insects. Bigger is better (rule 1) states that larger individuals tend to have higher fitness. In insects larger adult body size is expected to produce higher fitness in terms of higher fecundity in females and higher mating success in males, while potentially increasing longevity in both males and females. I show that there are limits to how large one can get, and stabilizing selection body is expected to be the norm in most populations, mediated through conflicting selection between large adults and costs of increased development time.

Hotter makes you smaller (rule 2), also known as the temperature-size rule in insects, states that within a population insects that develop in warmer regions will be smaller. The hotter makes you smaller rule is typically observed along a latitudinal gradient within the range of a species. This within species trend suggests strong associations between an organism size and its developmental environment, and suggests a possible resolution to of the paradox imposed by the bigger is better rule. While the temperature-size rule is well-established in natural and laboratory populations, I show that other factors such as predation and dispersal are likely determining latitudinal size patterns between species.

The final rule (rule 3), hotter is better, is the least well-established “rule” of the three rules. It states that warmer body temperatures will tend to lead higher performance in adult insects. Since dragonflies and damselflies are ectotherms, they are strongly dependent on ambient temperature to achieve optimal body temperatures, I show that hotter is not always better, and individuals with higher heating rates are favoured only in extreme temperature conditions, and otherwise thermal inertia is preferred, since the mean environment will match an individual’s optimal body temperature.

Popular Science Summary

For every type of animal there is an optimal size, but bigger is often better. 300 million years ago giant crow-sized dragonflies were the largest insects to ever exist. Higher atmospheric oxygen concentrations are thought to have allowed them to both fly and develop into giants. But today the largest dragonflies are no larger than a human hand and take decades to develop into adults. But why become crow-sized or hand-sized in the first place when there are significant risks to developing for decades?

Larger adult insects tend to survive longer, lay more eggs, or have more sex. If bigger is better, why no super giant dragonflies today? Even if dragonflies raised in chambers with high oxygen only become slightly larger. Long development times, is the likely answer, and being eaten as a child is a strong incentive to mature quickly into an adult.

Larger body size means larger muscles, more heat production, and a reduced surface-volume-ratio. Even under ideal conditions, the body temperatures of the largest extinct dragonflies may have exceeded 50 C. How super giant dragonflies were able to fly and not boil their insides is still an open question. If there is an optimal body size, there is also an optimal body temperature.

Surface-to-volume ratio is believed to be responsible for patterns in mammals like Bergman's rule, but for insects there is another rule called the temperature-size rule: insects raised at higher temperatures tend to be smaller. You might be thinking that the temperature-size rule sounds very similar to Bergman's rule (getting bigger as it gets colder), and dragonflies in the north tend to be larger than dragonflies in the south. But the reasons why mammals and dragonflies have larger body size in colder temperatures are different. For dragonflies and other insects, warmer temperatures lead to faster development times and smaller sizes. The temperature size-rule, however, is probably not enough to explain the distribution different sized dragonfly species. Or why large species are more likely to be found at temperate latitudes though. In chapter V, I explain why increased threats from birds and greater dispersive ability by larger species might explain why larger dragonflies have been more likely to move out of the tropics.

Over-heating is not the only challenge to being large. Dragonflies and other insects also rely on sunlight and outside temperatures to warm up on cold days, and larger body size will decrease the ability of an individual to reach an optimal body temperature quickly. Natural selection should favour individuals that find a balance between development time, heating gain/loss, and fitness.

Populärvetenskaplig Sammanfattning

För varje typ av djur finns en optimal storlek, men större är ofta bättre. För 300 miljoner år sedan var jättelika drakeflugor de största insekterna någonsin. Högre atmosfäriska syrekonzcentrationer tros ha tillåtit dem att både flyga och utvecklas till jättar. Men idag är de största sländorna inte större än en mänsklig hand och tar decennier att utvecklas till vuxna. Men varför blir kråkmått eller handstorlek i första hand när det finns betydande risker för att utvecklas i årtionden?

Större vuxna insekter tenderar att överleva längre, lägga mer ägg eller ha mer sex. Om större är bättre, varför är ingen super jätte libell idag? Även om sländor som lyfts upp i kamrar med högt syre blir de bara lite större. Långa utvecklingstider, är det troliga svaret, och att ätas som ett barn är ett starkt incitament för att mogna snabbt till en vuxen.

Större kroppsstorlek betyder större muskler, mer värmeproduktion och minskat ytvolymförhållande. Även under idealiska förhållanden kan kroppstemperaturerna hos de största utdöda sländorna ha överskridit 50 C. Hur super jätte sländor kunde flyga och inte koka sina insidor är fortfarande en öppen fråga. Om det finns en optimal kroppstorlek finns det också en optimal kroppstemperatur.

Förhållandet mellan ytan och volymen antas vara ansvarig för mönster i däggdjur som Bergmans regel, men för insekter finns en annan regel som kallas temperaturstorregeln: insekter som höjas vid högre temperaturer tenderar att vara mindre. Du kanske tänker på att temperaturstorregeln låter mycket lik Bergmans regel (blir större när det blir kallare), och sländor i norr tenderar att vara större än sländor i söder. Men anledningarna till att däggdjur och sländor har större kroppsstorlek vid kallare temperaturer är olika. För sländor och andra insekter leder varmare temperaturer till snabbare utvecklingstider och mindre storlekar. Temperaturstorregeln är emellertid förmögeligen inte tillräcklig för att förklara fördelningen av olika storlekar av dragonflyarter. Eller varför stora arter är mer benägna att hittas vid tempererade breddgrader. I kapitel V förklarar jag varför ökade hot från fåglar och större spridningsförmåga av större arter kan förklara varför större sländor har varit mer benägna att flytta ur troperna.

Överuppvärming är inte den enda utmaningen att vara stor. Dragonflies och andra insekter är också beroende av solljus och utetemperaturer för att värma upp på kalla dagar, och större kroppsstorlek minskar individens förmåga att nå en optimal kroppstemperatur snabbt. Naturligt urval bör gynna individer som finner en balans mellan utvecklingstiden, värmeförstärkning / förlust och fitness.

General Introduction

Kingsolver and Huey (2008) wrote the following three-rule haiku for selection and life history in insects:

bigger is better

hotter makes you smaller

hotter is better

It is not surprising that three rules can explain a lot of insect biology, since body size and temperature are two of the most important variables in biology, and size and temperature also interact and influence the evolution of the other. In this thesis, I will explore the implications of these rules in the evolution of insects.

Bigger is better (rule 1) states that larger individuals tend to have higher fitness. In insects larger adult body size is expected to produce higher fitness in terms of higher fecundity in females and higher mating success in males, while potentially increasing longevity in both males and females. However, there are, of course limits to how large one can get, and stabilizing selection on body size is expected to be the norm in most populations, although it is often not observed in nature (Kingsolver et al. 2001). Fossil lineages of many insects, such as dragonflies and damselflies, also show no net trend of increasing body size despite being good support for rule 1. How can we resolve the paradox of stasis in the fossil record and positive selection in natural populations? I will address these questions in **chapter III**.

Hotter makes you smaller (rule 2), also known as the temperature-size rule in insects, states that within a population insects that develop in warmer regions will be smaller. The hotter makes you smaller rule is typically observed along a latitudinal gradient within the range of a species. This within species trend suggests strong associations between an organism's size and its developmental environment, and suggests a possible resolution to the paradox imposed by the bigger is better rule. While the temperature-size rule is well-established in natural and laboratory populations, it is unknown whether this rule will create large-scale geographic patterns between species. In **chapter V**, I look at whether temperature or some other factor primarily shapes body size distributions in dragonflies and damselflies.

The final rule (rule 3), hotter is better, is the least well-established “rule” of the three rules. It states that warmer body temperatures will tend to lead higher performance in adult insects. Since insects are ectotherms, they are strongly dependent on ambient temperature to achieve optimal body temperatures, and it is generally the case that individuals within a population with higher body

temperatures will have better performance. Since the rate at which an insect can heat up (say after a cold night) is largely dependent on its body size and mass, rule 1 (bigger is better) and rule 3 (hotter is better) might seem to be opposed to one another. This creates a situation where the winning rule might depend on the environmental context (namely the ambient temperature). In the next section, I will outline how this thesis addresses these three rules of insect evolution and resolves some of the paradoxes these 3 rules create.

Thesis Aims

I address the complex relationship between body size and body temperature in ectothermic insects. Although the three rules of Kingsolver and Huey (2008) are fairly well established within populations, less research has been done on how these rules interact, and what are the broad patterns between species and across evolutionary time. While population-level biology is useful for understanding ecology, behavior, and selection, there is often a question as to whether the results of any population-level study are applicable over longer macroevolutionary time scales (Arnold *et al.* 2001). Studies attempting to connect macroevolution and population biology are rarely done in practice. However, such comparisons can be informative because what appears to be a paradox at one level can perhaps be resolved at another level. In this thesis, there will be several examples where micro- and macro-evolution inform each other.

All good research starts with methods development, and in **chapter I**, I present important developments in estimating selection on survival when recapture probabilities are not perfect. A large portion of my PhD studies, was spent measuring survival for small insects that were not easy to find (or recapture) in the field, and I address how I was able to measure survival selection, even when I was not completely sure an individual I was following was dead. This paper is published along with an R package **EasyMARK**, which allows for easier processing of capture history data.

From the cover art and article titles, it is obvious that almost all of this thesis is about dragonflies and damselflies. In **chapter II**, I present the **Odonate Phenotypic Database** (opdb)(Odonata being the name for the Order), which was a collaborative project between myself and the co-authors. The opdb is a resource for dragonfly and damselfly phenotypes. The database consists of a variety of morphological, behavioural, and bio-geographical information collected from various sources. The database is not intended for species identification, but for comparative analysis within the group or to be combined with data from other groups. The database is open to use to public and can be found at odonatephenotypicdatabase.org.

In **chapter III**, I integrate field data and phylogenetic comparative analyses to investigate the causes of body size evolution and stasis in dragonflies and damselflies. I address the paradox of if bigger is better, why is there widespread stasis in the fossil record? I find that there is strong evidence for bigger is better in the adult stage in dragonflies and damselflies, and I suggest that the lack of Cope's rule in this group results from conflicting selection between fitness advantages of large adult size and costs of long larval development. In **chapter IV**, I make a small commentary on Uyeda et al. (2011), and the misuse of the visual metaphor of the blunderbuss as a symbol of stasis and punctuated evolution.

In **chapter V**, I look into the factors influencing geographic body size patterns in dragonflies and damselflies. I find that the temperature-size rule (hotter is smaller), might not explain global geographic size differences between species. I find that temperature is not the primary force driving latitudinal body size trends in this group, and that ecological factors such as bird diversity and dispersal ability may play a larger role in shaping geographic body size patterns.

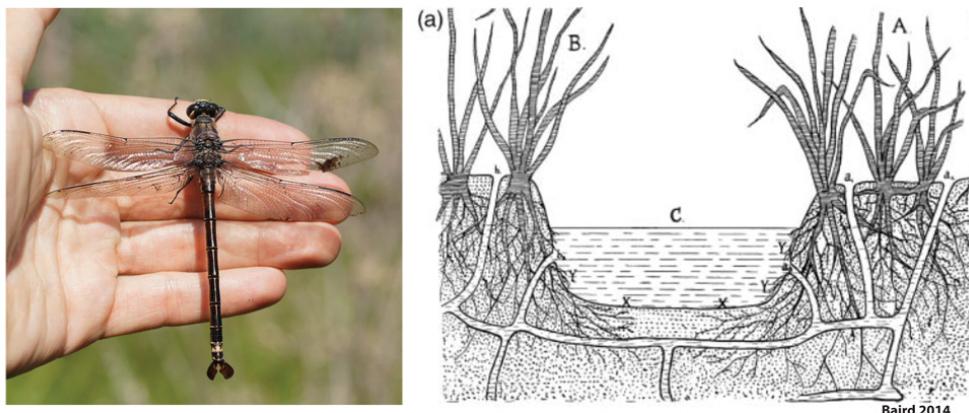
In **chapter VI**, I investigate the evolutionary potential of phenotypic plasticity in the thermal performance in fruit flies. I find that genetic variance in thermal performance curves might be low for male mating rate, making it unlikely that sexual selection through male mating success would be efficient in mediating evolutionary rescue through changed plasticity in response to changing temperatures. Finally in **chapter VII**, I examine the “hotter is better” rule in damselflies. In the study, I find that temperature extremes have a negative effect on survival and mating rate in our two studies species. I find that sexual selection on thermal plasticity increases in both high and low temperatures, implying that increased heating rate (hotter is better) is only adaptive in non-optimal temperature environments.

Background

Ecology of dragonflies and damselflies

Dragonflies and damselflies (odonates) are insects with complex life cycles, wherein growth and reproduction are divided roughly into four stages: 1. Egg 2. Larval 3. Immature 4. Adult. For simplicity it is helpful to informally call stages 1 and 2 the **larval stage** and stages 3 and 4 the **adult stage** (Stoks and Cordoba-Aguilar 2012). Events and conditions in the larval stage have been shown to have carry over effects and morphological and fitness consequences for the adult stage (Corbet 1999; Stoks and Cordoba-Aguilar 2012). In **chapter III**, I discuss how it is likely larval development time that counteracts the “bigger is better” rule in odonates.

The aquatic larval stage (of several molts) is specialized for fast growth, and can last months and years in some species. This is considerably longer than the adult stage, which usually lasts a few days or months. The largest dragonflies of the genus *Petalura* (200 mm body length) are purported to have larval stages lasting decades, uniquely living in burrows, which are probably predator defenses (Baird 2014)(Fig. 1).



This is the **largest extant dragonfly**. It has a development time of up to **10 years**.

It is also one of the only that **burrows** in the larval stage.



Figure 1. The Largest dragonfly.

The largest extant dragonfly of the genus *Petalura*. Known for having a long larval period of up to 10 years and being one of the few dragonflies that lives in burrows.

Growing larvae are subject to predation, parasites, and cannibalism. Larval activity increases mortality from predators (Mcpeek 2004), and time-constrained larvae increase their activity rates (Johansson and Rowe 1999). Larvae are also more active in turbid (cloudy) water than in clear water when fish predators present (van de Meutter *et al.* 2005). Gut parasites and water mites have been shown to reduce mass and adult survival (Braune and Rolff 2001; Tsubaki and Hooper 2004). The longer larvae spend underwater, the higher the probability of infection or predation (Corbet 1999).

The winged adult stage is specialized for dispersal and reproduction. Males can have territorial or opportunistic mating systems (scramble competition). Territories can be important for plant substrates and flowing water, as egg survival usually increases in areas with higher flow rate (Siva-Jothy *et al.* 1995). Sexual selection and visual signaling are important in many dragonflies and damselflies (Tynkkynen *et al.* 2005; Svensson *et al.* 2007; Svensson and Waller 2013). Some species of the genus *Calopteryx* have dark wing pigment, which has been shown to

be important for immunity function signaling (Siva-Jothy 2000). In **chapter VII**, I show that dark wing patches are positively related to lifespan and mating success in males.



Figure 2. Wing patch variation.

Variation in wing patch darkness in *Calopteryx splendens* males. Wing patches in this species likely signal condition and species identity to females. In **chapter VII**, I show that they are also positively related to heating rate.

Associations between wing pigmentation and male parasite resistance have been interpreted as supporting indicator models, or “good genes” models (Siva-Jothy 2000; Rantala *et al.* 2000). Other sexual selection studies indicate that wing pigmentation is important in male-male interference competition (Grether 1996; Tynkkynen *et al.* 2005), and it has also been demonstrated that pigmentation functions as a species recognition character and in sexual isolation, and shows a signature of latitudinal sorting, with the frequency of pigmented species higher in the temperate zone (Svensson & Waller 2013). In **chapter VII**, I show that males with darker wing patches also have higher heating rates.

Body size, oxygen, and fossils

In the fossil record, there are odonate (Protodonata) specimens with wing lengths of 350 mm (700 mm wingspans), which are comparable to the size of an adult crow (Harrison *et al.* 2010)(Fig. 3). Higher atmospheric oxygen levels 300 million years ago are thought to be in part responsible for the large body sizes seen in dragonflies and other insects (Harrison *et al.* 2010). This idea is also supported by experimental work on fruit flies (*Drosophila*), which if raised in experimentally

increased oxygen for multiple generations, evolve increased mean body size (Klok & Harrison 2009).

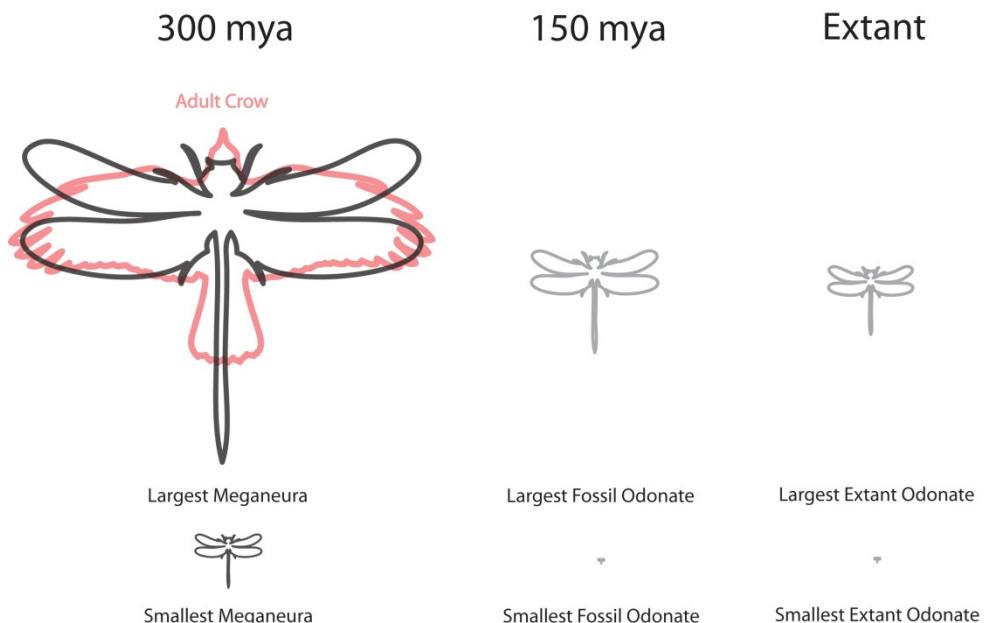


Figure 3. Variation in body size in the fossil record.

Variation in odonate body size found in the fossil record. It is thought that high oxygen content in the atmosphere may have allowed some ancient insects to grow to the size of small birds. However, the relationship between insect body size and oxygen is probably not as simple as once believed.

Figure 4 shows the oxygen percentage in the atmosphere over the last 350 million years (GEOCARBSULF model)(Berner 2009). The peak atmospheric oxygen in this reconstruction corresponds well to the peak seen in giant insects 300 million years ago (Clapham and Karr 2012). A 10% increase in oxygen concentration is thought to be especially important in the energetically intensive process of flight (Harrison *et al.* 2010). Body size also appears to be decoupled from oxygen levels after that initial peak (Clapham and Karr 2012), although other oxygen-record reconstructions might show a different pattern (Harrison *et al.* 2010).

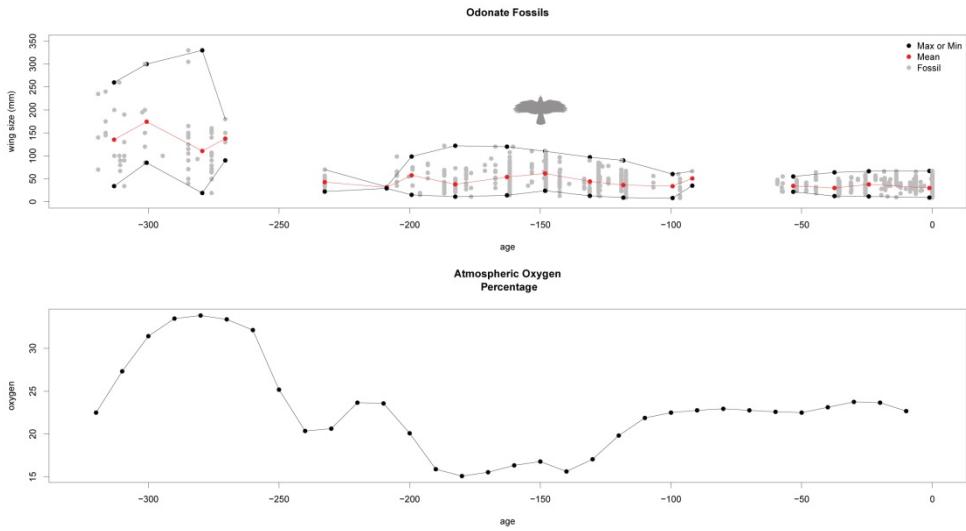


Figure 4. Wing size and oxygen.

The relationship between oxygen and maximum body size in odonates for the past 300 MYA. Oxygen levels predicts maximum body size towards the beginning of the time series (300 MYA), but after the evolution of birds (150 MYA), oxygen level and maximum odonate body size become decoupled.

In the fossil record, small and intermediate lineages have coexisted with the large ones. The response to high oxygen conditions is not uniformly to increase, and some laboratory studies on insects have shown that they actually became smaller in higher oxygen levels (Harrison *et al.* 2010). When oxygen is abundant it may allow insects to overcome some costly physiological constraints, and if growing larger has positive effects on survival, reproduction, or mating success, populations will evolve to become larger. No such evolutionary response would be expected, if there is no fitness benefits to being larger.

Lower atmospheric oxygen is not the only or even the most likely reason why maximum body size in odonates is presently smaller than in the past (Clapham and Karr 2012). Size-dependent predation by flying predators has been suggested as a constraint since 150 MYA (Clapham and Karr 2012), and it has been shown – in an extant population of two Calopterygid species (*Calopteryx splendens* and *C. virgo*) that larger individuals (longer wings) are indeed more likely to be predated by birds (Svensson and Friberg 2007). Odonates may also have occupied niches that are now occupied by vertebrates (Harrison *et al.* 2010). In chapter V, I show that bird diversity is negatively related to body size in present-day odonates even after accounting for other environmental variables, such as forest cover, temperature, and precipitation.

The comparative method

Often there are questions that cannot be answered by experimentation: When did a trait evolve? What is the rate of evolution? Are these two traits convergent? Why does this group have more species than another group? Phylogenetic comparative methods offer ways to address these questions and others (Pagel and Harvey 1988; Pagel 1999; Zsolt 2014).

Generally (although not exclusively), comparative methods use trait data from extant taxa and a phylogenetic tree to perform inferences or detect some pattern (Hansen and Martins 1996). Most comparative methods benefit from having a time-dated phylogenetic tree, which can usually be constructed from aligned genes and fossils after making some assumptions about the rate of mutation. Then typically some hierarchical clustering method is performed to produce a matrix of relatedness (Felsenstein 1984; Saitou and Nei 1987; Yang 1994). This relatedness and the divergence times are then in most cases incorporated into a covariance matrix of multi-variate normal distribution, and some inference is performed (Grafen 1989; Felsenstein 1997; Hansen and Martins 1996).

The inferred covariance matrix need not be perfect (i.e. it does not need to have a topology that exactly reflects evolutionary events) for non-independence to be adequately accounted for. Lacking a phylogeny, a taxonomy-level-based (family, genus, species) covariance matrix will usually suffice (Evans *et al.* 2012; Soul and Friedman 2015), although it is of course better to have a phylogeny that is as “true” as possible, although it should be emphasized that a phylogeny is always a hypothesis of the true underlying evolutionary relationships (Donoghue and Ackerly 1996). Uncertainty about divergence dates and topology can be incorporated if one takes a probabilistic and pragmatic perspective to tree building (Huelsenbeck *et al.* 2001; Pagel *et al.* 2004). While taxonomists strive to construct a fully dichotomous and trees with nodes that have low uncertainty, in some circumstances, a comparative biologist can be more pragmatic about uncertainty, since the exact relationship between taxa is not usually the main objective (Martins and Housworth 2002).

Modern probabilistic approaches to tree building can produce a posterior sample of trees, which can capture the uncertainty present in the topology and divergence dates (Drummond and Rambaut 2007). This sample of trees can then be used in any downstream comparative analysis. This will of course increase the uncertainty of any estimated parameters, but this uncertainty will reflect reality and might be acceptable. This pragmatic strategy allows one to make inferences and test models of evolution without having to wait for the “perfect phylogenetic tree.” It is this type of approach that I have employed in this thesis in **chapter III**.

For continuous characters, a simple model of evolution is a random walk process (Hansen 1997). A random walk can be modeled as draws from a normal distribution at each time step, which are then added (or subtracted) from the

previous time step. This process can be run several times, with each path taken representing a lineage drifting randomly though character space (Fig. 5). How quickly a trait evolves will depend on the rate of evolution, which is the variance of the normal distribution described above. The end result should produce a group of extant lineages with an expected phenotypic variance equal to the time multiplied by the rate (Fig. 5).

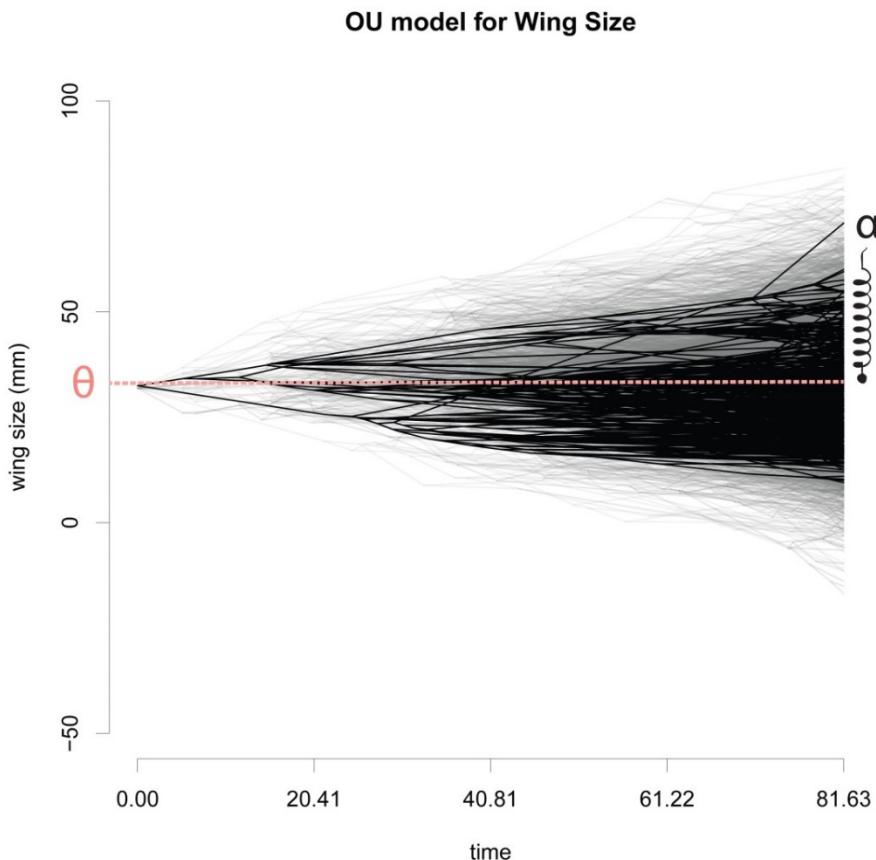


Figure 5. Brownian motion and pullbackforce.

An illustration of a Brownian motion (BM) and an Ornstein-Uhlenbeck (OU) process. BM is shown in light gray while the OU process is in black. The OU process is the same as the BM process except that an OU process includes a constraint terms (α) that acts as a spring pulling extreme values back closer to the lineage mean (θ) wing size.

This Brownian motion model of evolution assumes that there are no constraints on upper or lower bounds of this process. Constraints can however be modelled by adding an extra term to our model, which controls a pullback force as traits drift further away from the mean. This type of model is called Ornstein-Uhlenbeck (OU) model (Hansen 1997; Butler and King 2004).

For odonates, aside from the trivial constraint that body size cannot be negative, several constraints have been proposed on the upper limit of body size. About 300 mya and 150 mya odonates seem to have reached some local maxima in body size, according to the fossil record (Fig. 6). Since that time odonates have become smaller. This has led to the suggestion that birds are constraining maximum body size in this group, since they evolved around 150 mya. Another suggestion is that atmospheric oxygen levels are the limiting force (Harrison *et al.* 2010), since atmospheric oxygen levels were much higher about 300 mya (Fig. 5).

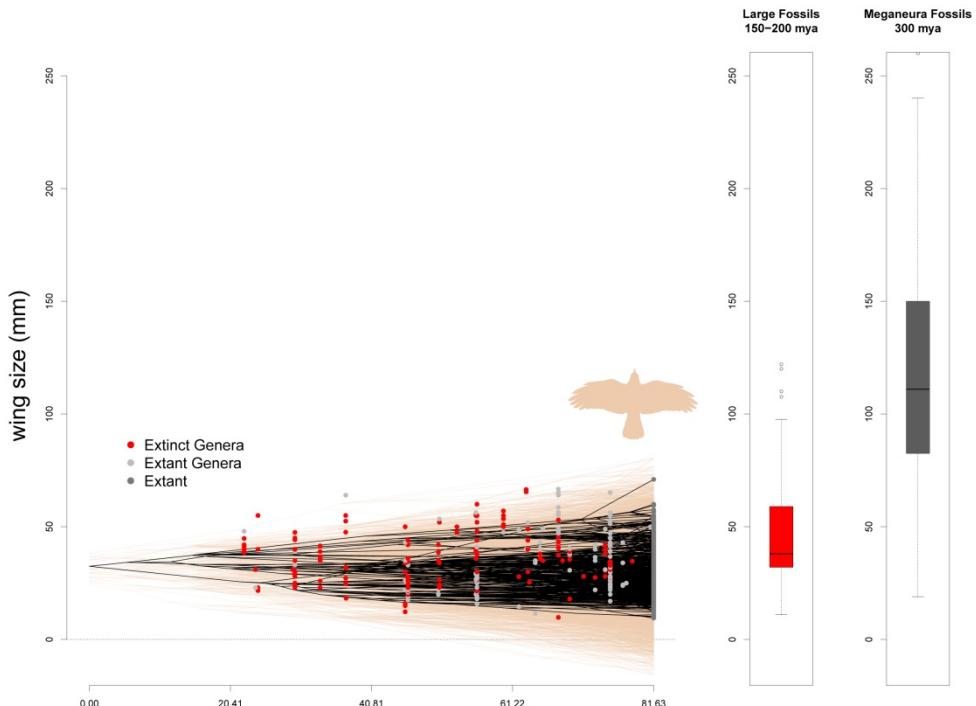


Figure 6. What is constraining body size in odonates?

Comparison of modern-age body sizes with ancient dragonflies and damselflies fossils. Here we see that even if constraints were relaxed (shown in pink) that at the current rate of evolution, dragonflies and damselflies have likely not had enough time to reach the very large sizes seen in the ancient past.

OU models are not able to assign causation to the constraints they measure, but they can tease out the rate of evolution (rate of change in the trait). In Fig. 6, I show that odonates in the modern age (the last 80-65 my since the last major extinction event), likely have not had enough time to reach the sizes observed in the fossil record 300 mya, but they nevertheless and with current selective constraints (caused by oxygen levels, birds, or some other ecological agent), they are likely to remain small until the constraints on their body size are relaxed (Fig. 6).

Evolution of thermal performance traits

Insects are considered thermal conformers because they allow their body temperature to closely match ambient temperature (Sanborn 2004; Angilletta 2009). In ectotherms, important physiological processes are at the mercy of ambient temperature, and from an enzymatic point of view, being an obligate thermal conformer might not always be optimal (Huey and Berrigan 2001; Blouin-Demers et al. 2005; Martin et al. 2008). Enzymes are less efficient at any non-optimal temperatures, so even though ectotherms will tend to match ambient temperatures without immediate mortality, we might still expect better thermoregulatory ability to be selected for in certain contexts, especially in extreme environments (Huey et al. 2009). While ectotherms have less control over their internal temperature than endotherms by definition, they typically still have the ability to thermoregulate to some degree via behaviour and physiology, and variation in this ability will differ between species and even within individuals in a population (Hoffmann et al. 2011; Somero 2010; Kearney et al. 2009; Buckley et al. 2015). Therefore, the dichotomy of being a thermoregulatory versus a thermal conformer is as false one, and even groups considered to be thermal conformers have fitness incentives to regulate their internal body temperatures to some extent (Huey and Slatkin 1976, Martin et al. 2008).

Climate change will likely increase seasonal variation in temperature in many areas, making the ability to thermoregulate increasingly important for ectotherms (Deutsch et al. 2008). Many temperate ectotherms appear to have maximum thermal tolerances that are 10–20 °C higher than required to withstand the average summer air temperatures where they live, which might reflect their evolutionary origin in tropical environments (Sanborn 2004; Sunday et al. 2011). There are many behavioral, morphological, and physiological adaptations that might allow ectotherms to survive unfavourable thermal environments (May 1979 ;Heinrich 2013), and the body temperatures of some insects have been recorded more than 35°C above ambient temperature (Sanborn 2004).

A large source of heat in insects is the flight muscles, and insects can increase heat generation through increased muscle activity (May 1979). Basking or changing body orientation to the sun is also a typical method that minimizes the exposed surface area and heat uptake (Porter 1982; Heinrich 2013). Color or melanism might also be a route to increasing the amount of radiation that gets turned into heat energy (True 2003; Trullas 2007; Svensson and Waller 2013). Finally, simply evolving a larger or smaller body size can change the surface-to-volume ratio of an insect, and effectively regulate its body temperature (Kingsolver and Huey 2008).

Thermal plasticity is a measure of an individual's ability to change its body temperature after a thermal stress (Trotta et al. 2006; Waller et al. 2017). It can be measured by the slope of heating or cooling rate after a thermal stress is applied.

This thermal performance measure captures any physiological differences between individuals in terms of their thermal plasticity, but also might capture any behavioral adjustments such as orientation and muscle activity.

The rate of evolution in response to selection on thermal plasticity depends on generation time, the additive genetic variances and covariances underlying phenotypic variation, and the strength of selection on traits that influence performance (Lynch and Walsh 1998), and a large number of studies have demonstrated that there is genetic variation in the slope of thermal reaction norms (Gutteling et al. 2007; Winterhalter and Mousseau 2007). Where environmental temperatures are largely constant, such as in the tropics, it might be expected that genetic variation in thermal traits might be low (Angilletta 2009). In other words, we might expect individuals in populations to be “canalized” by stabilizing into a certain thermal regimes and respond slowly to thermal stresses (Liefting et al. 2009). Conversely, in variable environments one might expect that selection be greater, and individuals might respond quickly to a thermal stress. Unfortunately, very few field studies have looked at selection on thermal traits in natural populations.

In **chapter VII**, I examine selection on thermal sensitivity in two species of damselflies common in southern Sweden, *Calopteryx splendens* (Harris 1780) and *Calopteryx virgo* (Linnaeus 1758). The genus *Calopteryx* did not evolve in Europe, but has migrated from tropical latitudes (Svensson and Waller 2013), and invaded southern Sweden around 12 000 years ago after the last glacial period (Misof et al. 2000; Clark et al. 2009).

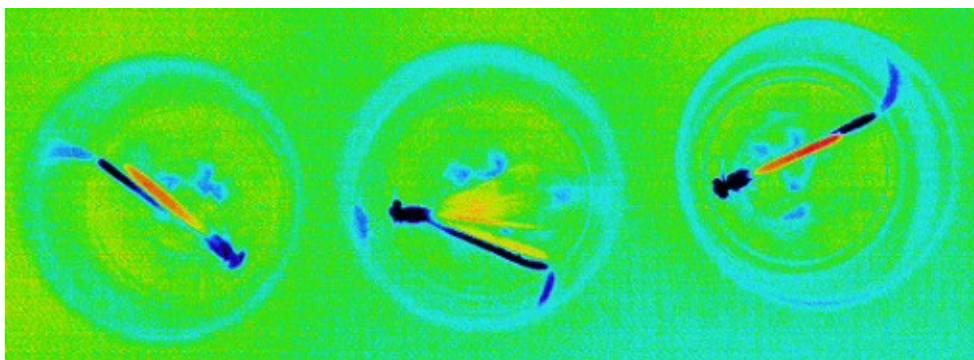


Figure 7. Thermal images.
Thermal images of damselflies taken in the field.

Materials and Methods

Measuring selection in natural populations

Simple predictions about expected evolutionary changes in population mean phenotypes can be made using the following equation, as developed from quantitative genetics (Lynch & Walsh 1998)

$$\Delta\hat{z} = G\beta$$

Where, $\Delta\hat{z}$ denotes a vector of changes in phenotypic means of the population, β denotes a directional selection gradient vector, G denotes the genetic variance-covariance matrix (Lande 1979). Depending on the structure of the G matrix, and making some assumptions about the multi-variate normality of the measured traits, we can use this information to predict the direction and magnitude of evolutionary change in natural populations (Lande and Arnold 1983).

It is important to emphasize that measuring selection (β) is different from measuring the reaction to selection ($\Delta\hat{z}$). If one knows $\Delta\hat{z}$ and G for a set of phenotypes, it is possible to solve for β , but such information is rarely available in natural populations . Lande and Arnold (1983), pointed out that fitness could be split into different components, such as survival, viability selection (survival to sexual maturity), mating success (the number of mates), or fecundity (the number of eggs). These components of fitness can then be used to quantify selection on the given set of phenotypes. The product of these different components of fitness can be combined to produce selection on total fitness. Such a fitness function could be written as a simple linear model:

$$w = \alpha + \beta z + \frac{1}{2}\gamma z^2$$

Where w is relative fitness, α is an intercept term, β is a linear selection gradient, z is a standardized trait vector and γ provides information about any curvature in the fitness surface (e. g. disruptive or stabilizing selection). The practical implementation of Lande and Arnold's method involves simply fitting a linear model with standardized response (survival ,mating success, fecundity) and explanatory (trait) variables values with quadratic terms (multiplied by two) (Stinchcombe *et al.* 2008). These straightforward statistical methods have stimulated evolutionary biologists to measure literally thousands selection coefficients using commonly available statistical software (Kingsolver *et al.* 2001).

It is important to remember that genetic covariances (G) between traits can change what the response to selection will ultimately be in the next generation. Additionally, un-measured components of total fitness or un-measured traits can

change the evolutionary response to selection (Grafen 1989; Hadfield 2008). For example, some individuals might die before a trait is measured or expressed (the so-called “invisible fraction”) (Nakagawa and Freckleton 2008). These types of issues might be especially common in animals with complex lifecycles, such as odonates, where selection in the adult life stage is easier to measure than selection in the aquatic larval stage.

Field work

Field work in this thesis was all done at the same location and on the same two species (*Calopteryx splendens* and *Calopteryx virgo*) (Fig. 8). For 5 years (2013–2017), field studies were performed in the river system Klingavälsån in the province of Skåne in southern Sweden at Sövdemölla Gård (Lat: 55.601492, Lon: 13.657340). Studies were conducted each summer (June and July). For each year, I collected morphological, life history, and thermal performance data on two damselfly species. My field work typically took place between 10.00 and 15.00 when activity levels were the highest. Damselflies at our field site have very low activity levels when temperature are below 15°C and during rainy and windy days (i.e. no flying or mating behaviour).

Males and females of both study species were captured with handnets and processed in the field. Individual damselflies were measured by placing their wings flat against two layers of transparency paper. A digital scan was then taken with a known scale and the body and wings traits were measured in ImageJ (Schneider et al. 2012; Waller and Svensson 2016; 2017). We measured the following traits: 1) front wing lengths and widths; 2) total body length; 3) abdomen length; 4) thorax length. The mass of each caught individual was taken to the nearest ± 0.002 grams. For *C. splendens* males patch length, width, and darkness were also measured from the scanned images. Patch darkness was taken as the mean intensity of a grayscale image with larger values meaning darker wing patches. Each individual was also marked with a unique 3-digit color code so that their survival could be tracked in the field (Fig. 9).

Throughout the 5 years of study, over 5000 individuals were captured and had their morphology measured. This large amount of data necessitated the implementation of automated procedures to be process this large number of damselflies. I developed a trained machine learning algorithm that was able to recognize the size and position of the wings of the scanned individual damselflies. (Fig. 10)



Figure 8. Field Assistants.
Many people helped collecting data from 2013-2017. Some are pictured here.



Figure 9. damselfly being painted.

A individual being painted with a color code so that it could be tracked in the field throughout its lifespan.

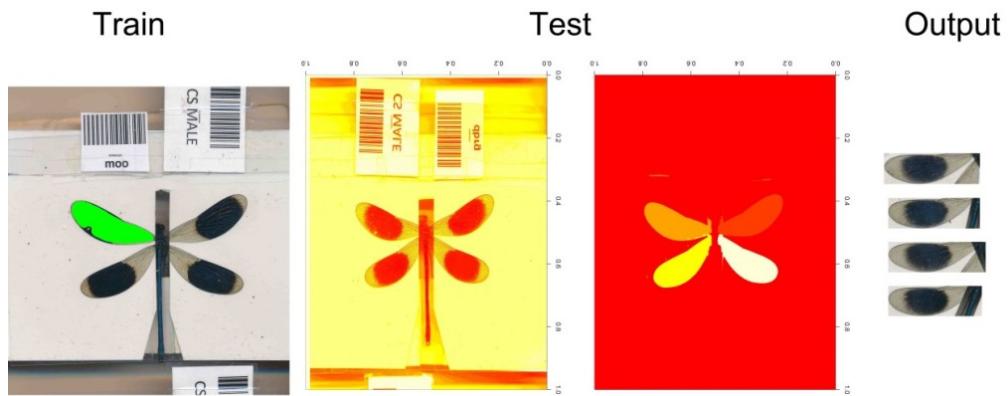


Figure 10. Image processing.

I trained a random forest algorithm to separate background pixels from wing pixels and also to determine the location of the wing (i.e. from, back or right, left). Output wings pictures were later reviewed for accuracy. Wing patches of *C. splendens* males were also able to be measured using a similar procedure.

Mark-recapture methods

After marking damselflies in the field, I attempted to track how long they were living through repeated attempts to recapture the marked individuals. Not every time I went looking for a marked individual, however, did I find it. This is known as imperfect detection or low recapture probability in the field of mark-recapture methods (Fig. 11). Gimenez et al. (2008) showed, using a field dataset, that even the mode of selection (directional or stabilizing) might be wrongly inferred when variation in recapture probability is ignored.

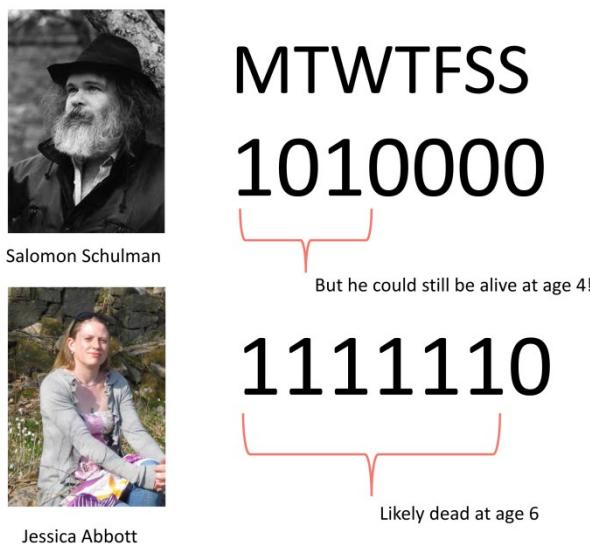


Figure 11. Mark-recapture methods.

Jessica Abbott (co-adviser of the author) and Salomon Schulman (famous doctor in Lund) are two individuals in my life with different resight probabilities (1 = seen, 0 = not seen, on the day of the week). In this scenario, Jessica has a higher likelihood of being dead on Sunday, since she has a higher recapture probability than Salomon who might still be alive on Thursday.

Although imperfect detection has therefore been recognized as a problem, the severity of this problem has never been quantified. In **chapter I**, I use simulations to determine the degree of the problem and develop a new R package to process capture history data.

One statistical solution to imperfect detection is capture-mark-recapture methods (Jolly 1982; Seber 1965; Cormack 1964; Lebreton et al. 1992; Seber & Schwarz 2002). To correct for imperfect detection, mark recapture methods use the extra information within individual capture histories to estimate recapture probability. For example, if an individual is seen on day 1, and then not seen on

day 2, but then is seen later on day 3, (a capture history of “101”) a researcher knows that individual was definitely alive on day 2, but was simply not detected

Building the odonate phenotypic database

The screenshot shows the homepage of the Odonate Phenotypic DataBase. On the left, there is a sidebar with a "Look Up" section containing dropdown menus for "Select family" (Calopterygidae), "Select genus" (Calopteryx), and "Select species" (Calopteryx splendens). Below this are "Download" buttons for "Complete Database" and "Variable Definitions". The main content area displays information for *Calopteryx splendens*. It includes sections for "Size ♂" (Body Length: 46.5 mm, Front Wing Length: -, Hind Wing Length: 31.5 mm), "Size ♀" (Body Length: -, Front Wing Length: -, Hind Wing Length: -), "Body Colors ♂" (Body Colors: green, Body Color Type(s): structural, Body Pattern Type(s): plain), and "Morphisms" (Strength of Sexual Dimorphism: strong, Polymorphisms by Sex: females, Polymorphisms by Region: -). The background features several small, semi-transparent dragonfly and damselfly illustrations.

Figure 12. Home page.
Home page of www.odonatephenotypicdatabase.org.

The Odonate Phenotypic Database (Fig. 12) is a resource for dragonfly and damselfly phenotypes (Odonates). In **chapter II**, I discuss construction of the database, which consists of a variety of morphological, behavioral, and biogeographical information collected from various sources in the literature. The database is not intended for species identification, but for comparative analysis within the group or to be combined with data from other groups. The database is provided along with a large tree (1322 taxa, 21% of known odonates) “The Odonate Super Tree”, which is discussed below.

I discuss how taxon-specific phenotypic databases as becoming an increasing valuable resource in comparative studies. Our database has at least some

phenotypic records for 1011 of around 6400 known species. The database is accessible at <http://www.odonatephenotypicdatabase.org/>.

Odonate super tree

For the comparative analyses done in this thesis (namely in **chapters III and V**, I needed a large phylogenetic tree of dragonflies and damselflies (Fig. 13), which was also time calibrated with fossils (Fig. 14).

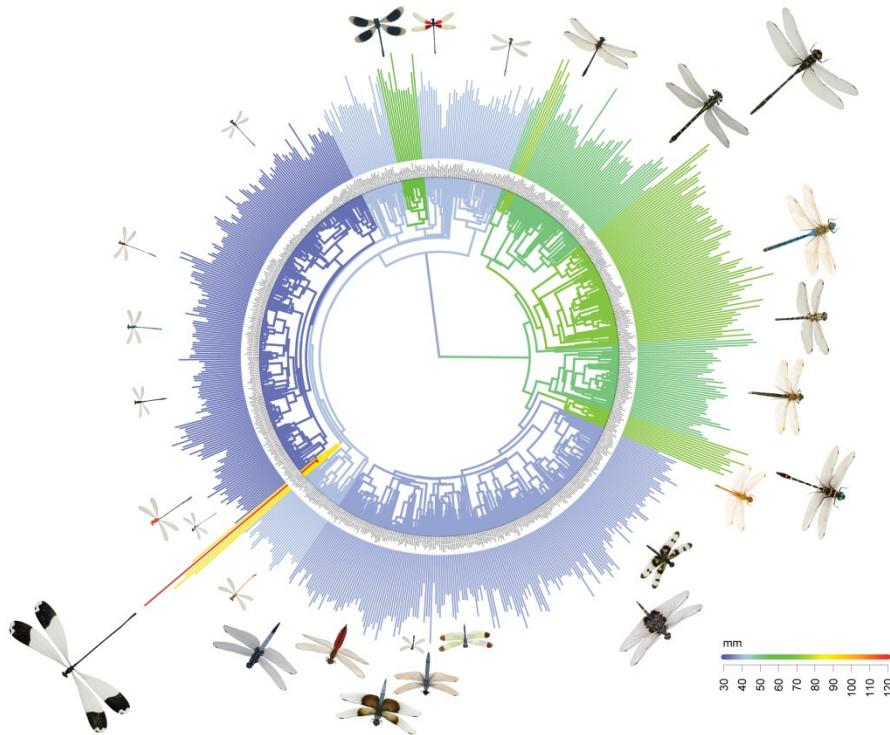


Figure 13. Odonate Super Tree.

The odonate super tree plotted with representative species. The body length of each species is also plotted. See chapter III, for more details.

The tree is comprised 1322 taxa, which is around 21% of known odonates ($n = 6400$). A total of 759 of these taxa were from the suborder Zygoptera (damselflies) and a total of 563) from Anisoptera (dragonflies). All 32 known odonate families are represented in this phylogeny.

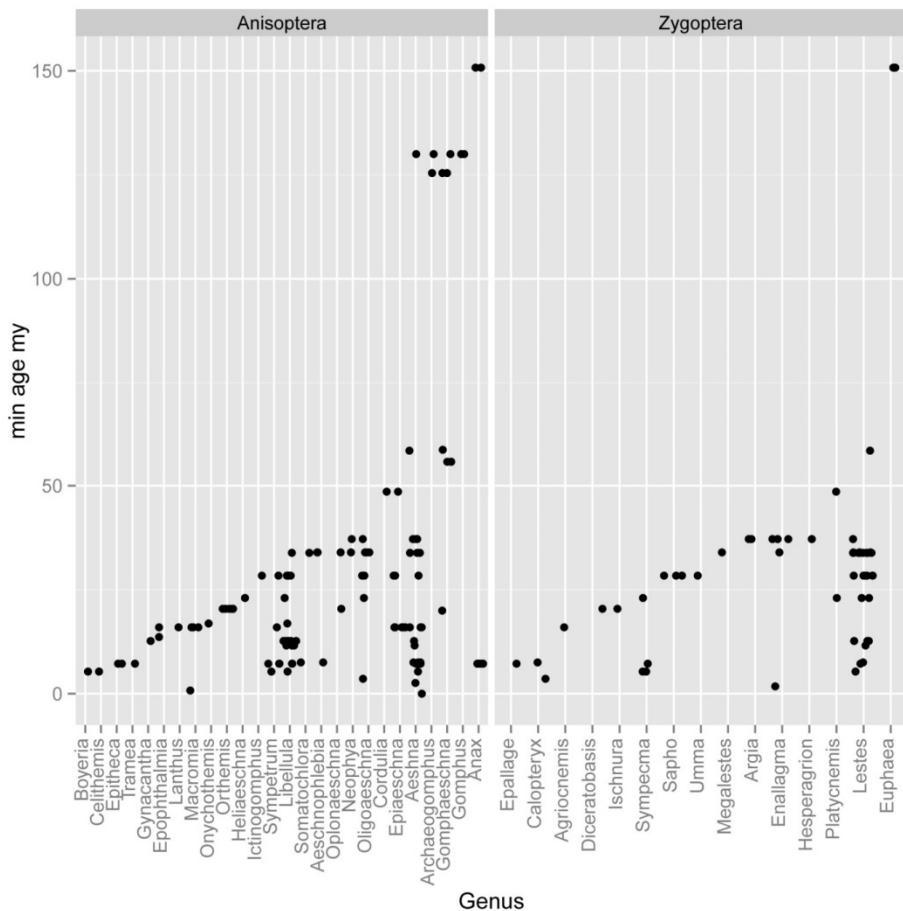


Figure 14. Fossils used for calibration.
Fossils used to time calibrate phylogenetic the odonate super tree.

Major Results

In **chapter I**, developed new methods in order to obtain selection gradients from mark-recapture models, so that comparable with other selection estimates, while still accounting for variation in recapture probability. Low recapture probability was a problem in my field work, since insects are difficult to track. I find that the main advantage of mark–recapture methods over other methods is the ability to control for hidden trait-dependent recapture probability, as it is often hard to tell *a priori* if trait dependence is an issue in a particular study. However, when trait-

dependent recapture probability is weak, traditional methods and mark–recapture methods perform similarly as long as recapture rates do not become too low, and the main problem of survival selection studies is still low statistical power. I developed the R package **EasyMARK** to facilitate future integration between MR methods and classical selection studies.

In **chapter II**, I discuss how we collected the data we later use in our comparative tests on dragonflies and damselflies. The Odonate Phenotypic Database is an online data resource for dragonfly and damselfly phenotypes (Insecta: Odonata). The database consists of a variety of morphological life-history behavioral traits, and biogeographical information collected from various sources in the literature. I see taxon-specific phenotypic databases becoming an increasing valuable resource in comparative studies. We have collected 34 phenotypic variables, and our database has at least some phenotypic records for 1011 of all 6400 known odonate species. The database is accessible at odonatephenotypicdatabase.org.

In **chapter III**, I integrate field data and phylogenetic comparative analyses to investigate the causes of body size evolution and stasis in dragonflies and damselflies. I address the paradox of if bigger is better, why is there widespread stasis in the fossil record? I find that there is strong evidence for “bigger is better” in the adult stage in dragonflies and damselflies, and I suggest that the lack of Cope’s rule in this group results from conflicting selection between fitness advantages of large adult size and costs of long larval development.

In **chapter IV**, I make a commentary on Uyeda et al. (2011), and the misuse of the visual metaphor of the blunderbuss as a symbol of stasis and punctuated evolution. that was originally put forward in Uyeda et al. (2011) “The million-year wait for macroevolutionary bursts”.

In chapter V, I look into the factors influencing geographic body size patterns in dragonflies and damselflies. I find that the temperature-size rule (“hotter is smaller”), might not explain global geographic size differences between species. In general, I find that temperature may not be the primary force driving latitudinal body size trends in this group, and that ecological factors such as birds and dispersal ability may play a larger role in shaping geographic body size patterns.

In **chapter VI**, I investigate the evolutionary potential of phenotypic plasticity in the thermal performance in fruit flies. I find that genetic variance in thermal performance curves might be low for male mating rate, making it unlikely that sexual selection through male mating success would be efficient in mediating evolutionary rescue through changed plasticity in response to changing temperatures.

Finally in **chapter VII**, I examine the “hotter is better” rule in damselflies. In the study, I find that temperature extremes have a negative effect on survival and mating rate in our two studies species, and that sexual selection on thermal plasticity increases in both high and low temperatures, implying that increased

heating rate (“hotter is better”) is only adaptive in non-optimal temperature environments.

Conclusions

In this thesis I set out to explore the 3 rules of insect evolution proposed by Kingsolver and Huey (2008):

bigger is better

hotter makes you smaller

hotter is better

Do the 3 rules hold up? What are the three rules after this thesis?

bigger is better, but time in the larval stage is limited.

hotter makes you smaller, but so maybe also do birds.

hotter is not always better, but thermoregulatory ability is favored in extreme conditions.

Acknowledgments

I moved to Lund from the US in August of 2009. At that time I did not really imagine that I would still be living in Sweden almost 10 years, and defending my thesis here. During this time many people have come and gone from Ekologihuset, but the great working atmosphere has stayed the same, and I would like to thank whoever is responsible for that, which is perhaps **Christer Löfstedt** or many of the seen (**Anne Fogelberg**) and unseen administrators, but probably everyone working here deserves a thank you.

My adviser **Erik Svensson** has been a huge part why my time in Lund has been so great. Truly a great adviser and recommended to anyone seeking a PhD. In Erik's lab I have gotten to travel to South Africa (2 times) and Argentina, and see more birds than I wanted to see. Throughout this PhD I relied heavily on Erik's experience, intuition, and surprising insights. My co-adviser **Jessica Abbott** has also been a great inspiration, and I have enjoyed working with her in the classroom, in the lab, and watching her on TV. I would also like to thank my official mentor **Jan-Åke Nilsson** and examiner **Anders Hedenström** who have always been very helpful during many doctoralsamtals.

I would like to thank all of the co-authors on the chapters of my thesis **Anna Kell**, **Mireia Ballestà**, **Aude Giraud**, **Max Tschol**, **Beatriz Willink**, **Jessica Abbot**, and **Erik Svensson**. And I would like to thank other co-authors as well **Anna Runemark**, **Tom Gosden**, and **Anna Norden**.

There were many people that I worked with during the summers of 2010 – 2017. **Sopia Engel** thank you for introducing me to damselflies and field work in Sweden in 2010. I would also like to thank the following people who helped with field work from 2013 to 2017: **Anna Kell**, **Mireia Ballestà**, **Miguel Gomez**, **Kaili Boarman**, **Shawn Kutch**, **Stephany Herman**, **Kajsa Svensson**, **Tammy Ho**, **Rasmus Johansson**, **Hanna Bensch**, **Chiara Despasqual**, **Johanna Eklund**, **Emil Fjellstedt**, **Qinyang Li**, **Eric Leach**, **Lisa Winberg**, **Rebecka Thisé**, and **Karolina Pehrson**.

And those individuals who helped contribute to the Odonate Phenotypic Database: **Kajsa Svensson**, **Robin Pranter**, **Lisa Winberg**, **Ev Poslin**, **Hanna Bensch**, and **Karolina Pehrson**.

There many current and past members of EXEB, which deserve acknowledgment either for listening to me speak, commenting on a manuscript, or for presenting great papers during lab meetings: **Sophia Engel**, **Machteld Verzijden**, **Maren Wellenreuther**, **Tom Gosden**, **Lesley Lancaster**, **Yuma Takahashi**, **Katrine Lund-Hansen**, **Alex Hegg**, **Viktor Nilsson**, **Anna Runemark**, **Jessica Abbott**, **Anna Norden**, **Katie Duryea**, **Stephen De Lisle**, **Maarit Mäenpää**, **Hanna Laakkonen**, **Tobias Uller**, **Nathalie Feiner**, **Alfredo Rago**, and **Reinder Radersma**.

I would also like to thank other PhD students who have been good friends and colleagues: **Katrine Lund-Hansen**, **Beatriz Willink**, **Lokeshwaran Manoharan**, **Fredrik Andreasson**, **Tom Evans**, **Linus Hedh**, **Alex Hegg**, **Marco Klein**, **Yesbol Manat**, **Cecilia Nilsson**, **Anna Nordén**, **Gabriel Norevik**, **Pablo Salmon**, **Utku Urhan**, **Kajsa Warfvinge**, **Saeed Alaei**, **Dafne Ram**, **Anna Drews**, **Julian Melgar**, **Kristin Rath**, **Jacob Roved**, **Maja Tarka**, **Therese Reber**, **Jonas Håkansson**, **Hanna Sigeman**, **Elin Videvall**, **Lovisa Nilsson**, **Martin Stervander**, **John Kirwan**, and **Shahriar Shakil**.

I would also like to thank members of Evolutionary Ecology, MEMEG, and other groups in the biology department: **Thomas Alerstam**, **Arne Andersson**, **Anders Brodin**, **Jon Loman**, **Per Lundberg**, **Jörgen Ripa**, **Susanne Åkesson**, **Staffan Bensch**, **Charlie Cornwellis**, **Philip Downing**, **Bengt Hansson**, **Dennis Hasselquist**, **Jane Jönsson**, **Julio Neto**, **Emily O'Connor**, **Suvi Ponnikas**, **Kristaps Sokolovskis**, **Maria Strandh**, and **Håkan Wallander**.

I would also like to thank **Petro Buyak** for painting the cover art.

I would finally like to thank **Liudmyla Lozinska** who has supported me and given great advice during the writing of this thesis.

References

- Angilletta, M. J. (2009). Thermal adaptation: a theoretical and empirical synthesis. Oxford University Press.
- Arnold, S. J., Pfrender, M. E., & A. G. Jones. (2001). The adaptive landscape as a conceptual bridge between micro-and macroevolution. *Genetica*. 112: 9-32.
- Baird, I. R. (2014). Larval burrow morphology and groundwater dependence in a mire-dwelling dragonfly, *Petalura gigantea* (Odonata: Petaluridae). *International journal of odonatology*. 17:101-121.
- Berner, R. A. (2009). Phanerozoic atmospheric oxygen: New results using the GEOCARBSULF model. *American Journal of Science*. 309:603-606.
- Blouin-Demers, G., and P. Nadeau. (2005). The cost–benefit model of thermoregulation does not predict lizard thermoregulatory behavior. *Ecology*. 86:560-566.
- Braune, P., and J. Rolff. (2001). Parasitism and survival in a damselfly: does host sex matter?. *Proceedings of the Royal Society of London B: Biological Sciences*. 268:1133-1137.
- Buckley, L. B., Ehrenberger, J. C., and M. J. Angilletta. (2015). Thermoregulatory behaviour limits local adaptation of thermal niches and confers sensitivity to climate change. *Functional Ecology*. 29:1038-1047.
- Butler, M. A., and A. A. King. (2004). Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *The American Naturalist*. 164:683-695.
- Clapham, M. E., and J. A. Karr. (2012). Environmental and biotic controls on the evolutionary history of insect body size. *Proceedings of the National Academy of Sciences*. 109:10927-10930.
- Clark, P. U., Dyke, A. S., Shakun, J. D., Carlson, A. E., Clark, J., Wohlfarth, B., Mitrovica, Steven W. Hostetler, and A. M. McCabe. (2009). The last glacial maximum. *Science*. 325:710-714.
- Corbet, P. S. (1999). Dragonflies: behaviour and ecology of Odonata. Harley books.
- Cormack, R. M. (1964). Estimates of survival from the sighting of marked animals. *Biometrika*. 51:429-438.
- Deutsch, C. A., Tewksbury, J. J., Huey, R. B., Sheldon, K. S., Ghalambor, C. K., Haak, D. C., and P. R. Martin. (2008). Impacts of climate warming on terrestrial ectotherms across latitude. *Proceedings of the National Academy of Sciences*. 105:6668-6672.
- Donoghue, M. J., and D. D. Ackerly. (1996). Phylogenetic uncertainties and sensitivity analyses in comparative biology. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*. 351:1241-1249.
- Drummond, A. J., and A. Rambaut. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology*. 7:214.
- Evans A. R., Jones D., Boyer A.G., Brown J. H., Costa D. P., Ernest S. M., Fitzgerald E. M., Fortelius M., Gittleman J. L., Hamilton M. J., and L.E. Harding (2012). The maximum rate of mammal evolution. *Proceedings of the National Academy of Sciences*. 109:4187-4190.

- Felsenstein, J. (1984). Distance methods for inferring phylogenies: a justification. *Evolution*. 38:16-24.
- Felsenstein, J. (1997). An alternating least squares approach to inferring phylogenies from pairwise distances. *Systematic biology*. 46:101-111.
- Garamszegi, L. Z. (Ed.). (2014). Modern phylogenetic comparative methods and their application in evolutionary biology: concepts and practice. Springer.
- Gimenez O., Viallefond A., Charmantier A., Pradel R., Cam E., Brown C. R., Anderson M.D., Brown M.B., Covas R., and J. M. Gaillard. (2008). The risk of flawed inference in evolutionary studies when detectability is less than one. *The American Naturalist*. 172:441-448.
- Grafen, A. (1989). The phylogenetic regression. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*. 326:119-157.
- Grether, G. F. (1996). Intrasexual competition alone favors a sexually dimorphic ornament in the rubyspot damselfly *Hetaerina americana*. *Evolution*. 50:1949-1957.
- Gutteling, E. W., Riksen, J. A. G., Bakker, J., and J. E. Kammenga. (2007). Mapping phenotypic plasticity and genotype-environment interactions affecting life-history traits in *Caenorhabditis elegans*. *Heredity*. 98:28-37.
- Hansen, T. F. (1997). Stabilizing selection and the comparative analysis of adaptation. *Evolution*. 51:1341-1351.
- Hansen, T. F., and E. P. Martins. (1996). Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. *Evolution*. 50:1404-1417.
- Harrison, J. F., Kaiser, A., and J. M. VandenBrooks. (2010). Atmospheric oxygen level and the evolution of insect body size. *Proceedings of the Royal Society of London B: Biological Sciences*. 277:1937-1946.
- Heinrich, B. (1996). The thermal warriors: strategies of insect survival. *Harvard University Press*.
- Hoffmann, A. A., Chown, S. L., and S. Clusella-Trullas. (2013). Upper thermal limits in terrestrial ectotherms: how constrained are they?. *Functional Ecology*, 27:934-949.
- Huelsenbeck, J. P., and F. Ronquist. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*. 17:754-755.
- Huey, R. B., and D. Berrigan. (2001). Temperature, demography, and ectotherm fitness. *The American Naturalist*. 158:204-210.
- Huey, R. B., and M. Slatkin. (1976). Cost and benefits of lizard thermoregulation. *The Quarterly Review of Biology*. 51:363-384.
- Huey, R. B., Deutsch, C. A., Tewksbury, J. J., Vitt, L. J., Hertz, P. E., Pérez, H. J. Á., and T. Garland. (2009). Why tropical forest lizards are vulnerable to climate warming. *Proceedings of the Royal Society of London B: Biological Sciences*, 276:1939-1948.
- Johansson, F., and L. Rowe. (1999). Life history and behavioral responses to time constraints in a damselfly. *Ecology*. 80:1242-1252.
- Jolly, G. M. (1982). Mark-recapture models with parameters constant in time. *Biometrics*. 30:301-321.

- Kearney, M., Shine, R., and W. P. Porter. (2009). The potential for behavioral thermoregulation to buffer cold-blooded animals against climate warming. *Proceedings of the National Academy of Sciences*. 106:3835-3840.
- Kingsolver, J. G., and R. B. Huey. (2008). Size, temperature, and fitness: three rules. *Evolutionary Ecology Research*. 10:251-268.
- Kingsolver J.G., Hoekstra H. E., Hoekstra J.M., Berrigan D., Vignieri S. N., Hill C. E., Hoang A., Gibert P., P. Beerli. (2001). The strength of phenotypic selection in natural populations. *The American Naturalist*. 157:245-261.
- Klok, C. J., Hubb, A. J., and J. F. Harrison. (2009). Single and multigenerational responses of body mass to atmospheric oxygen concentrations in *Drosophila melanogaster*: evidence for roles of plasticity and evolution. *Journal of evolutionary biology*. 22:2496-2504.
- Kuchta, S. R., and E. I. Svensson. (2014). Predator-mediated natural selection on the wings of the damselfly *Calopteryx splendens*: differences in selection among trait types. *The American Naturalist*. 184:91-109.
- Lande, R. (1979). Quantitative genetic analysis of multivariate evolution, applied to brain: body size allometry. *Evolution*. 33:402-416.
- Lande, R., and S. J. Arnold. (1983). The measurement of selection on correlated characters. *Evolution*. 37:1210-1226.
- Lebreton, J. D., Burnham, K. P., Clobert, J., and D. R. Anderson. (1992). Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecological monographs*. 62:67-118.
- Liefting, M., Hoffmann, A. A., and J. Ellers. (2009). Plasticity versus environmental canalization: population differences in thermal responses along a latitudinal gradient in *Drosophila serrata*. *Evolution*. 63:1954-1963.
- Lynch, M., and B. Walsh. (1998). Genetics and analysis of quantitative traits (Vol. 1, pp. 535-557). Sunderland, MA: Sinauer.
- Martin, T. L., and R. B. Huey. (2008). Why suboptimal is optimal: Jensen's inequality and ectotherm thermal preferences. *The American Naturalist*. 171:102-118.
- Martins, E. P., and E. A. Housworth (2002). Phylogeny shape and the phylogenetic comparative method. *Systematic Biology*. 51:873-880.
- May, M. L. (1979). Insect thermoregulation. *Annual review of entomology*. 24:313-349.
- McPeek, M. A. (2004). The growth/predation risk trade-off: so what is the mechanism?. *The American Naturalist*. 163:88-111.
- Misof, B., Anderson, C. L., and H. Hadrys. (2000). A phylogeny of the damselfly genus *Calopteryx* (Odonata) using mitochondrial 16S rDNA markers. *Molecular Phylogenetics and Evolution*. 15:5-14.
- Nakagawa, S., and R. P. Freckleton. (2008). Missing inaction: the dangers of ignoring missing data. *Trends in Ecology & Evolution*. 23:592-596.
- Pagel, M. (1999). Inferring the historical patterns of biological evolution. *Nature*. 401:877-884.
- Pagel, M. D., and P. H. Harvey. (1988). Recent developments in the analysis of comparative data. *The Quarterly Review of Biology*. 63:413-440.

- Pagel, M., Meade, A., and D. Barker. (2004). Bayesian estimation of ancestral character states on phylogenies. *Systematic biology*. 53:673-684.
- Porter, K. (1982). Basking behaviour in larvae of the butterfly *Euphydryas aurinia*. *Oikos*. 47:308-312.
- Rantala, M. J., Koskimäki, J., Taskinen, J., Tynkkynen, K., and J. Suhonen. (2000). Immunocompetence, developmental stability and wingspot size in the damselfly *Calopteryx splendens*. *Proceedings of the Royal Society of London B: Biological Sciences*. 267:2453-2457.
- Saitou, N., and M. Nei. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution*. 4:406-425.
- Sanborn, A. F. (2004). Thermoregulation and endothermy in the large western cicada *Tibicen cultriformis* (Hemiptera: Cicadidae). *Journal of Thermal Biology*. 29:97-101.
- Schneider, C. A., Rasband, W. S., and K. W. Eliceiri. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature methods*. 9:671-675.
- Seber, G. A. F., and E. D. Le Cren. (1967). Estimating population parameters from catches large relative to the population. *The Journal of Animal Ecology*. 44:631-643.
- Seber, G. A., and C. J. Schwarz. (2002). Capture-recapture: before and after EURING 2000. *Journal of Applied Statistics*. 29:5-18.
- Siva-Jothy, M. T. (2000). A mechanistic link between parasite resistance and expression of a sexually selected trait in a damselfly. *Proceedings of the Royal Society of London B: Biological Sciences*. 267:2523-2527.
- Siva-Jothy, M. T., Gibbons, D. W., and D. Pain. (1995). Female oviposition-site preference and egg hatching success in the damselfly *Calopteryx splendens xanthostoma*. *Behavioral Ecology and Sociobiology*. 37:39-44.
- Somero, G. N. (2010). The physiology of climate change: how potentials for acclimatization and genetic adaptation will determine winners and losers. *Journal of Experimental Biology*. 213:912-920.
- Soul, L. C., and M. Friedman. (2015). Taxonomy and phylogeny can yield comparable results in comparative paleontological analyses. *Systematic biology*. 64:608-620.
- Stinchcombe J. R., Agrawal A. F., Hohenlohe P. A., Arnold S.J., and M.W. Blows. (2008). Estimating nonlinear selection gradients using quadratic regression coefficients: double or nothing?. *Evolution*. 62.9:2435-2440.
- Stoks, R., and A. Córdoba-Aguilar. (2012). Evolutionary ecology of Odonata: a complex life cycle perspective. *Annual review of entomology*. 57:249-265.
- Sunday, J. M., Bates, A. E., and N. K. Dulvy. (2011). Global analysis of thermal tolerance and latitude in ectotherms. *Proceedings of the Royal Society of London B: Biological Sciences*. 278:1823-1830.
- Svensson, E. I., and M. Friberg. (2007). Selective predation on wing morphology in sympatric damselflies. *The American Naturalist*. 170:101-112.
- Svensson, E. I., and J. T. Waller. (2013). Ecology and sexual selection: evolution of wing pigmentation in calopterygid damselflies in relation to latitude, sexual dimorphism, and speciation. *The American Naturalist*. 182:174-195.

- Trotta, V., Calboli, F. C., Ziosi, M., Guerra, D., Pezzoli, M. C., David, J. R., and S. Cavicchi. (2006). Thermal plasticity in *Drosophila melanogaster*: a comparison of geographic populations. *BMC evolutionary biology*. 6:67-87.
- True, J. R. (2003). Insect melanism: the molecules matter. *Trends in ecology & evolution*. 18:640-647.
- Trullas, S. C., van Wyk, J. H., and J. R. Spotila. (2007). Thermal melanism in ectotherms. *Journal of Thermal Biology*. 32:235-245.
- Tsubaki, Y., and R. E. Hooper. (2004). Effects of eugregarine parasites on adult longevity in the polymorphic damselfly *Mnais costalis* Selys. *Ecological entomology*. 29:361-366.
- Tynkkynen, K., Kotiaho, J. S., Luojumäki, M., and J. Suhonen. (2005). Interspecific aggression causes negative selection on sexual characters. *Evolution*. 59:1838-1843.
- Van de Meutter, F., De Meester, L., and R. Stoks. (2005). Water turbidity affects predator-prey interactions in a fish-damselfly system. *Oecologia*. 144:327-336.
- Waller, J. T., Kell, A., Ballesta, M., Giraud, A., Abbott, J., and E. I. Svensson. (2017). Limited genetic variation for male mating success reveals low evolutionary potential for thermal plasticity in *Drosophila melanogaster*. *bioRxiv*, 166801.
- Winterhalter, W. E., and T. A. Mousseau. (2008). The strength of temperature-mediated selection on body size in a wild insect population. *Journal of Orthoptera Research*. 17:347-351.
- Yang, Z. (1994). Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *Journal of Molecular evolution*. 39:306-314.

Paper I



The measurement of selection when detection is imperfect: How good are naïve methods?

John Waller* and Erik I. Svensson

Evolutionary Ecology Unit, Department of Biology, Lund University, SE-223 62 Lund, Sweden

Summary

1. The life spans of animals can be measured in natural populations by uniquely marking individuals and then releasing them into the field. Selection on survival (a component of fitness) can subsequently be quantified by regressing the life spans of these marked individuals on their trait values. However, marked individuals are not always seen on every subsequent catching occasion, and for this reason, imperfect detection is considered a problem when estimating survival selection in natural populations.
2. Capture–mark–recapture methods have been advocated as a powerful means to correct for imperfect detection. Here, we use simulated and field data sets to evaluate the effect of assuming perfect detection ('naïve methods'), when detection is really imperfect. We compared the performance of the naïve methods with methods correcting for imperfect detection (mark–recapture methods, or MR).
3. Although the effects of trait-dependent recapture probability are mitigated when recapture probability is high, mark–recapture methods still provide the safest choice when recapture probability might be trait-dependent. In our simulations, mark–recapture methods had a power advantage over naïve methods, but all methods lost statistical power at low recapture probabilities.
4. The main advantage of mark–recapture methods over naïve methods is the ability to control for hidden trait-dependent recapture probability, as it is often hard to tell *a priori* if trait dependence is an issue in a particular study. However, when trait-dependent recapture probability is weak, naïve methods and mark–recapture methods perform similarly as long as recapture rates do not become too low, and the main problem of survival selection studies is still low statistical power. We provide a R package (EasyMARK) alongside with this paper to facilitate future integration between MR methods and classical selection studies. EasyMARK provides the opportunity to convert the regression coefficients from MR-approaches in to classical standardized selection gradients.

Key-words: capture–mark–recapture, directional selection, mark, natural selection, selection gradients, simulation, stabilizing selection

Introduction

The detection and quantification of natural and sexual selection in the field has a long history in evolutionary biology. A considerable milestone in this field was the highly influential publication by Russel Lande and Steve Arnold in the early nineteen eighties (Lande & Arnold 1983), and interest was further stimulated by John Endler's classical volume about natural selection in the wild a few years later in 1986 (Endler 1986). Since these early influential contributions, thousands of selection estimates have been published from plant and animal populations, and our knowledge about the strength, mode and occurrence of sexual and natural selection has increased considerably (Kingsolver *et al.* 2001, 2012). Several more recent meta-analyses and statistical developments have since been published in this very active research field. Recent studies deal with important methodological challenges such as critical

sample sizes and their possible effects on selection gradients (Kingsolver *et al.* 2001; Knapczyk & Conner 2007), whether observed magnitudes of selection can be considered weak or strong (Conner 2001; Kingsolver *et al.* 2001; Hereford, Hansen & Houle 2004), the role of temporal (Siepielski, DiBattista & Carlson 2009; Siepielski *et al.* 2011) and spatial variation in selection (Gosden & Svensson 2008; Calsbeek *et al.* 2012; Siepielski *et al.* 2013) and how measurement error of both phenotypic traits and fitness measures can effect selection estimates (Kingsolver & Diamond 2011; Morrissey & Hadfield 2012). Other important issues are the inferences that could be made from a purely correlative approach to studying selection (Mitchell-Olds & Shaw 1987) and the ecological causes of selection and the need to identify selective agents (Wade & Kalisz 1990; Svensson & Sinervo 2000; MacColl 2011).

Yet one issue in survival selection studies that has largely been ignored is imperfect detection. Clobert (1995) pointed out that evolutionary ecologists tended to ignore imperfect detection and assume any effect of unobserved individuals was negligible, using the last date seen in the field instead. Later,

*Correspondence author. E-mail: john.waller@biol.lu.se

Gimenez *et al.* (2008) showed, using a field data set, that even the mode of selection (directional or stabilizing) might be wrongly inferred when the recapture probability is less than one. Although imperfect detection has therefore been recognized as a problem, the severity of this problem has never been quantified using systematic simulations or in field studies, and it is very seldom discussed. This raises some important questions: Should past studies be reanalysed? How low does recapture probability have to be before it becomes a problem? In which direction would imperfect detection effect the estimates of selection – if any? How good are currently available statistical techniques at correcting such errors?

One statistical solution to imperfect detection is capture-mark-recapture methods (MR). To correct for imperfect detection, MR uses the extra information within individual capture histories to estimate recapture probability. For example, if an individual is seen on day 1, and then not seen on day 2, but then is seen later on day 3, (a capture history of '101') a researcher knows that individual was definitely alive on day 2, but was simply not detected. Describing MR methods in detail and their many implementations is outside the scope of this paper. Instead, we point readers to the key literature in the field (Cormack 1964; Seber 1965; Jolly 1982; Lebreton *et al.* 1992; Seber & Schwarz 2002; Williams, Nichols & Conroy 2002).

Here, our primary aim is instead to quantify the risk of flawed inference by comparing the estimates from two 'naïve' regression methods versus using the MR-approach (the 'informed method'). With the two naïve methods, we use the minimum life span of individuals in the field as the fitness measure. This is a common fitness measure used by evolutionary ecologists. We evaluate these different approaches on both real and simulated data sets. We primarily focus on stabilizing selection, as this mode of selection is of principal interest on both micro- and macro-evolutionary time-scales (Haller & Hendry 2013). Stabilizing selection is also notoriously difficult to estimate, due to low statistical power in most field studies with limited sample sizes (Kingsolver *et al.* 2001) and the fact that such stabilizing selection tends to 'erase its traces' by removing variation around the fitness optimum (Haller & Hendry 2013).

Our main finding is that all methods lose statistical power with decreasing recapture probability, and trait-dependent recapture probability becomes an issue mainly when recapture probabilities are trait-dependent and low. Note that we use 'trait dependence' in the broad sense, such as when recapture probabilities are heterogeneous in some way. Such heterogeneous recapture probabilities are not restricted to quantitative phenotypic traits, but could also include time or discrete phenotypic categories, such as sex effects. Nevertheless, it is often hard to tell *a priori* if trait dependence or low recapture probability is an issue within a data set, making the use of MR methods the safest choice even in scenarios when it may only have low impact on the conclusions. We provide R scripts and a new R package EasyMARK (Supporting: EasyMARK) along with this article to facilitate research in this area.

Methods

SIMULATING INDIVIDUAL CAPTURE HISTORIES

We developed a simple mark–recapture simulator, which produced individual capture histories under a range of different conditions. The simulator was written in the R programming language v3.0.2 (R Development Core Team 2008). The situation we model can be thought of as a short-lived organism (e.g. an insect), with no or negligible senescence. For more information, see Supporting Information: Simulating Individual Capture Histories (Fig. S1). Simulations were run in parallel using the r package foreach v1.4.1 (Revolution Analytics & Weston 2013) on a Windows 7, Intel Core i7-3930K CPU @ 3.20 GHz and 32 GB of RAM.

Two scenarios were simulated as follows: (i) linear trait-dependent recapture probability and (ii) negative quadratic ('stabilizing') selection (Table 1). These data sets will be referred to as trait-dependent recapture probability and stabilizing selection, respectively. These two scenarios represent two opposite cases:

- (1) trait-dependent simulations: only trait-dependent recapture probability and no actual survival selection (Table 1).
- (2) stabilizing selection simulations: only survival selection and imperfect detection, but no trait dependence on recapture probability (Table 1).

Our trait-dependent simulations had the following form:

$$\text{logit}(p) = I_p + b_p z \quad \text{eqn 1}$$

Here, p is survival probability, I_p is an intercept term, z is a normally distributed vector of trait values (mean = 0, var = 1), and b_p is the linear trait dependence coefficient. The linear trait dependence coefficient (b_p) was set in our simulator at 0.05, 0.15 and 0.25, which we refer to as 'low', 'medium' and 'high' trait dependence, respectively (but see Figs S6–S7, where we also tested with higher values). Unfortunately, it is generally unknown what a typical or reasonable value in practice for this coefficient would be. The intercept term (I_p) controlled the mean recapture probability. Mean recapture probability was varied using the intercept from 0.1 to 0.9 at a 0.1 interval, resulting in 9 mean recapture probability values (i.e. using values $I_p = -2.2, -1.39, -0.85, -0.41, 0, 0.41, 0.85, 1.39, 2.2$). The normally distributed continuous trait (z) had mean zero and variance of one. Survival probability (Φ) was set constant at 0.70 and was not dependent on the trait (z). These recapture probabilities (p) and survival probabilities (Φ) for each individual were then used to create individual capture histories.

Our stabilizing selection simulations had the following form:

$$\text{logit}(\Phi) = I_\Phi + b_\Phi z + g_\Phi z^2 \quad \text{eqn 2}$$

Here, Φ is the survival probability, I_Φ is an intercept term, z is a normally distributed vector of standardized phenotypic trait values (mean = 0, var = 1), b_Φ is the linear coefficient, and g_Φ is the quadratic coefficient. When simulating stabilizing selection, the linear coefficient (b_Φ) was set to zero, which is equivalent to assuming that the popula-

Table 1. Summary of the simulations and assumptions of the two scenarios

Simulation	Selection of Survival	Recapture Probability
Stabilizing selection	Yes	Not dependent on trait
Trait-dependent recap. prob	No	Dependent on trait

tion sits on its adaptive peak and is no longer subject to directional selection. The quadratic coefficient (g_Φ) was set at values of -0.03 , -0.07 and -0.11 . This corresponds to transformed values of gamma (γ) of -0.06 , -0.12 , -0.18 in terms of classical variance-standardized quadratic selection gradients (Lande & Arnold 1983; Janzen & Stern 1998; Stinchcombe *et al.* 2008; Morrissey & Sakrejda 2013; M. Morrissey, pers. comm.). See Supporting Information: Comparing Estimates.

Our simulated selection strengths are similar in magnitude to those that have been estimated previously in natural populations and which were included in meta-analyses. These estimates should therefore presumably reflect the true average strength of selection in natural populations (Hoekstra *et al.* 2001; Kingsolver *et al.* 2001). An intercept (I_Φ) of one was used in survival simulations, resulting in a mean life span of about 3 days (occasions), maximum life span of around 20 days, and an average daily survival probability of approximately 0.70 (Fig. S1). The recapture probability (p) was varied from 0.1 to 0.9 at a 0.1 interval, resulting in nine recapture probability values, which were constant and not dependent on the trait (z).

Three sample sizes were used for both data sets: 100, 250, and 500 individuals. A 1000 runs were made for each combination of recapture probability, sample size, and selection or trait dependence strength.

Minimum life span was calculated as the time elapsed from marking to the last day detected. For example, individuals with capture histories of '1010100', '1111100' and '1000100' would all have minimum life spans of 5 days. The full output and associated R scripts will be uploaded to DRYAD (<http://datadryad.org/>). For more information, see Supporting: Simulating Individual Capture Histories. Our simulated life spans resulted in a non-normal and highly skewed survival distribution (Fig. S1), that was very similar to the distribution of estimated life spans from the field data set. An implementation of this simulator is also available in our R package EasyMARK. See Supporting: EasyMARK.

FITTING THE MODELS TO THE SIMULATED DATA

All analyses were conducted in R. We fitted two general linear model types from the R base and MASS packages with different assumptions about the underlying distribution of errors (normal and negative binomial) and a mark–recapture model from the R package mra v2.13 (McDonald 2012; Table 2). Hereafter, these three approaches will be referred to as LA ('Lande-Arnold'), NB (negative binomial), and MR (mark–recapture). Note that we use the term 'Lande-Arnold method' here to mean a regression model with a Gaussian error structure. The LA- and NB-approaches can both be seen as naïve linear methods which do not take recapture probability into account and which only use minimum life span as the fitness measure. In contrast, the MR-method is a more informed method, which corrects for recapture probability, and the fact that individuals that are still alive might not be observed every time they are looked for by the investigator in the field. A general goal in this study is to compare these two naïve methods with the MR-method, and to understand how robust the former are with respect to deviations from the assumptions of perfect detection, as past survival selection studies have often assumed perfect detection.

There are very close relationships between per-interval logistic-scale regression coefficients of survival, and log-scale regression coefficients, such as those generated by our NB analysis, and selection gradients acting through longevity (Morrissey & Sakrejda 2013; M.B. Morrissey & I.B.J. Goudie, unpubl. data; M. Morrissey, pers. comm.; See Supporting information for our implementation). We applied the formulae given in Morrissey and Goudie equations 7 and

Table 2. Summary of methods and their abbreviations in this study

Method	Full Name	Fitness Measure	Recapture Probability
LA	Lande-Arnold	Minimum Life span	Ignores
NB	Negative Binomial	Minimum Life span	Ignores
MR	Mark–Recapture	Survival Probability	Corrects

8 to the logistic and log-scale regression coefficients from our MR and NB analysis to render them directly comparable to the selection gradients estimated by LA.

For LA, but not for NB or MR, the fitness response variable of each individual was divided by the sample mean fitness (Lande & Arnold 1983). In our case, this meant that the minimum life span of each individual was divided by the mean life span of the sample. Additionally, for LA, but not for NB or MR, we multiplied the resulting quadratic coefficient by two (Stinchcombe *et al.* 2008). Furthermore, for NB and MR, but not for LA, we needed to transform the estimates numerically to the data scale (Morrissey & Sakrejda 2013; M.B. Morrissey & I.B.J. Goudie, unpublished manuscript; M. Morrissey, pers. comm.). We will use beta (β) and gamma (γ) to denote estimates that represent selection gradients interpretable within the classical quantitative genetics framework. We address these transformation issues more in Supporting information: Comparing Estimates.

Mark–recapture was implemented with the R package mra (McDonald 2012). In addition to the assumptions of standard MR models (Cormack 1964; Seber 1965; Jolly 1982; Lebreton *et al.* 1992), we did not assume any time effects in our mark–recapture model. We direct readers to mra's documentation (McDonald 2012) and the source literature (Cormack 1964; Seber 1965; Jolly 1982; Lebreton *et al.* 1992) for more details. We chose the negative binomial over a Poisson distribution because we observed zero-inflation and over-dispersion when attempting to fit a Poisson distribution (Fig. S1; Hoef & Boveng 2007).

ANALYSIS OF THE SIMULATED DATA

With the simulated data sets, we were interested in which method (LA, NB, MR) would perform better in our two scenarios (i) inferring correctly that there is no survival selection in the case of the trait-dependent recapture data sets and (ii) inferring correctly stabilizing selection in the case where there was no trait dependence on recapture probability, but only stabilizing selection on survival probability.

With the trait-dependent simulations, for LA and NB we fitted two model types: the linear model ($w = I_w + b_w z$) and intercept only model ($w = I_w$), the latter assuming no fitness dependence on the trait (i.e. trait neutrality). Here relative longevity (w) is minimum life span divided by the mean life span of the sampled population. We counted a run as failed for LA and NB if the linear model ($w = I_w + b_w z$) was the top performing model by ΔAIC value > 2 . For MR, we fitted four model types.

Two of these models included b_Φ terms:

$$\text{logit}(p) = I_p; \text{logit}(\Phi) = I_\Phi + b_\Phi z \quad \text{eqn 3}$$

$$\text{logit}(p) = I_p + b_p z; \text{logit}(\Phi) = I_\Phi + b_\Phi z \quad \text{eqn 4}$$

Two models did not include the b_Φ terms:

$$\text{logit}(p) = I_p + b_p z; \text{logit}(\Phi) = I_\Phi \quad \text{eqn 5}$$

$$\text{logit}(p) = I_p; \text{logit}(\Phi) = I_\Phi \quad \text{eqn 6}$$

For MR, we counted a trial as failed if a model with a linear survival term (b_{Φ}) was the top performing model by at least two ΔAIC greater than a model without that term.

For the stabilizing selection simulations, for LA and NB we compared nested models in ordered complexity: Intercept only ($w = I_w$), intercept plus linear term ($w = I_w + b_w z$), and intercept and linear plus quadratic term ($w = I_w + b_w z + g_w z^2$). We considered the run successful, for LA and NB, if the model with the g_w term was the best model by at least two ΔAIC. For MR, we compared three models:

$$\text{logit}(p) = I_p; \text{logit}(\Phi) = I_\Phi + b_\Phi z + g_\Phi z^2 \quad \text{eqn 7}$$

$$\text{logit}(p) = I_p; \text{logit}(\Phi) = I_\Phi + b_\Phi z \quad \text{eqn 8}$$

$$\text{logit}(p) = I_p; \text{logit}(\Phi) = I_\Phi \quad \text{eqn 9}$$

Similarly, we considered a run successful, for MR, if the model with the g_Φ term was the best model by at least two ΔAIC than the next best model.

Transformed estimate values from each method (LA, NB, MR) were all compared at each recapture probability and trait dependence or selection strength. For the trait-dependent recapture probability simulations, we looked at estimates of linear selection (β), and for the stabilizing selection simulations, we looked at estimates of gamma (γ).

QUANTIFYING SURVIVAL SELECTION ON CALOPTERYX DAMSELFLIES IN THE FIELD

We also evaluated the three approaches on a field data set of marked individuals of damselflies (Insecta: Odonata). This field data set comes from an individually marked population of the banded demoiselle (*Calopteryx splendens*) in the southern part of the Swedish province Skåne at Södtemölla gård near Sövdesjön. This study population is located within the same river system (Klingavälsån) as previous work that has been performed on *C. splendens* further upstream (Svensson, Eroukhmanoff & Friberg 2006; Svensson & Friberg 2007; Wellenreuther, Vercken & Svensson 2010; Wellenreuther, Larson & Svensson 2012). All data in this study comes from the field season of 2012 (June and July). From this data set, we only analysed male *C. splendens* individuals, which at this site had a recapture probability of only around 10%, and a daily mean survival probability of around 80% (see Supporting information: Field Data set, Fig. S2).

Table 3. Results from the field dataset. Survival selection for total body length (tbl) and front wing length (fwl). Note that all MR models also included a term controlling for trait-dependent recapture probability $\text{logit}(p) = I_p + \text{tbl} + \text{fwl}$. The delta column shows the ΔAIC from the top rank model. Selection estimates were taken from the full model for each method: NB and LA: $w = I_w + \text{tbl} + \text{fwl} + \text{tbl}^2 + \text{fwl}^2$. MR: $\text{logit}(p) = I_p + \text{tbl} + \text{fwl}; \text{logit}(\Phi) = I_\Phi + \text{tbl} + \text{fwl} + \text{tbl}^2 + \text{fwl}^2$. See Supporting: Analysis of the Field Dataset for more details

MR	MR.		LA.		NB.				
	delta	LA	delta	NB	delta	Term	MR.est	LA.est	NB.est
$I + \text{fwl}$	0	$I + \text{fwl}$	0	$I + \text{fwl}$	0	tbl	-0.012	0.184	0.134
$I + \text{tbl} + \text{fwl}$	1.9	$I + \text{tbl}$	0.2	$I + \text{tbl}$	0.7	fwl	0.34	0.181	0.217
$I + \text{tbl} + \text{fwl} + \text{fwl}^2$	2.3	$I + \text{tbl} + \text{fwl}$	0.9	$I + \text{tbl} + \text{fwl}$	1.4	tbl^2	-0.028	-0.118	-0.208
$I + \text{tbl} + \text{fwl} + \text{tbl}^2$	3.2	$I + \text{tbl} + \text{fwl} + \text{tbl}^2$	2.5	$I + \text{tbl} + \text{fwl} + \text{tbl}^2$	2.1	fwl^2	-0.018	-0.004	0.043
$I + \text{tbl} + \text{fwl} + \text{tbl}^2 + \text{fwl}^2$	4.2	$I + \text{tbl} + \text{fwl} + \text{fwl}^2$	2.9	I					
$I + \text{tbl}$	7.8	I	3	$I + \text{tbl} + \text{fwl} + \text{fwl}^2$	3.1				
I	10.5	$I + \text{tbl} + \text{fwl} + \text{tbl}^2 + \text{fwl}^2$	4.5	$I + \text{tbl} + \text{fwl} + \text{tbl}^2 + \text{fwl}^2$	4.1				

SELECTION ANALYSIS OF THE FIELD DATA

The *C. splendens* field data set was analysed in a similar manner as the simulated data. We primarily focus on two traits: total body length (tbl) and front wing length (fwl), although we also followed up this analysis by estimating selection for a total of 11 phenotypic traits in these damselflies. We fitted the methods described above (LA, NB, MR) to the individual recapture histories, while also accounting for trait-dependent recapture probability on tbl and fwl (Table 3). See Supporting information: Analysis of Field Data set.

Results

SIMULATED DATA SETS

With the stabilizing selection data set, all methods (LA, NB, MR) lost statistical power to infer the correct form of selection with decreasing recapture probability (Fig. 1). Under conditions of weak selection and low sample sizes, no method performed well in detecting stabilizing selection, even when it was present (Fig. 1). In contrast, under strong selection and at high recapture probabilities, all methods performed similarly (Fig. 1), although MR generally performed better at inferring stabilizing selection across most recapture probabilities (Fig. 1).

Estimates of stabilizing selection between the methods (LA, NB, MR) were largely in agreement (Fig. 2). However, the estimates from NB and LA became stronger (more negative) with decreasing recapture probability (Fig. 2). This is likely reflects that minimum observed longevity divided by true longevity varies with true longevity (M. Morrissey, pers. comm.). In addition, MR estimates of stabilizing selection tended to have lower variance than the naïve methods.

In the simulations where we incorporated trait-dependent recapture probabilities, all methods (LA, NB, MR) performed similarly at high recapture probabilities, but LA and NB began to fail at low recapture probabilities and high trait dependence (Figs 3, S6–S7). Although the estimates of β , from all three methods were similar at high recapture probabilities, LA and NB were biased at lower recapture probabilities and when trait dependence was high (Figs 4, S7).

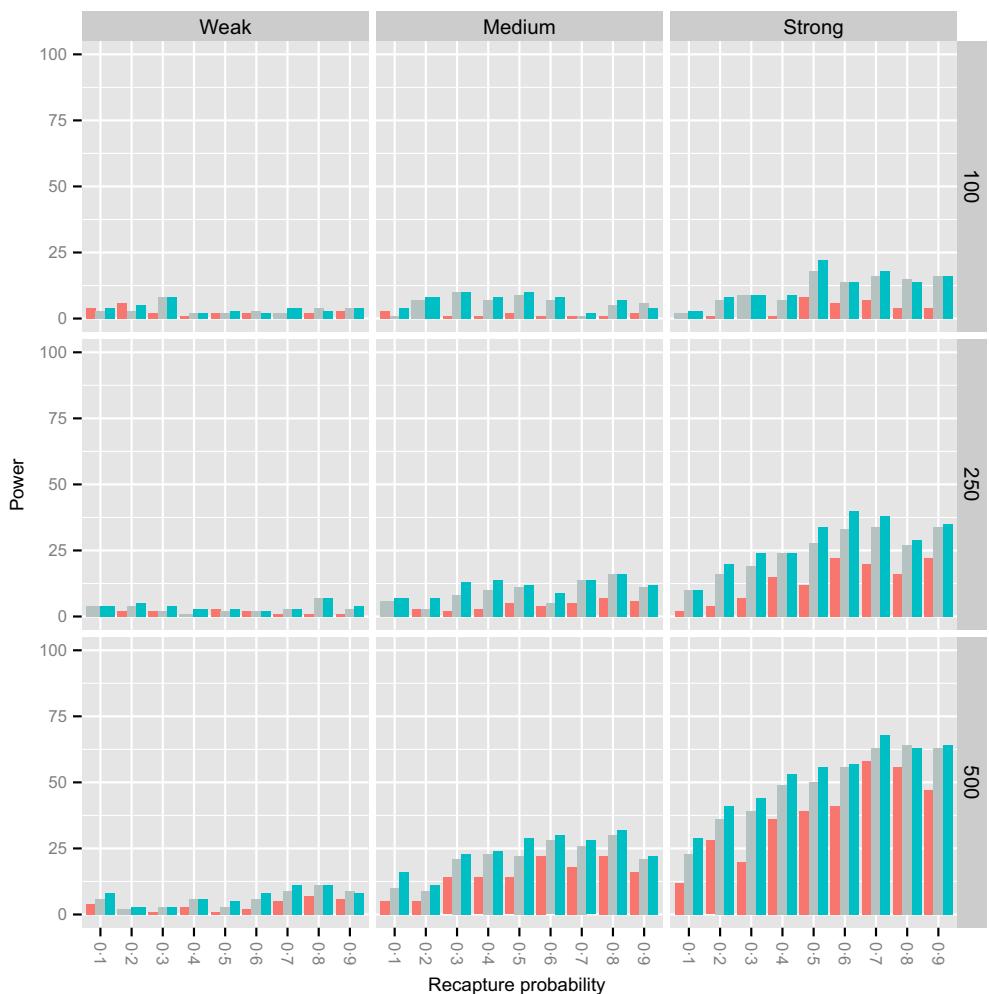


Fig. 1. Power to detect stabilizing selection in the simulated data set using three different methods: Lande–Arnold (LA; red), negative binomial (NB; gray) and mark–recapture (MR; blue). Each bar shows the percentage of times stabilizing selection was detected in our simulations under different conditions of recapture probability and selection strength. The right panel shows the sample sizes used (100, 250, 500 individuals). The upper panel shows the true underlying stabilizing selection gradient (γ : weak = -0.06 , medium = -0.12 , strong = -0.18). The x-axes show how we varied recapture probability from 0.1 (10% daily recapture probability) to 0.9 (nearly perfect detection). The y-axes show the percentage of runs when stabilizing selection was detected using the AIC-criterion ($\Delta\text{AIC} > 2$). Each bar represents 1000 runs. b_{ϕ} values were set to 0 for all simulations in this figure, which is equivalent to assume that directional selection is absent and the population is sitting on its adaptive peak.

FIELD DATA SET

Based on AIC, all methods (LA, NB, MR) ranked the ‘model: $I + \text{fwl}$ ’ highest, that is, linear selection only on front wing length (fwl; Table 3). The naïve methods (LA, NB) tended to favour a models including total body length (tbl) more than MR (Table 3). Additionally, transformed

estimates of linear and quadratic selection on total body length and front wing length tended to differ between the naïve methods (LA, NB) and MR, suggesting that there might be some trait-dependent effects on recapture probability in this data set (Tables 3, S1). Including an interaction between fwl and tbl did not improve model fit, so it was excluded for simplicity. Comparing selection estimates

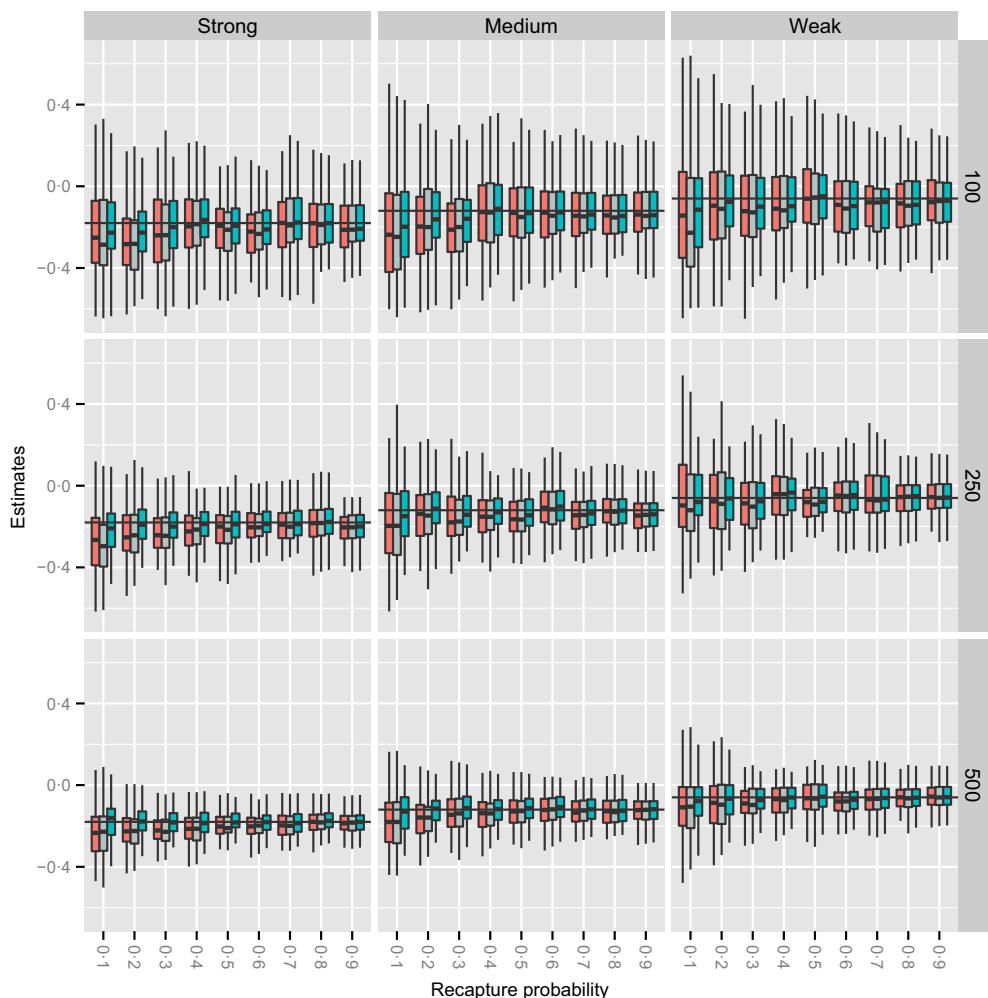


Fig. 2. Estimates of stabilizing selection in the simulated data set using three different methods: Lande–Arnold (LA; red), negative binomial (NB; grey) and mark–recapture (MR; blue). The upper panel shows the three true underlying stabilizing selection gradient categories (γ : weak = -0.06 , medium = -0.12 , strong = -0.18). Each box is 1000 simulated runs. The solid line indicates the true underlying selection gradient. The upper and lower limits of the boxplot give the 25th and 75th percentiles, respectively.

obtained from MR with those from the LA-approach for 11 traits in *C. splendens* revealed some concordance between the two (Fig. 5).

Discussion

Here, we have shown that the main advantage of MR methods is their ability to control for trait-dependent recapture probability (Figs 3, S6–S7). Although trait-dependent effects are mitigated at high recapture probabilities, MR methods are likely the safest choice, since it is often unclear *a priori* how

strongly recapture probability depends on a given trait (Figs S6–S7). There are also certainly power advantages of MR over naïve methods at most recapture probabilities. However, we note that even MR loses significant power at low recapture probabilities (Figs 1, S3–S4).

Estimates of stabilizing selection by our naïve methods (LA, NB) were somewhat biased (becoming stronger and more negative) with decreasing recapture probability (Fig. 2), while MR estimates were largely unaffected and unbiased. This negative bias produced by naïve methods became particularly pronounced when stabilizing selection became stronger and

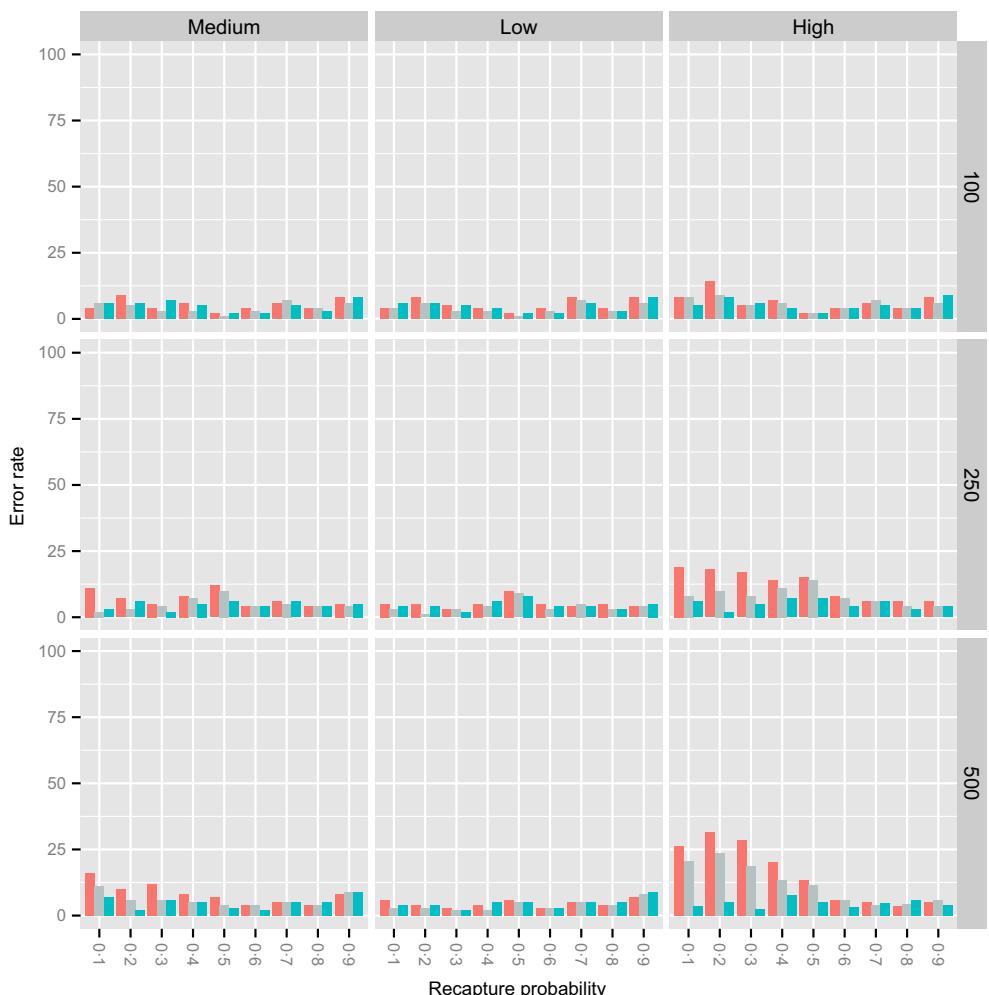


Fig. 3. Error rate from the trait dependence data set, using three different methods: Lande–Arnold (LA; red), negative binomial (NB; grey) and mark–recapture (MR; blue). Each bar represents the percentage of times a method failed our performance test. For a method to fail our performance test, it had to significantly prefer ($\Delta\text{AIC} > 2$) a model with a linear survival term (b_p). The trait dependence term (b_p) 0.05, 0.15 and 0.25, which we refer to as ‘Low’, ‘Medium’ and ‘High’ dependence respectively (See Fig. S7 for higher values). We see here the naïve methods (NB and LA) fail at high trait dependence and low recapture probabilities. The x-axes show how we varied recapture probability from 0.1 (10% daily recapture probability) to 0.9 (nearly perfect detection). Each bar represents 1000 runs.

when recapture probabilities lower (see Fig. S4), although the length of the study period also affected bias in stabilizing selection estimates in complex ways (Fig. S9). We note that such bias could potentially have affected the results and influenced the conclusions from past meta-analyses of the strength of stabilizing selection (e.g. Kingsolver *et al.* 2001). However, without empirical field data from natural populations about the magnitude recapture probabilities it is impossible to judge how severe this potential issue might be. Although not presented here, it is very likely that naïve methods might also become

biased when estimating other modes of selection (directional and disruptive).

When recapture probability was not dependent on any phenotypic traits, all methods (LA, NB, MR) performed somewhat similarly, especially at high recapture probability (Fig. 3). However, no study can usually be sure *a priori* that there is not significant trait-dependent recapture probability present when recapture probability is less than one (Figs 3, S6–S7). Therefore, the safest choice would be to utilize MR methods even in scenarios where it might only have limited

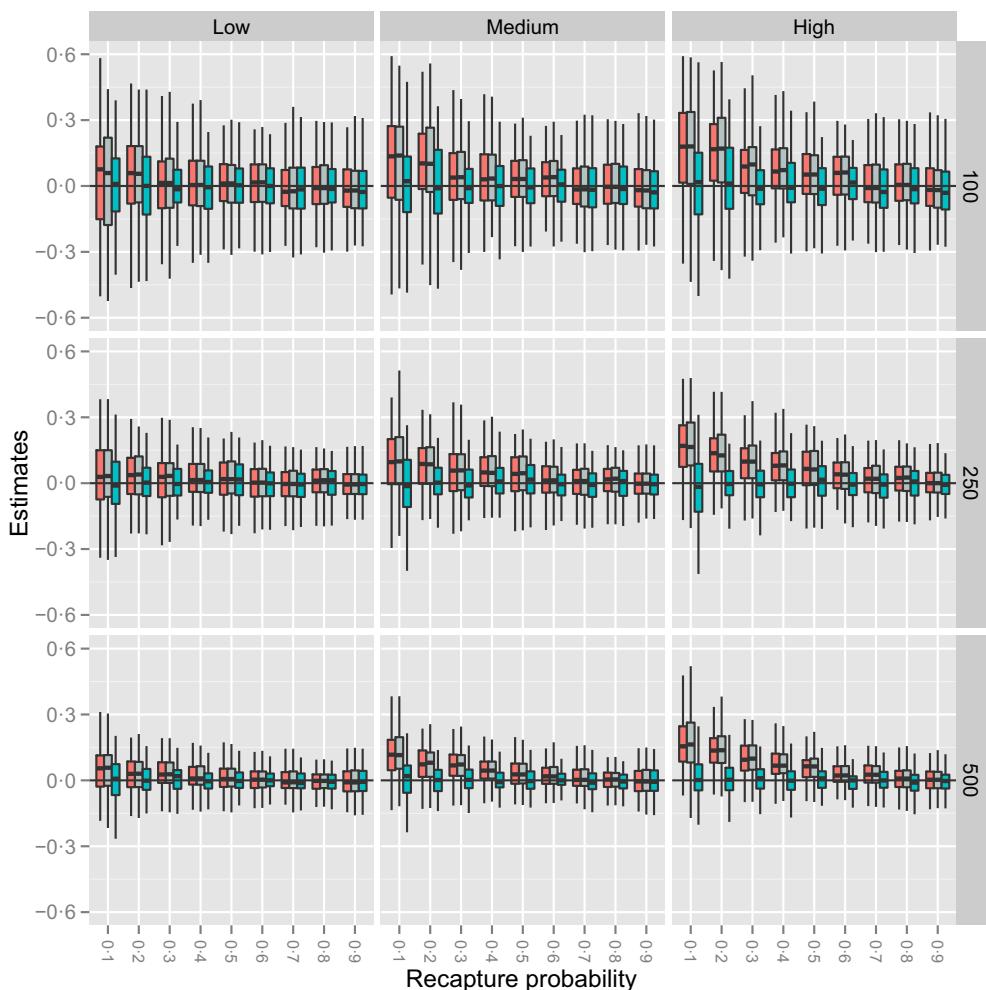


Fig. 4. Estimates of linear selection from the simulations incorporating trait-dependent recapture probability, using three different methods: Lande–Arnold (LA; red), negative binomial (NB; grey) and mark–recapture (MR; blue). Since our test with the mark–recapture method had two models with a b_Φ term, we present only estimates from the model that accounts for trait-dependent recapture probability ($\text{logit}(p) = I_p + b_p z$; $\text{logit}(\Phi) = I_\Phi + b_\Phi z$). The true underlying selection ($\beta = 0$) is indicated with a solid line. b_p : Weak = 0.05, medium = 0.15, strong = 0.25. Each box is 1000 simulated estimates. The upper and lower limits of the boxplot give the 25th and 75th percentiles, respectively.

impact on the conclusions of the study, such as with high recapture probability and low sample sizes.

An alternative procedure for assessing whether MR methods are necessary would be to fit initially a naïve method, and if survival significantly depends of the observed traits, one could verify using MR methods that the observed effect is not simply an effect of trait-dependent recapture probability. However, even this procedure would not control for alternative scenarios in which recapture probability and survival probability are dependent on the observed trait but in opposite directions.

Also in these types of scenarios, MR methods are still more powerful than naïve methods at most recapture probabilities (Fig. 1).

Mean survival probability also affected statistical power to detect selection. The simulations we have presented in this study were generated with a mean daily survival probability of 70%, but when this value was lowered to 30%, all methods (LA, NB, MR) lost power across all recapture probabilities (see Figs S3–S4). Similarly ending the study period early, before all individuals have died, reduces power in all methods,

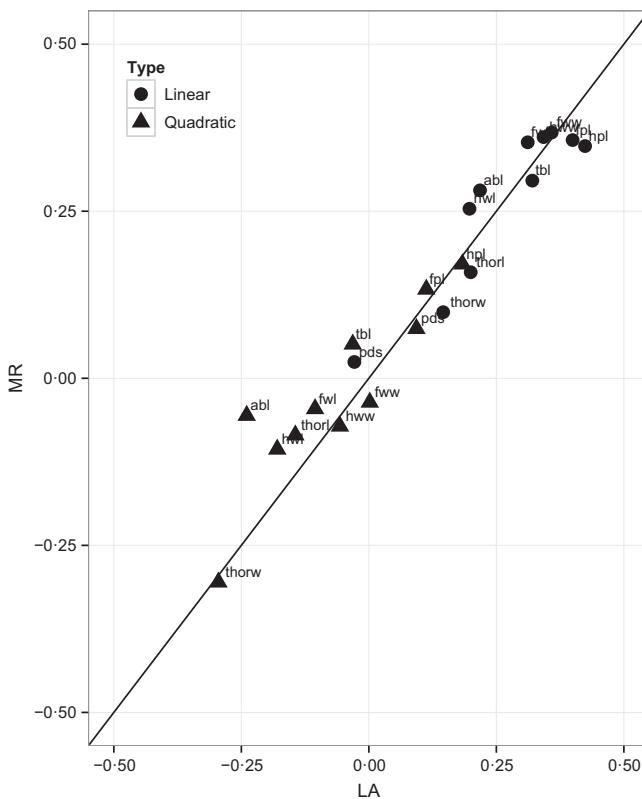


Fig. 5. Estimates of selection on eleven phenotypic traits from an empirical field data set (males in the banded demoiselle *Calopteryx splendens*). Shown is a comparison of the estimates from the classical LA approach (assuming perfect detection) and MR (taking imperfect detection into account). Each point represents an estimate of selection on an individual trait using either approach. The line represents a hypothetical perfect concordance between the approaches (slope = 1, intercept = 0). Estimates were produced from univariate models including only the focal trait (e.g. for LA: $w = I + \text{trait} + \text{trait}^2$; and for MR: $\text{logit}(p) = I + \text{trait}$, $\text{logit}(\Phi) = I + \text{trait} + \text{trait}^2$) and no other terms, so that a total of 22 separate models were estimated.

but affects naïve estimates of selection more severely (see Figs S8–S9).

The vast majority of past survival selection studies have assumed perfect detection (Conner 2001; Kingsolver *et al.* 2001). Given the weak statistical power of most past selection studies and the limited samples sizes in these (Fig. 1), the general picture of the strength of survival selection is in nature (Kingsolver *et al.* 2001) is unlikely to change with the implementation of MR methods. This is of course, assuming that trait-dependent recapture probability is not a widespread and strong force in these historical data sets, a topic that deserves further empirical attention. Of course, recapture probabilities need not only to be *trait*-dependent. Other heterogeneous aspects of recapture probability can cause problems when they are ignored. Future individual studies will nevertheless benefit from utilizing MR methods.

Comparing the selection estimates between the three methods is challenging. The coefficients typically estimated using statistical methods such as MR and NB, will not agree and are not strictly comparable with selection gradients used in quantitative genetic theory (e.g. LA; Morrissey & Sakrejda 2013; M.B. Morrissey & I.B.J. Goudie, unpubl. data; M. Morrissey, pers. comm.). All estimates used in this study have therefore been transformed to become comparable between the different

methods on different scales and equivalent to selection gradients in the classical quantitative genetic framework (M.B. Morrissey & I.B.J. Goudie, unpubl. data).

Our analyses based on the field data set of *C. splendens* males were broadly in agreement with the results and conclusions from our simulations (Table 1; Fig. 5). Since our field data set had low recapture probability, at around 10%, and low to intermediate sample size ($N = 276$), our results are consistent with the situation in our simulations with low sample size and low recapture probabilities. One concern is that initial guesses about recapture probability are by no means intuitive and easily understood outside the MR-framework. With our field data set, for example, our initial guess was that we would have a daily recapture probability of about 50%, given a relatively high fraction of marked individuals that were seen at least once again in the field after the marking occasion. However, after analysing the individual field histories using MR, we found that the daily recapture probability turned out to be only 10%. We made our guess after several years of study and first-hand experience in the field. We found that recapture probability was trait-dependent for total body length (tbl) and front wing length (fwl) but with opposite sign ($b_{p, \text{tbl}} = 0.281$, $b_{p, \text{fwl}} = -0.482$; Table S1). This caused MR to prefer 'model: $I + \text{fwl}$ ' more strongly than our naïve methods, which also

ranked ‘model: $I + fwl$ ’ highest, but with lower ΔAIC . While all methods ranked the ‘model: $I + fwl$ ’ highest, the estimates of γ and β , along with ΔAIC values diverged enough from the MR results for some concern.

Conclusions

All statistical methods (MR, LA, NB) lost power with decreasing recapture probability. Generally, no statistical method can replace the value of choosing a good study system with high recapture probability and sufficiently large sample sizes, in combination with careful study designs. Here we find that MR methods, as expected, are probably the best statistical choice in situations where recapture probability is low or when recapture probability is trait-dependent. The two naïve methods (NB and LA) that assumed perfect detection and used minimum life span performed well at high recapture probabilities. However, we stress that one can never be certain that there is not high trait dependence present within their data without the use of MR methods (Figs S6–S7). Although at present mark-recapture methods are generally more difficult to implement than our naïve methods (LA, NB), their ease of implementation are likely to improve in the future (Gimenez *et al.* 2006; Gimenez *et al.* 2007; Gimenez, Grégoire & Lenormand 2009; King 2012). We offer our R package EasyMARK as a step in this direction.

Acknowledgements

We are grateful to Thomas Gosden, Lesley Lancaster, Maren Wellenreuther, and Beatriz Willink for constructive criticisms on the first draft of this manuscript. We also thank Johanna Eklund and Anna Norden for help in collecting the field data. We appreciate the comments and suggestions of the two reviewers of this manuscript, and would especially like to thank M. Morrissey whose comments and suggestions greatly improved the current study. Funding for this study was provided by grants from the ‘Carl Tryggers Foundation’ (CTS; contract number: 11:458) and The Swedish Research Council (VR; contract number: 621-2012-3768) to E.I.S.

Data accessibility

The simulation and field data sets along with associated R scripts can be accessed via DRYAD (<http://datadryad.org/>). Additional figures and supporting Results and discussion can be found in the online supporting information doi:10.5061/dryad.bcd0d8).

References

- Calsbeck, R., Gosden, T.P., Kuchta, S.R. & Svensson, E.I. (2012) Fluctuating selection and dynamic adaptive landscapes. *The Adaptive Landscape in Evolutionary Biology* (eds E.I. Svensson & R. Calsbeck), pp. 89–110. Oxford University Press, Oxford, UK.
- Clobert, J. (1995) Capture-recapture and evolutionary ecology: a difficult wedging? *Journal of Applied Statistics*, **22**, 989–1008.
- Conner, J.K. (2001) How strong is natural selection? *Trends in Ecology and Evolution*, **16**, 215–217.
- Cormack, R.M. (1964) Estimates of survival from sighting of marked animals. *Biometrika*, **51**, 429–430.
- Endler, J.A. (1986) *Natural Selection in The Wild*. Princeton University Press, Princeton, NJ, USA.
- Gimenez, O., Grégoire, A. & Lenormand, T. (2009) Estimating and visualizing fitness surfaces using mark-recapture data. *Evolution*, **63**, 3097–3105.
- Gimenez, O., Covas, R., Brown, C.R., Anderson, M.D., Brown, M.B. & Lenormand, T. (2006) Nonparametric estimation of natural selection on a quantitative trait using mark-recapture data. *Evolution*, **60**, 460–466.
- Gimenez, O., Rossi, V., Choquet, R., Dehais, C., Doris, B., Varella, H. & Pradel, R. (2007) State-space modelling of data on marked individuals. *Ecological Modelling*, **206**, 431–438.
- Gimenez, O., Viallefond, A., Charmantier, A., Pradel, R., Cam, E., Brown, C.R. *et al.* (2008) The risk of flawed inference in evolutionary studies when detectability is less than one. *American Naturalist*, **172**, 441–448.
- Gosden, T.P. & Svensson, E.I. (2008) Spatial and temporal dynamics in a sexual selection mosaic. *Evolution*, **62**, 845–856.
- Haller, B.C. & Hendry, A.P. (2013) Solving the paradox of stasis: squashed stabilizing selection and the limits of detection. *Evolution*, **68**, 483–500.
- Herford, J., Hansen, T.F. & Houle, D. (2004) Comparing strengths of directional selection: how strong is strong? *Evolution*, **58**, 2133–2143.
- Hoef, J.M.V. & Boveng, P.L. (2007) Quasi-poisson vs. negative binomial regression: how should we model overdispersed count data? *Ecology*, **88**, 2766–2772.
- Hoekstra, H.E., Hoekstra, J.M., Berrigan, D., Vignieri, S.N., Hoang, A., Hill, C.E., Beerli, P. & Kingsolver, J.G. (2001) Strength and tempo of directional selection in the wild. *Proceedings of the National Academy of Sciences*, **98**, 9157–9160.
- Janzen, F.J. & Stern, H.S. (1998) Logistic regression for empirical studies of multivariate selection. *Evolution*, **52**, 1564–1571.
- Jolly, G.M. (1982) Mark recapture models with parameters constant in time. *Biometrics*, **38**, 301–321.
- King, R. (2012) A review of Bayesian state-space modelling of capture–recapture–recovery data. *Interface Focus*, **2**, 190–204.
- Kingsolver, J.G. & Diamond, S.E. (2011) Phenotypic selection in natural populations: what limits directional selection? *American Naturalist*, **177**, 346–357.
- Kingsolver, J.G., Hoekstra, H.E., Hoekstra, J.M., Berrigan, D., Vignieri, S.N., Hill, C.E., Hoang, A., Gilbert, P. & Beerli, P. (2001) The strength of phenotypic selection in natural populations. *American Naturalist*, **157**, 245–261.
- Kingsolver, J.G., Diamond, S.E., Siepielski, A.M. & Carlson, S.M. (2012) Synthetic analyses of phenotypic selection in natural populations: lessons, limitations and future directions. *Evolutionary Ecology*, **26**, 1101–1118.
- Knapczyk, F.N. & Conner, J.K. (2007) Estimates of the average strength of natural selection are not inflated by sampling error or publication bias. *American Naturalist*, **170**, 501–508.
- Lande, R. & Arnold, S.J. (1983) The measurement of selection on correlated characters. *Evolution*, **37**, 1210–1226.
- Lebreton, J.D., Burnham, K.P., Clobert, J. & Anderson, D.R. (1992) Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecological Monographs*, **62**, 67–118.
- MacColl, A.D. (2011) The ecological causes of evolution. *Trends in Ecology and Evolution*, **26**, 514–522.
- McDonald, T. (2012) *mra: Analysis of Mark-Recapture data*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://CRAN.R-project.org/package=mra> [accessed 20 May 2014]
- Mitchell-Olds, T. & Shaw, R.G. (1987) Regression analysis of natural selection: statistical inference and biological interpretation. *Evolution*, **41**, 1149–1161.
- Morrissey, M.B. & Hadfield, J.D. (2012) Directional selection in temporally replicated studies is remarkably consistent. *Evolution*, **66**, 435–442.
- Morrissey, M.B. & Sakrejda, K. (2013) Unification of regression-based methods for the analysis of natural selection. *Evolution*, **67**, 2094–2100.
- R Development Core Team (2008) *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org> [accessed 20 May 2014]
- Revolution Analytics & Weston, S. (2013) *foreach: Foreach Looping Construct for R*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://CRAN.R-project.org/package=foreach> [accessed 20 May 2014]
- Seber, G.A.F. (1965) A note on multiple-recapture census. *Biometrika*, **52**, 249–250.
- Seber, G.A.F. & Schwarz, C.J. (2002) Capture-recapture: before and after EUR-ING 2000. *Journal of Applied Statistics*, **29**, 5–18.
- Siepielski, A.M., DiBattista, J.D. & Carlson, S.M. (2009) It’s about time: the temporal dynamics of phenotypic selection in the wild. *Ecology Letters*, **12**, 1261–1276.
- Siepielski, A.M., DiBattista, J.D., Evans, J.A. & Carlson, S.M. (2011) Differences in the temporal dynamics of phenotypic selection among fitness components in the wild. *Proceedings of the Royal Society B-Biological Sciences*, **278**, 1572–1580.
- Siepielski, A.M., Gotanda, K.M., Morrissey, M.B., Diamond, S.E., DiBattista, J.D. & Carlson, S.M. (2013) The spatial patterns of directional phenotypic selection. *Ecology Letters*, **16**, 1382–1392.

- Stinchcombe, J.R., Agrawal, A.F., Hohenlohe, P.A., Arnold, S.J. & Blows, M.W. (2008) Estimating nonlinear selection gradients using quadratic regression coefficients: double or nothing? *Evolution*, **62**, 2435–2440.
- Svensson, E.I., Eroukhmanoff, F. & Friberg, M. (2006) Effects of natural and sexual selection on adaptive population divergence and premating isolation in a damselfly. *Evolution*, **60**, 1242–1253.
- Svensson, E.I. & Friberg, M. (2007) Selective predation on wing morphology in sympatric damselflies. *American Naturalist*, **170**, 101–112.
- Svensson, E. & Sinervo, B. (2000) Experimental excursions on adaptive landscapes: density-dependent selection on egg size. *Evolution*, **54**, 1396–1403.
- Wade, M.J. & Kalisz, S.M. (1990) The causes of natural selection. *Evolution*, **44**, 1947–1955.
- Wellenreuther, M., Larson, K.W. & Svensson, E.I. (2012) Climatic niche divergence or conservatism? Environmental niches and range limits in ecologically similar damselflies. *Ecology*, **93**, 1353–1366.
- Wellenreuther, M., Vercken, E. & Svensson, E.I. (2010) A role for ecology in male mate discrimination of immigrant females in *Calopteryx* damselflies? *Biological Journal of the Linnean Society*, **100**, 506–518.
- Williams, B.K., Nichols, J.D. & Conroy, M.J. (2002) *Analysis and Management of Animal Populations*. Academic Press, San Diego, CA, USA.

Received 18 September 2015; accepted 16 October 2015

Handling Editor: Robert B. O'Hara

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Fig. S1. Histograms comparing minimum lifespan values generated from a random simulation and from the field dataset of damselflies (from male banded demoiselles, *Calopteryx splendens*).

Fig. S2. Quantifying morphological phenotypic variation in the banded demoiselle (*Calopteryx splendens*) from photographs.

Fig. S3. Power to detect stabilizing selection with low mean survival probability (0.3) using three different methods: Lande-Arnold

(LA; red), negative binomial (NB; gray), and mark-recapture (MR; blue).

Fig. S4. Estimates of stabilizing selection with low mean survival probability (0.3) using three different methods: Lande-Arnold (LA), negative binomial (NB), and mark-recapture (MR).

Fig. S5. False positive rates (Type I error) of our three methods (LA, NB, MR).

Fig. S6. Failure rates (Type II error) from additional trait-dependent simulations, using three different methods: Lande-Arnold (LA; red), negative binomial (NB; gray), and mark-recapture (MR; blue).

Fig. S7. Estimates of β from the trait dependent simulations using three different methods: Lande-Arnold (LA), negative binomial (NB), and mark-recapture (MR).

Fig. S8. Power to detect stabilizing selection when the study period is ended early: Lande-Arnold (LA), negative binomial (NB), and mark-recapture (MR). LA (red), NB (gray), MR (blue).

Fig. S9. Estimates of stabilizing selection when the study period is ended early: Lande-Arnold (LA), negative binomial (NB), and mark-recapture (MR).

Table S1. Output from our mark-recapture analysis on the field dataset.

Data S1. Simulation and field datasets.

Data S2. R scripts used in Study.

Paper II



The Odonate Phenotypic Database: a new open data resource for comparative studies of an old insect order

Authors

John Waller^{1,2}, Beatriz Willink¹, Maximilian Tschol¹, Erik I. Svensson^{1,2}

Affiliations

1. Lund University, Department of Biology

2. Authors for correspondence: erik.svensson@biol.lu.se and john.waller@biol.lu.se

Abstract

The Odonate Phenotypic Database is an online data resource for dragonfly and damselfly phenotypes (Insecta: Odonata). The database consists of a variety of morphological life-history behavioral traits, and biogeographical information collected from various sources in the literature. We see taxon-specific phenotypic databases as becoming an increasing valuable resource in comparative studies. Odonata is a relatively small insect order that currently consists of about 6400 species belonging to 32 families. Our database has at least some phenotypic records for 1011 of all 6400 known odonate species. The database is accessible at <http://www.odonatephenotypicdatabase.org/>.

Background & Summary

The Odonate Phenotypic Database is an online data resource for dragonfly and damselfly phenotypes (Insecta: Odonata). The database consists of a variety of morphological life-history behavioral traits, and biogeographical information collected from various sources in the literature. The database is not intended for species identification, but for comparative analysis within this insect group or to be combined with data from other taxonomic groups. The database is provided along with a large phylogenetic tree (1322 taxa, 21% of known odonates): “The Odonate Super Tree”. This phylogenetic tree was built using DNA-sequences from GenBank in combination with a traditional (morphologically-based) odonate taxonomy (see Waller and Svensson 2017).

Comparative analyses are becoming an increasing common part of evolutionary studies, as researchers attempt to bridge the gap between microevolutionary processes and macroevolutionary patterns (Arnold et al. 2001, Estes et al. 2007; Uyeda et al. 2011; Arnold 2014). Almost all comparative analysis rely on high-quality phenotypic data collected from the literature, and often a large amount of time is spent collecting such data. Alternatively, a variable is obtained from live individuals or museum specimens, but an important covariate is often needed from the literature (such as behaviour, habitat, or body size). It seems in the best interest of those working in the field to collect and curate these phenotypic observations, so they can be used in the future and combined with other sources of phenotypic information.

Paradoxically, as a community we have been much more successful at storing and cataloguing genotypes, and DNA-sequences often exist in GenBank for many species, but not even simple phenotypic data (such as body size) exists in an easily accessible form for many organismal groups. This lack of information is likely due to there being no clear structure in which to store phenotypic data. Phenotypic databases, because of the high-dimensional nature of most phenotypes, are also very different from a genetic database such as GenBank. Phenomics, as it has been termed, will always have to prioritize what aspects of the phenotypes to measure (Houle et al. 2010).

From an evolutionary viewpoint, phenotypes are arguably just as important and interesting as genotypes, if not more so (Houle et al. 2010, Kühl et al. 2013, Laughlin and Messier 2015), as selection operates on phenotypes, regardless of their genetic basis (Lande and Arnold 1983). Moreover, the increasingly integrative research practices in evolutionary biology will need not only access to high-quality genomic, molecular and phylogenetic resources, but will also need high-quality phenotypic and biogeographic data, fossil information for time-calibration of phylogenetic trees and other general data provided by biodiversity

informatics (Losos et al. 2013). Therefore, the difficulty of the task and the size of the project should not dictate the creation of structures needed to store the data.

One way forward is to create taxon-specific phenotypic databases, as we have done here. Having such databases available that focus on a certain taxonomic group, also allows the recorded phenotypes to be tailored to fit the needs of the specific group and have the advantages that trait definitions are less ambiguous (i.e., there is no need for a wing length variable in a mammal database).

Examples of such open taxon-specific databases with various forms of phenotypic, biogeographic and phylogenetic data include AmphiBIO for amphibian ecological traits (Oliveira et al. 2017), panTHERIA (Jones et al. 2009) and EltonTraits 1.0 (Wilman et al. 2014) for various mammalian traits, TRY for plants (Kattge et al. 2011) and birds (Dyer et al. 2017). However, in the case of animals, they are largely focused on vertebrate groups, while the most speciose animal group – the insects – do not have any such open database available, to our knowledge.

We see taxon-specific phenotypic databases as becoming an increasing valuable resource in comparative studies. Since our research background and expertise is in the order Odonata (dragonflies and damselflies), we have collected 34 phenotypic variables we see as useful to the research community (Table 1). Odonata is a relatively small insect order that currently consists of about 6400 species belonging to 32 families. Odonata are characterized morphologically highly conserved with respect to their overall external morphology (all representatives have six legs and four wings), but they also show considerable diversity in terms of wing and body colouration and shape (Fig. 1). Our database has at least some phenotypic records for 1011 of all 6400 known odonate species. The database is accessible at <http://www.odonatephenoypicdatabase.org/>.

Tabel 1.

Current approximate coverage of variables as of the publication of this document.

Table	Variables	Description	Coverage
Taxonomy	GenusSpecies Genus Species Family SubOrder	Taxonomy of record. Names are taken from the World Odonata List.	15%
Size	body_lengths forewing_lengths hind_wing_lengths	Body size in mm of species record.	13%
Body Colors	body_colors body_colorotypes body_patterns		15%
Behaviour	mate_guarding flight_mode territoriality	Mating and flight behaviour of species in general.	15%
Location and Habitat	continents aquatic_habitats climates ecozones habitat_openness	Location and habitat climates taken from range maps usually.	15%
Morphisms	sex_polymorphisms sex_dimorphisms geo_polymorphisms	Polymorphisms and strength of sexual dimorphism.	15%
Wing Pigment or Color	has_wing_pigment wing_pigment_extent_discrete wing_pigment_extent_continuous wing_pigment_pattern wing_pigment_symmetry wing_pigment_dimorphism wing_pigment_color wing_pigment_placement wing_pigment_color_type	Wing color and pigment variables.	15%

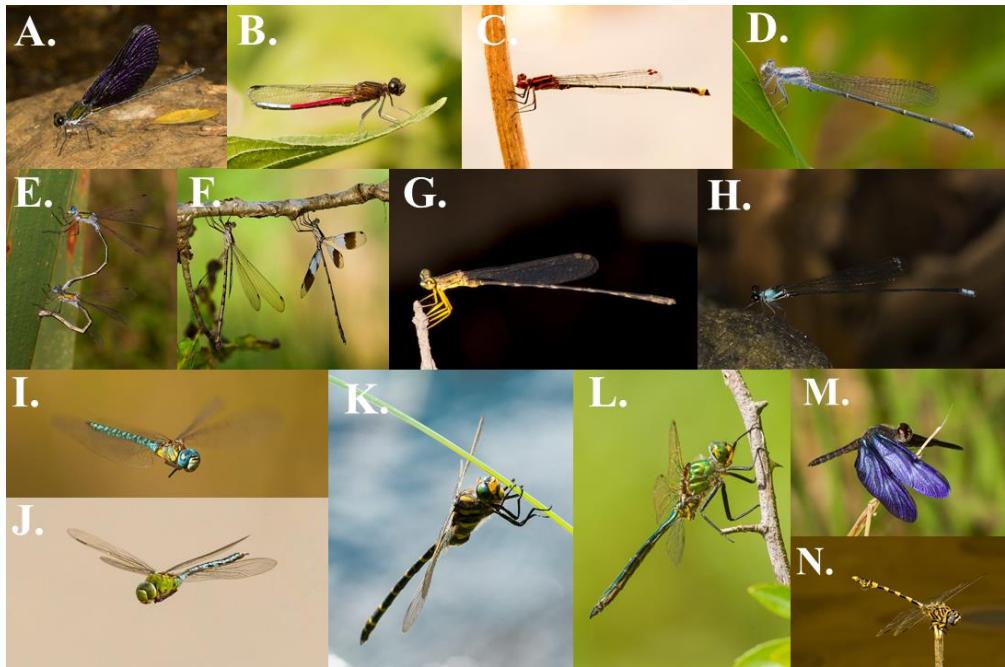


Figure 1 Phenotypic and taxonomic diversity

Phenotypic and taxonomic diversity of 12 representative families of Odonata, an insect order which currently encompasses c. a. 6400 species and a total of 32 families. All 32 odonate families are included in our molecular and time-calibrated phylogeny (Fig. 2; Waller and Svensson 2017). **A.** Family Calopterygidae: *Sapho orichalcea* (Cameroon, Africa, January 2017). **B.** Family Chlorocyphidae: *Chlorocypha curta* (Cameroon, Africa, January 2017). **C-D.** Family Coenagrionidae: **C.** *Acanthagrion adustum* (Guyana, South America, January 2015). **D.** *Argia moesta* (Texas, North America, April 2012). **E.** Family Lestidae: *Lestes sponsa* (Sweden, Europe, July 2010). **F.** Family Synlestidae: *Chlorolestes tessellatus* (Eastern Cape, South Africa, Africa, April 2010). **G.** Family Platycnemidae: *Copera congoensis* (Cameroon, Africa, February 2017). **H.** Family Protoneuridae: *Elattoneura balli* (Cameroon, Africa, January 2017). **I-J.** Family Aeshnidae: **I.** *Aeshna affinis* (Sweden, Europe, August 2010). **J.** *Anax imperator* (Sweden, Europe, August 2015). **K.** Family Cordulegasteridae: *Cordulegaster boltonii* (Sweden, Europe, July 2016). **L.** Family Corduliidae: *Somatochlora flavomaculata* (Sweden, Europe, June 2014). **M.** Family Libellulidae: *Zenithoptera fasciata* (Guyana, South America, January 2015). **N.** Family Gomphidae: *Ictinogomphus ferox* (Namibia, Africa, April 2017). All photographs by Erik Svensson.

Methods

Phenotypic data has been collected from the scientific literature and field guides. These literature sources are listed in Table 2. Phenotypes were scored by following a specific set of instructions for each variable. The descriptions of each variable can be found Data Records section below. The construction of the Odonate Super Tree has been described elsewhere (see Waller and Svensson 2017).

Table 2.

The primary references used to gather the data. Most of time these are field guides. Below is a list of resources that the database currently is based on.

Authors	Title and Publisher	Year
Theischinger and Hawking	Complete Field Guide to Dragonflies of Australia; CSIRO Publishing	2006
Kawashima and Futahashi	Dragonflies of Japan; Bunichi-Sogo Syuppan	2012
Dijkstra and Lewington	Field Guide to the Dragonflies of Britain and Europe; British Wildlife Publishing	2006
Paulson	Dragonflies and Damselflies of the West; Princeton University Press	2009
Samways	Dragonflies and Damselflies of South Africa; Penguin Random House South Africa	2008
Tarboton and Tarboton	A Fieldguide to the Dragonflies of South Africa; Warwick & Michèle Tarboton	2002
Heckman	Encyclopedia of South American Aquatic Insects: Odonata-Anisoptera; Springer	2006
Paulson	Dragonflies and Damselflies of the East; Princeton University Press	2011
Tang Hung Bun et al	A Photographic Guide to the Dragonflies of Singapore; Raffles Museum of Biodiversity Research	2010
Subramanian	Dragonflies and Damselflies of Peninsular India: A Field Guide; Indian academy of Sciences	2005
Hamalainen and Pinratana	Atlas of the Dragonflies of Thailand Distribution Maps by Provinces; Brothers of St. Gabriel in Thailand	1999
Esquivel	Dragonflies and damselflies of Middle America and the Caribbean; INBio	2006
Garrison et al	Dragonfly Genera of the New World An Illustrated and Annotated Key to the Anisoptera; John Hopkins University Press	2010
Nair	Dragonflies and Damselflies of Orissa and Eastern; Orissa Wildlife Organisation	2011
Tarboton and Tarboton	A Fieldguide to the Dragonflies of South Africa; Tarboton and Tarboton	2005
Okudaira et al	Dragonflies of the Japanese Archipelago in Color; Hokkaido UP	2001
Lencioni	Damselflies of Brazil: An illustrated identification guide 2 Coenagrionidae; F.A.A Lencioni	2006
Suhling and Martens	Dragonflies and Damselflies of Namibia; Gamsberg Macmillan	2007
Askew	The Dragonflies of Europe; Great Horkeley	1988
Dunkle	Dragonflies through Binoculars: A Field Guide to Dragonflies of North America (Butterflies Through Binoculars); Oxford University Press	2000
Manolis	Dragonflies and Damselflies of California; University of California Press	2003
Abbott	Dragonflies and Damselflies (Odonata) of Texas Volume 5; John C. Abbot	2011
Dijkstra and Clausnitzer	The Dragonflies and Damselflies of Eastern Africa; RMCA	2014
Tze-wai et al	The Dragonflies of Hong Kong; Cosmos	2011
Michalski	The Dragonflies and Damselflies of Trinidad and Tobago; Kanduanum Books	2015
Kompier	A Guide to the Dragonflies and Damselflies of the Serra dos Orgaos South-eastern Brazil; Regua publications	2015
Marinov and Waqa-Sakiti	An Illustrated Guide to Dragonflies of Viti Levu Fiji; The University of the South Pacific Institute of Applied Sciences	2013
Polhemus and Asquith	Hawaiian Damselflies: A Field Identification Guide; Bishop Museum	1996
Biggs	Common Dragonflies of California; Azalea Creek Publishing	2000

Code availability

All code used to generate the website and Odonate Super Tree are available in the supplementary material and on github (<https://github.com/jhnwllr/shiny-server/tree/master/odonates>).

Data Records

The following file have been deposited in Dryad. Additionally, up-to-date data can be explored and downloaded at: <http://www.odonatephenotypicdatabase.org/>.

1. opdb.csv
 - A flat data file 39 variables of all of the phenotypic data within the database. See Table 1 and Fig. 2 for a description and coverage of variables.
2. variable definitions.pdf A file that describes each of the variables within the opdb.csv and how they were collected.

Technical Validation

All of the phenotypic data were collected from published field guides or reliable internet sources. The field guides are listed in Table 2. All of the field guides have been published by respected odonatologists and experts on species identification. Our database is also not static, and additional data will be added as it becomes available. We encourage readers to contribute to this database by contacting the two authors for correspondence. We will accept data both from already published sources (e. g. scientific papers) even if it has already been deposited in other databases such as Dryad, as well as data that is not intended to be published elsewhere, as long as it can tailored to the format of the Odonate Phenotypic Database. Each species has a reference list, which lists the references from which the data was gathered, so it is possible to check each entry against these primary sources.

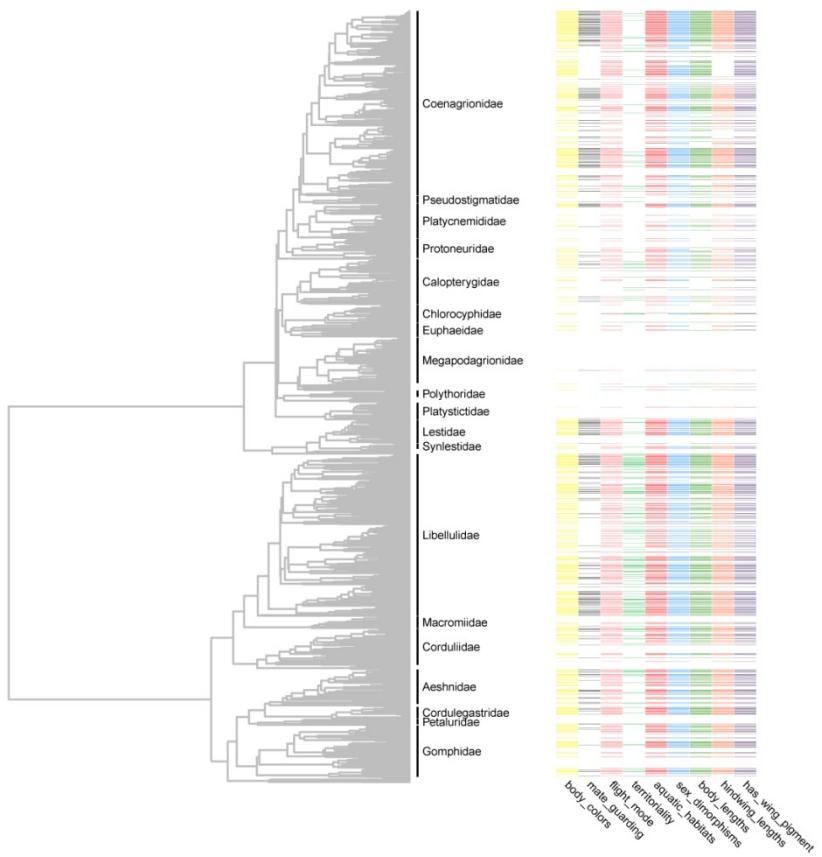


Figure 2

Our molecular time-calibrated phylogenetic tree, encompassing 1322 dragonfly and damselfly species (Waller and Svensson 2017a,b). Some representative data is annotated on the tree showing coverage of some of the variables in the database.

Usage Notes

The database is intended to be used in comparative analyses of odonate trait evolution. We therefore provide a previously published phylogenetic tree (Waller and Svensson 2017) along with the phenotypic data. Past problems that we have addressed using the phylogenetic comparative methods one these and similar data include the relationship between latitude and wing pigmentation (Svensson and Waller 2013) and macroevolutionary dynamics of body size (Waller and Svensson 2017). We have also investigated diversification dynamics (speciation and extinction rates) in relation to body size and wing pigmentation (Svensson and Waller 2013; Waller and Svensson 2017). Other potentially interesting future questions to pursue, include the evolutionary and ecological consequences of colour polymorphism in females (Svensson et al. 2005; Le Rouzic et al. 2015; Willink and Svensson 2017), biogeographical and ecological influences on various phenotypic traits such as body size and colouration and the role of climatic niche conservatism in the evolution and biogeography of odonates (Wiens et al. 2005; Wellenreuther et al. 2012).

Acknowledgements

We would like to acknowledge the students and laboratory assistants that helped us to collect the data from the literature, and in particular Anna Kell, Ev Poslin, Hanna Bensch, Robin Pranter, Kajsa Svensson, Lisa Winberg, Karolina Pehrson, Mireia Balesta, Tammy Ho. We would like to also thank the many authors of the field guides from which our database references. Our database is by no means nor intends to be a substitute for these valuable books.

Author contributions

John Waller: constructed the website <http://www.odonatephenotypicdatabase.org/> as part of his PhD-work, collected data, helped create variable definitions, wrote this manuscript.

Beatriz Willink: collected data, helped plan and organize data collection.

Maximilian Tschol: collected wing pigment data, created variable definitions.

Erik I. Svensson: created variable definitions, supervised data collection, came up with the original idea to create the database.

Competing interests

We declare no competing interests and the data published within this database is done so without bias.

References

- Arnold, S. J. (2014). Phenotypic Evolution: The Ongoing Synthesis. *The American Naturalist*. 183:729-746.
- Arnold, S.J., M. E. Pfrender, and A.G. Jones. (2001). The adaptive landscape as a conceptual bridge between micro-and macroevolution. *Genetica* 112: 9-32.
- Dyer, E. E., Redding, D. W., and T.M. Blackburn. (2017) The global avian invasions atlas, a database of alien bird distributions worldwide. *Scientific Data* 4, sdata201741.
- Estes, S., and S. J. Arnold. (2007). Resolving the paradox of stasis: models with stabilizing selection explain evolutionary divergence on all timescales. *The American Naturalist* 169.2: 227-244.
- Houle, D., Diddahally, R. G., and S. Omholt (2010). Phenomics: the next challenge. *Nature Reviews Genetics*. 11.12: 855.
- Jones K.E., Bielby J., Cardillo M., Fritz S.A., O'Dell J., Orme C.D., Safi K., Sechrest W., Boakes E.H., Carbone C., and C. Connolly. (2009). PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology*. 90:2648-2648.
- Kattge, J., Diaz, S., Lavorel, S., Prentice, I.C., Leadley, P., Bönisch, G., Garnier, E., Westoby, M., Reich, P.B., Wright, I.J. and J.H.C. Cornelissen. (2011). TRY—a global database of plant traits. *Global change biology*. 17:2905-2935.
- Kühl, H.S. and T. Burghardt. (2013). Animal biometrics: quantifying and detecting phenotypic appearance. *Trends in ecology & evolution*. 28:432-441.
- Lande, R. and S.J., Arnold. (1983). The measurement of selection on correlated characters. *Evolution*. 37:1210-1226.
- Laughlin, D.C. and J. Messier. (2015). Fitness of multidimensional phenotypes in dynamic adaptive landscapes. *Trends in Ecology & Evolution*. 30:487-496.
- Losos, J.B., Arnold, S.J., Bejerano, G., Brodie III, E.D., Hibbett, D., Hoekstra, H.E., Mindell, D.P., Monteiro, A., Moritz, C., Orr, H.A. and D.A. Petrov. (2013). Evolutionary biology for the 21st century. *PLoS Biology*. 11: p.e1001466.
- Oliveira, B.F., São-Pedro, V.A., Santos-Barrera, G., Penone, C., and G.C. Costa. (2017). AmphiBIO, a global database for amphibian ecological traits. *Scientific Data*. 4: p.sdata2017123.
- Svensson, E.I. and J.T. Waller. (2013). Ecology and sexual selection: evolution of wing pigmentation in calopterygid damselflies in relation to latitude, sexual dimorphism, and speciation. *The American Naturalist*. 182:174-195.

- Uyeda, J.C., Hansen, T.F., Arnold, S.J. and J. Pienaar. (2011). The million-year wait for macroevolutionary bursts. *Proceedings of the National Academy of Sciences*. 108:15908-15913.
- Waller J.T. and E.I. Svensson. (2017a). Data from: Body size evolution in an old insect order: no evidence for Cope's Rule in spite of fitness benefits of large size. *Evolution*. <http://dx.doi.org/10.5061/dryad.755d6>.
- Waller, J.T. and E.I. Svensson. (2017b). Body size evolution in an old insect order: No evidence for Cope's Rule in spite of fitness benefits of large size. *Evolution*. 71:2178-2231.
- Wellenreuther, M., Larson, K.W., and E.I. Svensson. (2012). Climatic niche divergence or conservatism? Environmental niches and range limits in ecologically similar damselflies. *Ecology*. 93:1353-1366.
- Wiens, J.J. and C.H. Graham. (2005). Niche conservatism: integrating evolution, ecology, and conservation biology. *Annu. Rev. Ecol. Evol. Syst.* 36:519-539.
- Willink, B. and E.I. Svensson. (2017). Intra-and intersexual differences in parasite resistance and female fitness tolerance in a polymorphic insect. *Proc. R. Soc. B* 284:2346-2407.
- Wilman, H., Belmaker, J., Simpson, J., de la Rosa, C., Rivadeneira, M. M., and W. Jetz (2014). EltonTraits 1.0: Species-level foraging attributes of the world's birds and mammals. *Ecology*. 95:2027-2027.

Paper III





Body size evolution in an old insect order: No evidence for Cope's Rule in spite of fitness benefits of large size

John T. Waller^{1,2} and Erik I. Svensson¹

¹Evolutionary Ecology Unit, Department of Biology, Lund University, SE-223 62 Lund, Sweden

²E-mail: john.waller@biol.lu.se

Received November 21, 2016

Accepted June 20, 2017

We integrate field data and phylogenetic comparative analyses to investigate causes of body size evolution and stasis in an old insect order: odonates ("dragonflies and damselflies"). Fossil evidence for "Cope's Rule" in odonates is weak or nonexistent since the last major extinction event 65 million years ago, yet selection studies show consistent positive selection for increased body size among adults. In particular, we find that large males in natural populations of the banded demoiselle (*Calopteryx splendens*) over several generations have consistent fitness benefits both in terms of survival and mating success. Additionally, there was no evidence for stabilizing or conflicting selection between fitness components within the adult life-stage. This lack of stabilizing selection during the adult life-stage was independently supported by a literature survey on different male and female fitness components from several odonate species. We did detect several significant body size shifts among extant taxa using comparative methods and a large new molecular phylogeny for odonates. We suggest that the lack of Cope's rule in odonates results from conflicting selection between fitness advantages of large adult size and costs of long larval development. We also discuss competing explanations for body size stasis in this insect group.

KEY WORDS: Body size, comparative methods, natural selection, Odonata, phylogenetics, sexual selection, stasis.

Body size influences many aspects of an organism's life history (Cardillo et al. 2005), and larger body size is typically associated with fitness benefits (Blanckenhorn 2000). In spite of these fitness benefits, many populations and lineages remain small over long evolutionary time scales (Blanckenhorn 2000; Kingsolver and Pfennig 2004). On paleontological time scales, Cope's Rule states that within-lineage body size tends to increase over time (Gould 1997; Alroy 1998; Gotanda et al. 2015; Bokma et al. 2016). Support for Cope's Rule comes most famously from comparative analyses of extant mammalian taxa (Baker et al. 2015). Based on the results from meta-analyses of natural and sexual selection in natural populations (Kingsolver et al. 2001), it has been suggested that Cope's Rule might largely or entirely be a result of positive selection of large size at the intraspecific level and in local populations (Kingsolver and Pfennig 2004; Kingsolver and Diamond 2011).

Some lineages, however, do not show evidence for Cope's Rule and the expected evolutionary increase in mean body size,

in spite of consistent directional selection for larger size (Gotanda et al. 2015; Rollinson and Rowe 2015). Body size evolution in some lineages is often slow, potentially being an example of evolutionary stasis (Gotanda et al. 2015; Rollinson and Rowe 2015). The problem of evolutionary stasis (or at least apparent stasis in the fossil record) has been discussed for a long time in the evolutionary biology literature, and a variety of explanations have been put forward to explain such patterns (e.g., Charlesworth et al. 1982; Merilä et al. 2001; Hansen and Houle 2004; Estes and Arnold 2007). Controversies remain whether stasis is a real phenomenon or a statistical artefact. The two main explanations for stasis are either based on some forms of genetic constraints or opposing selection pressures (Hansen and Houle 2004). A third pattern of body size evolution is punctuated change associated with cladogenesis as lineages split (Mattila and Bokma 2008). Understanding how and which of these different modes of body size evolution are operating on extant taxa and were operating on extinct taxa in the past remains a major challenge in evolutionary biology.

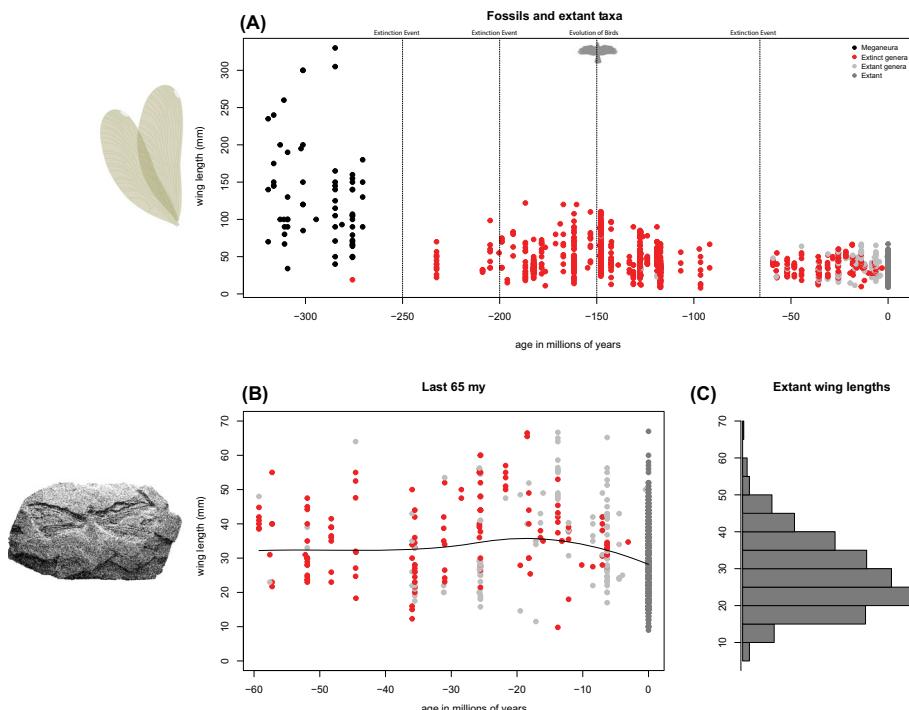


Figure 1. (A) Odonate fossil data over the last 350 million years. Wing measurements and dates were taken from Clapham and Karr (2012). The genus *Meganeura* are shown in black, extinct genera in red, extant genera in gray, and extant taxa in dark gray. Major extinction events are indicated with dotted lines. (B) Over the last 65 my, since the last major extinction event, mean wing length has remained relatively constant (indicated with a lowess curve). The down-tick seen in this graph is probably an artifact of lowess overfitting and not representative of real evolution of smaller wing length. Smaller extinct species may be underrepresented, since they might fossilize at a lower rate. (C) The histogram shows the distribution of extant wing sizes, which is slightly right-skewed, which might indicate size-dependent diversification in the presence of a lower bound (Clauset and Erwin 2008).

Hone and Benton (2005) outlined several reasons why lineages might remain small, despite selection for larger body size, and their possible explanations include increased development times and increased susceptibility to extinction of large taxa (Damuth 1981; Millar and Zammuto 1983; Damuth 1993).

The old insect order Odonata (Anisoptera: “dragonflies”; Zygoptera: “damselflies”) consists of more than 6000 extant species. This insect order was present already during the lower Permian about 250 mya, and they might have evolved even earlier (Grimaldi and Engel 2005). Odonates have complex life cycles, with an aquatic larval stage and a terrestrial adult reproductive stage, and final adult structural body size is therefore fixed as they emerge from the larval stage (Corbet 1999; Stoks and Cordoba-Aguilar 2012). Environmental conditions in the larval stage have

been shown to have strong carry-over effects with important fitness consequences for the adult stage (Corbet 1999; Stoks and Cordoba-Aguilar 2012). The aquatic larval stage can last years in some species (Corbet 1999), and it is usually considerably longer than the adult stage, which can last as little as a few days or weeks (Corbet 1999).

Crow-sized dragonfly relatives of modern odonates (Protodonata) existed between the late Carboniferous and Late Permian, but these large insects went extinct around 247 mya (Grimaldi and Engel 2005) (Figs. 1, S1). As atmospheric oxygen levels were considerably higher during this period (30–35%) compared to today (about 21%), such physiological limits have been suggested to limit body size evolution of tracheal-breathing insects (Harrison et al. 2010). This oxygen hypothesis is also partly

supported by experimental work on fruit flies (*Drosophila*), which if raised and exposed to experimentally increased oxygen levels for multiple generations evolve increased body sizes (Klok and Harrison 2009). However, it has also been argued that predation or niche filling by birds are major selection pressures that have counteracted the evolution of larger body sizes in insects during the last 150 mya (Clapham and Karr 2012).

While there have been large increases and decreases in maximum body size in the odonate fossil record within individual lineages, whole-lineage mean body size has changed very little since the end of the Cretaceous (Fig. 1; data taken from Clapham and Karr 2012). This lack of any directional trends suggests that body size evolution in odonates have largely followed an unbiased random walk process since the last mass extinction at the K-T boundary 65 mya. Such apparent stasis at the level of the entire order—as suggested by the odonate fossil record—is somewhat paradoxical in light of the common observation of positive selection for large body size in the adult stage in most odonate species studied to date (Honek 1999; Sokolovska et al. 2000). Hence, there appears to be body size stasis in this and many other animal groups, when we would rather expect to see Cope's Rule, based on the positive bias in the direction of selection on body size (cf. Kingsolver and Pfennig 2004), that is the idea that “bigger is better” (Blanckenhorn 2000; Kingsolver and Huey 2008).

In this article, we argue that slow body size evolution in odonates is likely to result from conflicting selection pressures that change sign during the life cycle, in line with some previous suggestions (Thompson and Fincke 2002; Nijhout et al. 2010). The problem of evolutionary stasis has been discussed at depth in the evolutionary literature (Merilä et al. 2001; Hansen and Houle 2004; Voje 2016). Here, we show using combination of field studies and phylogenetic comparative analyses that several of the proposed mechanisms that could generate stasis are unlikely to explain slow body size evolution in odonates. We generated a large new molecular phylogeny of odonates for this purpose and use modern comparative methods to investigate the mode and dynamics of body size evolution in this insect order. We especially evaluate five not mutually explanations for stasis: (1) stabilizing selection in the adult stage; (2) fluctuating selection in the adult stage; (3) conflicting selection in the adult stage (pleiotropic constraints between adult fitness components); (4) species-selection against large taxa, and (5) conflicting selection between the adult and larval stage.

Based on the results in our study, we suggest that conflicting selection between adult body size and costs of long development time is the most likely explanation for stasis in this insect order (cf. Schlüter et al. 1991). Our favored explanation for slow body size evolution in odonates is potentially also applicable to other insect groups and ectotherms that show no evidence of “Cope's Rule” and no evidence for apparent directional evolutionary size

trends. Conflicting selection pressures can thus be divided into selection against long larval development time in opposition against selection for large body size in the adult stage through increased fecundity, survival, and mating success (Schlüter et al. 1991). We briefly discuss the various other explanations for stasis in body size evolution that have been proposed, and the implications of our results for other organismal groups.

Materials and Methods

GENERAL PROCEDURES: FIELD WORK, SELECTION ANALYSES, LITERATURE DATA, AND PHYLOGENY RECONSTRUCTION

Unless otherwise stated, all statistical analyses in this article were performed using the R programming environment (R Development Core Team 2013). The phylogenetic tree that we use in this study was constructed from odonate DNA sequences downloaded GenBank (Fig. S2). The resulting phylogeny comprised 1322 taxa, about 21% of all known extant odonate species (about 6400 in total) (Schorr and Paulson 2015). See our Supplementary Material for a detailed description of the construction of phylogenetic trees used in this study (Figs. S2–S4), and how we accounted for phylogenetic uncertainty (Fig. S5).

Survival and sexual selection estimates come from a mark-recapture study of males in the banded demoiselle (*Calopteryx splendens*), that took place in southern Sweden from 2011–2015 (see Supplementary Material; Figs. S6, S7; Table S1). Briefly, we performed daily field work during the summer months (June and July) 2011–2015. We captured as many males as possible, marked them individually, and recorded survival (longevity) and mating success (fraction of males obtaining copulations). We also measured several morphological traits, of which we used forewing length and total body length (both measured in mm) in our analyses of natural and sexual selection in this study. We subsequently used individual variation in these two traits and fitness components (longevity and male mating success) to quantify natural and sexual selection in the field (Lande and Arnold 1983; Gosden and Svensson 2008; Stinchcombe et al. 2008). More detailed methodological information about general field work procedures, population locations, and the quantification of natural and sexual selection in this species can also be found in the Supplementary Material and in our previous publications (Svensson et al. 2004, 2006, 2014, 2016; Waller and Svensson 2016).

Body length and wing length data were extracted from the literature (mainly from national, regional and local field guides, complemented with scientific articles, See Table S2) for each odonate species in the World Odonata List (Schorr and Paulson 2015). All these trait data and ecological variables have been compiled into a database called the Odonate Phenotypic Database (opdb), which will become publicly available in the near future).

More detailed information can be found in the Supplementary Material (Fig. S8).

We collected data on larval development time in relation to latitude and body length data of European odonates from the literature (Sandhall 1987). To quantify the relationship between body size and development time (in years), we conducted a phylogenetic generalized least square (PGLS) regression using the corPGLS function in ape (Paradis et al. 2004) with a lambda parameter estimated be 0.89. Latitudinal data for these species were obtained from the distribution maps in Dijkstra and Lewington (2006). More information can be found in the Supplementary Material.

We compiled and updated literature data from a previously published quantitative review about the relationship between body size and various fitness components across several odonate species (Sokolovska et al. 2000). To evaluate the possible role of stabilizing selection on body size from these correlation estimates, we simulated the distribution of correlation coefficients that we would expect to sample under a scenario where only stabilizing selection was operating and only a correlation coefficient was estimated (See Supplementary Material for explanation).

DIVERSIFICATION IN RELATION TO BODY SIZE

We estimated the speciation rate in relation to body size of each odonate family for which we had sufficient (> 10 taxa) phylogenetic and phenotypic data ($N = 21$ families), assuming a yule model or pure speciation model (Paradis et al. 2004). A yule model was fitted for each family tree with missing species added. We used mean body length per family, calculated using the body lengths from all our species in our dataset.

The procedure described above was preferred over fitting all the data in a single model, such as using the QuaSSE procedure (FitzJohn 2010). We had a large number of missing tip taxa, and QuaSSE can only deal with randomly missing tip taxa (Fitzjohn 2010). QuaSSE is also known to have high type I error rates when the rates on the tree vary (Rabosky and Goldberg 2015). However, for completeness, we also present the results of a fitted QuaSSE model (sigmoid), although we emphasize that these results should be interpreted with the above caveats in mind (Fig. S9). We were unable to reliably estimate extinction rates, as there were many missing terminal taxa in our phylogeny (FitzJohn et al. 2009). In these analyses, we therefore assumed that the extinction rates were constant across the phylogeny.

COMPARATIVE ANALYSES OF BODY SIZE EVOLUTION

First, we fitted a simple single-optimum Ornstein-Uhlenbeck (OU) model of evolution to our body size data. Our goal here was simply to detect if there was any evolutionary constraints on interspecific body size variance accumulation in this lineage

(Butler and King 2004). This single-optimum OU-model was first compared to a Brownian Motion (BM) model where interspecific variance in body sizes is supposed to increase at a linear rate over time (Hansen 1997). A significant alpha parameter in an OU-model suggests the presence of a significant pullback force that constrains maximum or minimum body size, and is often interpreted as an indication that some constraints are operating on the lineage as a whole (Hansen et al. 2008; O'Meara and Beaulieu 2014). The fit of this simple OU-model with one optimum was subsequently compared to the fit of a BM-model using the Akaike Information Criterion (AIC).

We complemented this single-optimum OU-model by using a newly developed phylogenetic comparative method that automatically detects evolutionary shifts in OU optima for a continuously varying trait (the *lou* R-package; Khabbazian et al. 2016). We used the *estimate_shift_configuration* function in this package to document the most likely shifts in optima in our body length data on the phylogenetic tree (Khabbazian et al. 2016). Shifts in body size optima would indicate a changing adaptive landscape over evolutionary time and suggest cases of positive or negative body selection pressures toward different optima. The *lou* R-package offers several computational advantages for large trees (such as ours, with > 700 species), compared to similar methods such as surface and bayou (Ingram and Mahler 2013; Uyeda and Harmon 2014), which often take a long time to converge. For further details regarding our phylogeny and our comparative analyses, see Supplementary Material.

Results

NATURAL AND SEXUAL SELECTION ON BODY SIZE IN THE FIELD

Standardized linear univariate selection gradients (β) estimated from how forewing length of *C. splendens* males affected survival suggested that natural selection on wing length was positive in four out of the five seasons (Fig. 2). The strength of selection varied only marginally across years and did not change sign (ANCOVA; Forewing length: $F_1 = 13.8, P < 0.001$; Year: $F_4 = 0.001, P = 1.00$, Forewing length \times Year: $F_4 = 2.08, P = 0.085$). Selection on body size was positive from 2011 to 2014 (Fig. 2). The exception to this positive survival selection was observed in 2015, when natural selection on forewing length was close to zero and not significant (Fig. 2). Combining these natural selection estimates for all years showed significant and consistent directional selection toward longer forewing length, but no evidence of significant quadratic selection (Fig. 2). A similar picture emerged when we estimated survival selection on total body length, where selection gradients were positive in all five years and highly significant when years were combined ($\beta = 0.147; P < 0.001$; Fig. 2). Similar to the situation for wing length, there was no significant

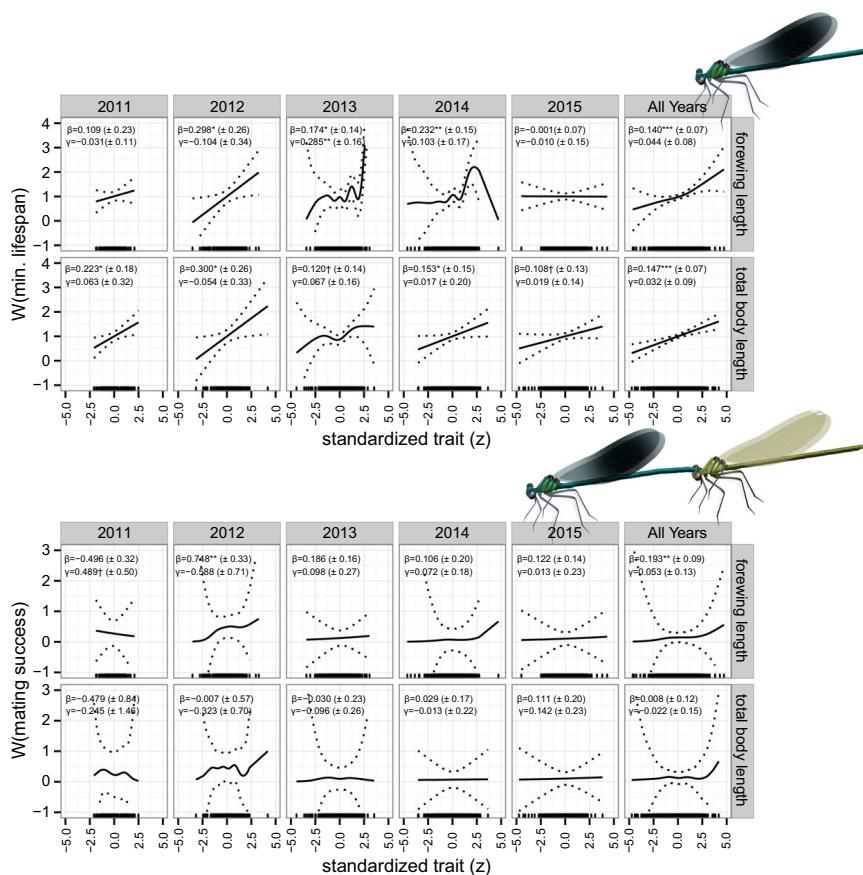


Figure 2. Natural and sexual selection in adult male banded demoiselles (*Calopteryx splendens*) from 2011 to 2015. Cubic splines were fit to each year's fitness data (Schluter 1988). Fitness functions for natural selection (minimum lifespan) are shown in the upper panel, and fitness surfaces for binary mating success (sexual selection) in the lower panel. Cubic splines were fit using the `gam` function in the package `mgcv` (Wood 2011) and smoothing with method `GCV.Cp`. R code for this figure is included in the supplement. Linear selection (β) and quadratic coefficients (γ) are listed in the top corner. Significance is marked ($^{\dagger} < 0.1$, $*$ < 0.05 , $^{**} < 0.01$, $^{***} < 0.001$). In most years, selection is positive for wing length, both in terms of survival and mating success. There is very little evidence for stabilizing selection, as often our quadratic coefficients show even slight positive curvature. See supplement for more detailed statistics and a similar analysis with a composite body size measure (PC1).

evidence for any negative curvature in the fitness function of body length, that is no evidence for stabilizing selection (γ) on body size (Fig. 2).

For sexual selection (male mating success), we found that the univariate selection gradients for forewing length were also mainly positive in four out of five years (Fig. 2). The exception to this general pattern was in 2011, when we found a negative

selection gradient for forewing length, although this was far from significant (Fig. 2). Sexual selection on body length showed a less consistent pattern, with negative directionality the first three years, and positive the last two years, although notably these differences were not significant, neither alone or when years were combined (Fig. 2). Just like in the case for natural selection (see above), there was no evidence for any significant stabilizing

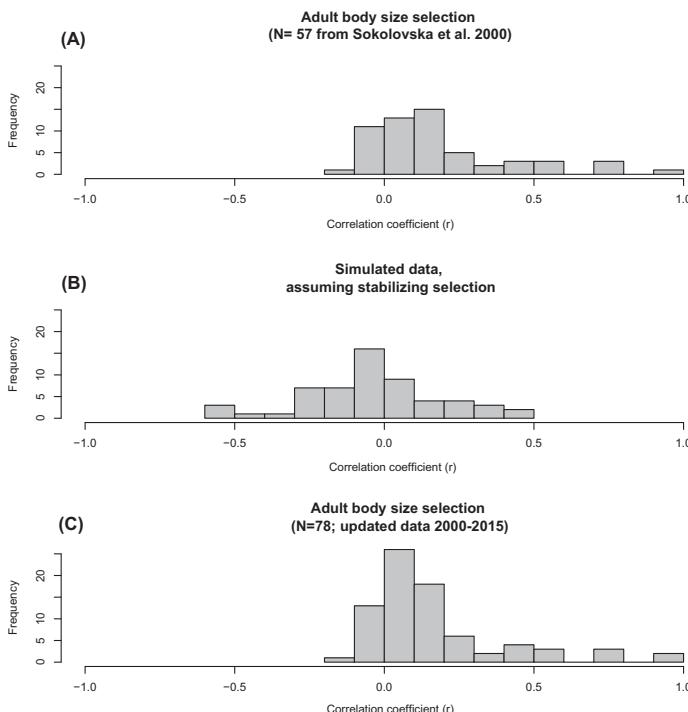


Figure 3. Correlation coefficients from studies obtained from a literature survey of body size effects on fitness in odonates (mating success, survival, and fecundity). (A) The distribution coefficients from the original review by Sokolovska et al. (2000) that contained 57 estimates. (B) The symmetrical distribution of correlation coefficients that would be expected if all populations were sitting on their adult optima (adaptive peaks), assuming a moderate level of stabilizing selection (i.e., $\beta \sim N(0, 0.1)$ and $\gamma \sim N(-0.15, 0.1)$, See R code in supplement for details). (C) Updated dataset since Sokolovska et al. (2000) ($N = 78$). Rather than a symmetric distribution around zero, the distribution is strongly right-skewed, revealing that most populations studied to date are currently not occupying their adult adaptive peaks. Instead, adults are on average smaller than their adult fitness optima.

sexual selection on either forewing length or body length (i.e., no significant negative γ). If anything, there was a small but non-significant positive curvature in our fitness function, consistent with disruptive rather than stabilizing selection (Fig. 2). Similar results were obtained when we instead estimated selection on the first principal component (PC1) of five morphological traits (fore- and hindwing length, total body length, abdomen length, thorax length), where PC1 was used as a composite measure of selection on overall size (Fig. S7).

LINKING BODY SIZE TO ADULT FITNESS

Our updated literature review on odonate body size and fitness produced 78 estimates (correlation coefficients) of the relationship between body size and various male and female fitness

components (Fig. 3). Of the 78 correlation coefficients, 33 come from independent studies, 26 from different species, and 48 of the coefficients from study sites in temperate regions. These fitness components include the relationship between male mating success, survival, and female fecundity on body size (Fig. 3). Overall, this analysis confirmed the pattern found in the original analysis from 2000 (Sokolovska et al. 2000). The distribution of these 78 correlation coefficients reported in the literature was strongly and positively skewed to the right, with 67 estimates being positive and only 11 estimates being negative. This differs significantly from the expectation if we would assume stabilizing selection in most taxa, which instead predicts that the values should be symmetrically distributed around zero (sign test: $P < 0.001$; Fig. 3B). Similarly, the original dataset reported by Sokolovska

et al. (2000) shows a significant positive skew, compared to the null expectation (48 out of 57 estimates were positive; $P < 0.001$, using a sign test). Note that these positive relationships between large body size include several different fitness components and come from both sexes (i.e., male mating success, female fecundity, and adult survival). Although not all estimates were positive, those that were negative were typically low in magnitude (Fig. 3).

Our simulated dataset (Fig. 3B) showed that if these correlation coefficients had been sampled from the same underlying fitness functions with a stable optimum (i.e., all populations in the literature would be occupying their adult optima or fluctuating around those optima), we would instead have expected to find a symmetrical pattern of correlation coefficients around zero (Fig. 3B). Similarly, disruptive selection would also have produced a symmetrical pattern (Fig. 3B). The strongly right-skewed distribution (Fig. 3 A, C) therefore suggests that stabilizing selection on body size in the adult life-stage can be rejected, or it is at least not very common. It does not rule out the possibility, however, that some populations could be experiencing stabilizing selection in the adult stage, but these are exceptions to the general rule. Stabilizing selection on adult body size is thus not common, since large body size in odonates, and many other ectotherms is almost always associated with higher fitness, irrespective of fitness components and sex (cf. Kingsolver and Huey 2008).

DEVELOPMENT TIME AND BODY SIZE

We found a significant and positive relationship between development time and body size in the European odonates ($N = 64$ species; Fig. 4). This relationship was not confounded by latitude or phylogeny, both of which were accounted for in a phylogenetic generalized least square (PGLS) analysis (Lambda: 0.89; Effect of size = 0.040, SE = ± 0.01, $P = 0.002$; effect of latitude = 0.053, SE = ± 0.01, $P = 0.0003$). Thus, after controlling for phylogeny, European odonates increased their developmental time at higher latitudes and with larger adult body size. We emphasize that at a given latitude, development time increases with body size, and the pattern we see in Fig. 4 is not simply due to confounding variation in mean range latitude.

OU-MODELS: ONE OR MULTIPLE BODY SIZE OPTIMA?

Using phenotypic data and our novel molecular phylogeny of extant odonate species, we found evidence for a significant constraining force on maximum and minimum body sizes (Fig. 5A). A single-optimum OU-model (fit to the wing length dataset) had a significantly better fit than a simple BM-model (BM: AIC = 4807.86, $\sigma^2 = 1.08$, root = 30.9, log-likelihood = -2401.93; OU: AIC = 4774.10, $\alpha = 0.008$, $\sigma^2 = 1.36$, root = 31.3, log-likelihood = -2384.05) (Fig. 5A). Note that not only were body

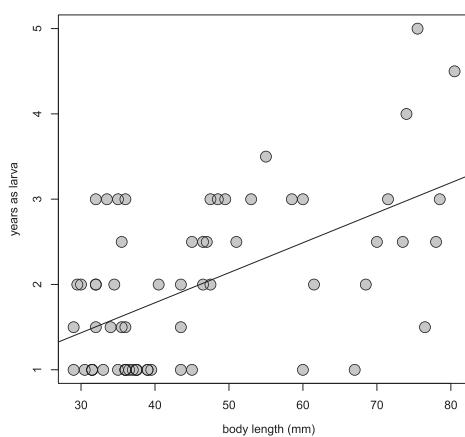


Figure 4. The positive relationship between development time and body size (body length) in the European odonates (Phylogenetic Least Square analysis (PGLS); $N = 64$; Effects of body size: size = 0.054, $P < 0.001$; effects of latitude: lat = 0.065, $P = 0.0002$). Thus, there is a significant effect of adult body size on larval development time that holds even when we control for phylogenetic nonindependence and geographical variation due to latitude. Species that have noninteger developmental times (e.g. 1.5 years) are those where the literature gives ranges in years of the development time. See Dryad supplement for access to the full dataset.

size (wing length) variance constrained across the last 65 million years since the Cretaceous (Fig. 5A), but there was no evidence for any directional change in mean wing length, either upwards or downwards (Fig. 1). Hence, these data suggest some constraining force keeping interspecific wing length variance lower than expected from a random diffusion process (Figs. 1, 5A). These results are supported by comparing wing length trends within genera with both fossil representatives and extant taxa (Fig. 5B). Of these 25 taxa, 10 (40%) show wing length declines up to the present, whereas 15 (60%) show wing length increases. Hence, evolutionary wing length increases and decreases were approximately equally common (Fig. 5B).

Even though average body length did not change across all odonate taxa, we were able to infer multiple likely shifts in body length optima on this phylogenetic tree (Fig. 6) (see Supplementary Material for comparison with simpler models presented above). Thus, body length has therefore changed significantly during evolution, both upwards and downwards in different families compared to an assumed neutral background rate of body length evolution (Fig. 6). This suggests the presence of multiple body size optima in odonates, which are likely to be related to local

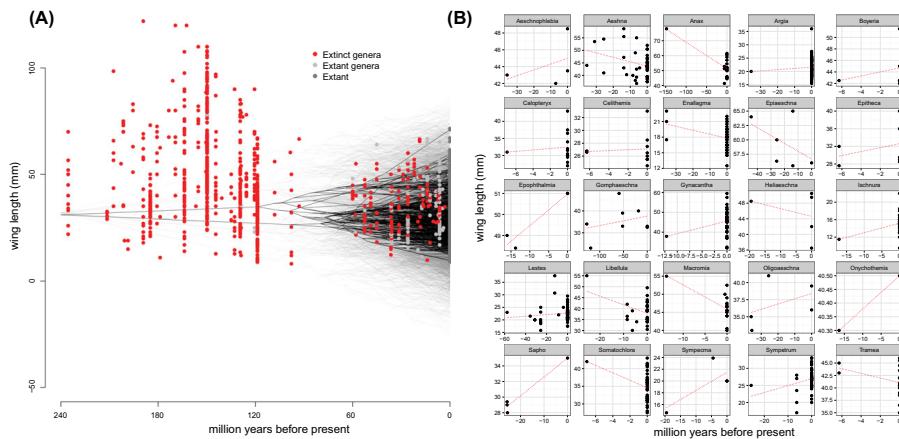


Figure 5. (A) Observed data (black) compared to expected evolutionary increase in interspecific body size variance, obtained from simulations of a Brownian Motion (BM) process (gray). The simulated pure BM-process predicts significantly more interspecific variance in wing length than we observe in this dataset. This suggests the presence of a significant constraining force that limits maximum and minimum body size in odonates. This was further confirmed by an Ohrnstein-Uhlenbeck (OU)-model that had a significantly better fit to these data than a simple BM model (BM: $AIC = 4807.86$, $\sigma^2 = 1.08$, root = 30.9, log-likelihood = -2401.93 , ntanax = 780; OU: $AIC = 4774.10$, $\alpha = 0.008$, $\sigma^2 = 1.36$, root = 31.3, log-likelihood = -2384.05 , ntanax = 780). Fossils are shown as points (red: extinct fossils, gray: fossil from extant genera, dark gray: extant taxa). A completely unconstrained model, however, still does not reach the sizes observed 150 my ago, as many fossils are larger than the extent of the unconstrained simulations. (B) Body size trends within extant odonate genera for which we also have fossil data. Here, simple linear relationships were fit to each of the extant genera for which we have fossil data. Of the 25 genera that are shown, 10 (40%) decline in body size up to the present, whereas 15 (60%) increase in body size. Hence, there is no obvious evolutionary trend for a directional bias toward either smaller or larger wing size over evolutionary time (i.e., no support for Cope's Rule).

environmental conditions, correlated with latitude, temperature, humidity, or other abiotic and biotic factors (Fig. 6).

DIVERSIFICATION IN RELATION TO BODY SIZE

Across odonate families, we found a negative, although nonsignificant relationship between speciation rate and mean family body length (linear regression: body length effect = -0.00019 , SE = ± 0.00013 , $P = 0.18$; Fig. 7). In this analysis, we did not control for phylogenetic nonindependence (Fig. 7), but the above conclusion remained when we included suborder (i.e., dragonfly or damselfly) as a random effect (linear-mixed effect model: $\beta_{bodyLength} = -0.00012$ SE = ± 0.00014 , $P = 0.40$). Dragonflies showed a positive relationship and damselflies a negative relationship between body length and speciation rate (Fig. 7). However, a formal statistical test for any such interaction revealed no significant difference between the two suborders in this relationship (linear regression: $\beta_{bodyLength} = 0.00009$, SE = ± 0.0004 , $P = 0.75$; $\beta_{Suborder} = 0.023$, SE = ± 0.020 , $P = 0.26$; $\beta_{bodyLength \times Suborder} = -0.00025$, SE = ± 0.00033 , $P = 0.46$). Finally, we note that the species distribution of wing lengths in odonates shows a slightly right-skewed pattern

(as shown in the histogram in Fig. 1C). Such right-skewed distributions have been suggested to indicate species selection against larger body size (Clauset and Erwin 2008).

We also fitted a QuaSSE-model (FitzJohn 2010) to these data. This analysis suggested a trend of decreasing speciation rate with body length (QuaSSE: $l.y0 = 0.25$, $l.y1 = 0.23$, $l.xmid = 45.62$, $l.r = 0.97$, $m.c = 0.19$, drift = 0.00, diffusion = 2.64), as indicated by a positive and significant $l.r$ term (full: log-likelihood = -22720 , $AIC = 45452$; constant_lr: log-likelihood = -22732 , $AIC = 45469$) (Fig. S9). However, we note that QuaSSE-models have high type I error (false positives) when diversification rates vary across the tree (Rabosky 2009; Rabosky and Goldberg 2015), which is especially likely to be the case in large phylogenetic trees, as we studied here.

Discussion

Here, we have combined fossil data (Figs. 1, 5B), phylogenetic comparative analyses (Figs. 4–6) and measurements of various fitness components from extant odonate species and several

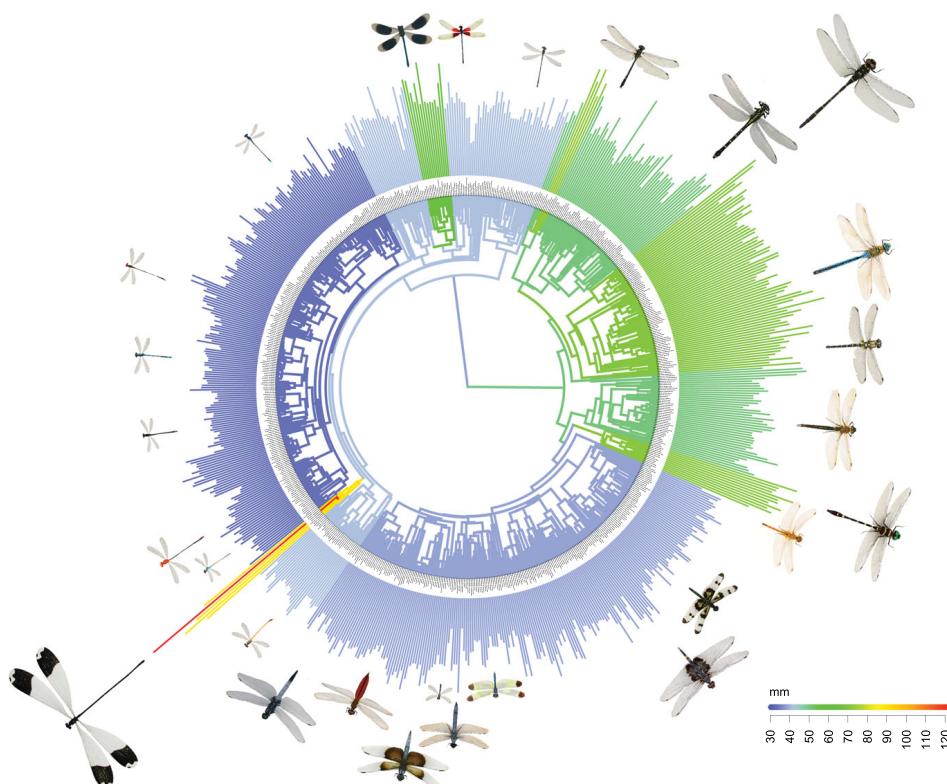


Figure 6. Body size evolution in odonates inferred from a multiple-optima Ornstein-Uhlenbeck (OU: optima shifts = 10, number of taxa = 728) model. Bars on the tips represent total body lengths. Warm colors (red, green, yellow) represent significant positive shifts from a neutral background process, whereas cold colors (blue, green, violet) represent shifts toward smaller body sizes and warmer colors (red, orange, yellow) represent shifts toward larger body sizes. These body size shifts are indicative of past positive and negative evolution toward either increased or decreased body size, revealing that evolution in this group has been completely neutral. We used the estimate_shift_configuration function of the l1ou package (Khabbazian et al. 2016) to fit our body size data to our phylogenetic tree. Photos of some selected taxa are roughly to scale (See Supplementary Material for more statistical details).

populations in the field (Figs. 2 and 3) (Weber and Agrawal 2012). Our general goal was to understand the causes of body size evolution in this ancient insect order. Our combined results suggest consistent directional selection (for these fitness components, at least) toward larger adult body size in the majority of extant populations and species studied to date (Figs. 2 and 3).

As a case in point, we found that selection on *Calopteryx splendens* male body size (as measured by both forewing length and total body length) was positively biased for both mating success and survival (Fig. 2). Although both natural and sexual selection varied marginally in magnitude between years, the long-term

trend was clearly and significantly positive over the five seasons that this study took place (Fig. 2). The estimated strengths of natural and sexual selection on body size in this study are similar in magnitude and sometimes larger than estimates of the strength of selection in natural populations from previous meta-analyses (Fig. 2; cf. Kingsolver et al. 2001).

The narrow-sense heritability (h^2) of morphological traits in natural populations (like wing length and body size) is on average 0.40–0.50 (Mousseau 1987), but see Hansen et al. 2011 for a critical discussion about the relationship between heritability and evolvability. Morphological traits typically have higher

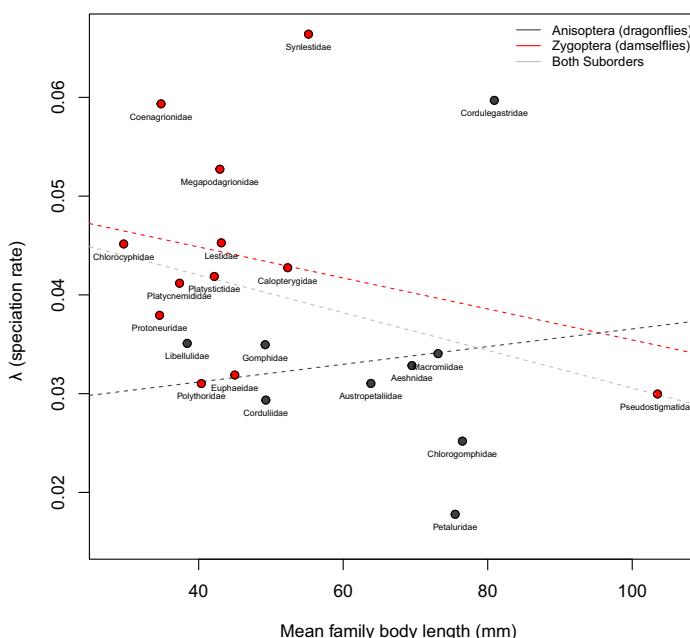


Figure 7. Speciation rate (λ) and body size (mm) for two odonate suborders: Anisoptera (dragonflies) and Zygoptera (damselflies). A pure birth model (i.e., assuming no extinction) was fit for each family. The estimate from each model was compared to mean family body length. λ decreases with increasing mean family body size, but this decrease is not significant within or across suborders, whether phylogenetic nonindependence was controlled for or not. Extinction rates cannot be estimated reliably using this phylogeny because of too many missing tip taxa. However, missing tip taxon was added to their families before estimating λ for each family.

heritabilities than physiological, behavioral, and life-history traits (Mosseau 1987). Assuming similar heritabilities for the traits included this study (or even more moderate levels), we would expect rapid increases in body size over a few thousand generations, given the strengths of natural and sexual selection we have documented in *C. splendens* (Fig. 2; cf. Lande and Arnold 1983). However, such expected rapid increases from a univariate selection model would assume that selection in the adult stage is not counteracted at the larval stage. Since adult and larval phenotypic traits are likely to be at least partly correlated with each other, there is potential for antagonism between the sign of selection on body size during the adult and the juvenile life-stages (cf. Walsh and Blows 2009).

Although conflicting selection *between* life stages can potentially be important in odonates, we have not found any strong evidence for conflicting selection *within* the adult life-stage (Fig. 1, 5). For example, we have found no evidence of any trade-off between fitness components within the adult life-stage, which would manifest itself in the form of different signs of sexual

selection and natural selection on body size (cf. Fig. 2). Instead, natural and sexual selection on male body size was typically concordant and of the same sign (Fig. 2). In other words, large males had both higher survival and higher mating success and there was no sign of any trade-off between natural and sexual selection in these data (Fig. 2).

Body size evolution could also potentially be constrained by intralocus sexual conflict (Bonduriansky and Chenoweth 2009; Cox and Calsbeek 2009), for example if males and females are selected toward different body size optima (Calsbeek and Sinervo 2004). However, we consider such intralocus sexual conflict on body size also an unlikely explanation for stasis in the case of odonates. Both male and female fitness components are typically positively correlated with body size, across most odonate species that have been studied to date (Sokolovska et al. 2000) (Fig. 3). Moreover, a long-term study of sexual selection in males and fecundity selection in females of the Coenagrionid damselfly *Ischnura elegans* found no evidence for ongoing intralocus sexual conflict on body size (Gosden and Svensson 2008). Instead,

selection gradients in the two sexes of *I. elegans* were positively correlated with each other at the local population level, suggesting sexually concordant selection rather than intralocus sexual conflict over body size (Gosden and Svensson 2008). In summary, available data in this study (Figs. 2 and 3) as well as previous studies (cited above) do not provide any empirical support for either stabilizing or conflicting selection on body size within the adult life-stage, nor any important role for intralocus sexual conflict.

Fluctuating selection pressures between generations around a stable long-term body size optimum is another potential explanation for stasis (cf. Estes and Arnold 2007). Although year-to-year variation in selection pressures could potentially result in a long-term stable adult body size optimum, such fluctuating selection is also unlikely to explain stasis on body size either for odonates in general (Fig. 3) or for *C. splendens* in particular (Fig. 2). Fluctuating selection around a long-term optimum is expected to produce a symmetrical distribution of correlation coefficients centered around zero, which contrasts strongly with the observed right-wing skew and bias toward positive selection for size (Figs. 2 and 3). In *C. splendens*, natural and sexual selection favors large body size, and adults are typically smaller than their fitness optima in most years (Fig. 2).

Therefore we can, based on these results, largely rule out conflicting selection between adult fitness components or stabilizing selection during the adult life-stage (see above). Instead, we suggest that selection for large adults is opposed by conflicting selection from predation, resource competition, or other ecological factors operating during the larval stage. Growing a larger body size will, all else being equal, require a longer developmental period (Fig. 4). A longer development period will result in a longer period of predator exposure and increased mortality risk. Such conflicting selection pressures that vary in sign between different ontogenetic stages during the life cycle have been found in studies of birds, fish, and insects (Schlüter et al. 1991; Barrett et al. 2008). In the case of odonates, we note that there is considerable scatter around the regression line of how the length of the larval period affects adult body size (Fig. 4). Thus, although the length of the larval period within a given latitudinal band and controlling for phylogeny increases with adult body size, different taxa show considerable deviations from the overall positive relationship (Fig. 4). We suggest that such deviations from the overall positive relationship reflect species-specific trade-offs and differences in larval ecology and predator avoidance behaviors. For instance, larval burrowing behaviors in some long-lived dragonfly species (Baird 2014) could reduce mortality rates from predation during the larval period and thus partly mitigate the costs of a long developmental period. Moreover, some odonate species have invaded novel habitats with either reduced overall predation risk or changed predator faunas, which could also obscure the overall

positive relationship across species. One example is *Enallagma*-damselflies that moved from lakes with primarily fish predators to lakes with primarily dragonfly predators (McPeek et al. 1996).

Stabilizing selection on pleiotropically linked traits, either between or within life-stages have been suggested as a potential explanation for stasis (Hansen and Houle 2004). Here, we argue that such pleiotropic constraints between fitness components in the adult stage are unlikely to explain stasis in odonates, although we see a potential role of such pleiotropic constraint between life-stages, in the form of conflicting selection between the adult and the larval stages. Estes and Arnold (2007) argue that pleiotropic constraints are unlikely to explain long-term stasis for two reasons: (1) Populations often respond rapidly to changing environmental conditions and (2) Pleiotropic constraints can themselves evolve and so are unlikely to account for long periods of stasis that characterize many fossil lineages. In Figures 1 and 7, we show that there has indeed been significant past evolutionary shifts in body size, implying that any pleiotropic constraints have been overcome several times in the past.

Recently, Rollinson and Rowe (2015) proposed another model to explain stasis in organismal size that is also based on conflicting selection between life-stages. Using a meta-analysis from a large dataset of selection estimates on body size, they showed that (1) juveniles are typically under strong directional selection for larger size, (2) there are typically negative genetic and/or phenotypic correlations between juvenile size and parental fecundity, suggestive of parent-offspring conflict, and (3) directional selection on adult size is often weak or close to zero (Rollinson and Rowe 2015). While their model seemingly might contradict our explanation for stasis, we note that their dataset only included vertebrates and plants, but no insects (Rollinson and Rowe 2015). In contrast to their study and conclusions, here we have shown that adult body size is typically under strong positive directional natural and sexual selection (Figs. 2 and 3). Fitness benefits of large size is likely to be similar in many other insect species (Bonduriansky 2001). Moreover, organisms with complex life-cycles typically experience radically different juvenile and adult environments, especially semi-aquatic groups like odonates that change from aquatic to a terrestrial environments after emergence. Given the conflicting selection between fitness benefits of large adult size and the costs of a long development time (Fig. 4), we suggest that our explanation of stasis might be more generally applicable to organisms with complex life-cycles, than the parent-offspring conflict model favored by Rollinson and Rowe (2015).

Temperature (latitude and altitude), food resources, and seasonality have all been experimentally demonstrated to affect larval growth rate and final adult body sizes in odonates (Nylin and Gotthard 1998; McPeek 2004; Van Doorslaer and Stoks 2005; De Block et al. 2008; De Block and Stoks 2008; Örtman et al. 2012). Larger European odonate species spend more time in the

larval period than smaller species, and this positive relationship between adult body size and development time holds even after controlling for phylogenetic nonindependence and latitudinal position (Johansson 2003; Sniegula et al. 2016) (Fig. 4). In Fig. 4, we note some small species taking several years to develop (*Coenagrion armatum*, *Leucorrhinia albifrons*), and some rather large odonates developing quickly within a year (*Anax imperator*, *Aeshna mixta*; Sandhall 1987), but this variation is largely driven by differences in mean latitude, with r-squared increasing from 0.29 (without incorporating latitude) to 0.44 (when incorporating latitude). Body sizes in other insects has been shown to increase at an exponential rate during ontogeny, meaning that even a small increase in development time should lead to much larger size (Nijhout et al. 2006), with as much as 90 percent of the body size increase being gained in the last larval stage (Nijhout et al. 2010).

The odonate fossil data indicate relative stasis in mean body size at the order level over the last 65 million years (Fig. 5A, B). For continuous characters, a simple model of evolution is an unbounded random walk process (Brownian motion; BM) (Hansen 1997). BM-models assume that there are no evolutionary constraints on the upper and lower bounds of phenotypic traits (Hansen 1997). Evolutionary constraints can however be quantified and detected by adding an extra term (α), which measures a pullback force as traits move further away from an inferred optimum (Hansen 1997; Butler and King 2004) (Fig. 5B). The existence of such a pullback force has often been interpreted as a signature of long-term stabilizing selection (O'Meara and Beaulieu 2014).

Using data from our time-calibrated odonate phylogeny from extant odonate taxa and body size data, we detected a significant pullback force constraining the accumulation of interspecific body size variance (Fig. 5A), with maximum body size among extant odonate taxa being 10–15 mm smaller than predicted by Brownian motion (Fig. 5A). This suggests the existence of either selective or genetic constraints on maximum and minimum body size in extant odonates (Fig. 5A). We note, however, that the approach to an adaptive optimum in quantitative genetic models of phenotypic evolution is almost instantaneous on a geological time scale, whereas OU models typically capture more long-term constraining forces that are integrated over millions of years (Hansen 1997; O'Meara and Beaulieu 2014). Moreover, OU models cannot, in themselves, distinguish between “true” stabilizing selection within a life-stage, which has been suggested as an explanation for long-term stasis (Charlesworth et al. 1982) from conflicting selection between life stages (Schlüter et al. 1991). As we have argued above, conflicting selection between life stages is a more likely explanation for relative stasis in this insect order than is stabilizing selection within the adult life-stage (Figs. 1–5). To distinguish true stabilizing selection from conflicting selection between life stages, phylogenetic comparative methods are

clearly not sufficient but should ideally be complemented with experimental and mechanistic studies of carefully chosen species (Weber and Agrawal 2012).

Body size optima could shift, however, particularly in conjunction with speciation (Gould 1988). Using a multiple-optima OU-model we were able to locate several significant shifts in body size optima among extant odonate lineages (Fig. 6). These significant shifts imply past selection toward smaller or larger body size in some lineages. These comparative results from extant taxa are also consistent with our analyses on the dynamics of body size evolution among fossil data (Fig. 5B). The putative ecological causes driving these shifts will be discussed in more detail in a forthcoming article. These shifts suggest either that selection for large body size among adults is not always positive (*contra* Figs. 2 and 3), or that selection pressures during the larval stage have in some cases opposed the fitness benefits of large adult size. Future studies of both larval and adult fitness costs and benefits of size and development times, particularly in tropical odonates, would be needed to quantify the net selection pressures in different lineages. We emphasize that these different optima are unlikely to represent optimal body sizes for single species or populations. Rather, these optima should be interpreted as broad adaptive zones that cannot be well explained by a single-mean multivariate Gaussian distribution or an OU-model assuming a single optimum shared by multiple species (cf. Fig. 5). Adding additional taxa to our incomplete phylogeny is likely to reveal additional optima shifts in the future.

Several ecological factors might put an upper limit of body size in odonates. First, because the demand for oxygen in flying insects is likely large, atmospheric oxygen has been suggested as a major factor controlling maximum body size in insects (Harrison et al. 2010). The peak atmospheric oxygen in historical reconstructions corresponds well to the peak seen in giant insects 300 mya (Clapham and Karr 2012). High atmospheric oxygen levels during the Carboniferous have been suggested to be important in the energetically intensive process of flight (Harrison et al. 2010). However, body size was subsequently decoupled from atmospheric oxygen levels after the initial oxygen peak (Clapham and Karr 2012), although other oxygen-record reconstructions might show a different pattern (Harrison et al. 2010). Another selective agent that has been suggested to operate against large size is size-dependent predation by flying predators such as birds, which evolved during the last 150 million years (Clapham and Karr 2012). Interestingly, two independent studies of an extant populations of Calopterygid odonate species (*Calopteryx splendens* and *C. virgo*) have shown that individuals with longer wings are more likely to be predated by birds (Svensson and Friberg 2007; Kuchta and Svensson 2014). Previously, odonates and other large insects may also have occupied niches that are now occupied by vertebrates such as birds and bats (Harrison et al. 2010) (Fig. S1).

Body size distributions within taxonomic groups are often asymmetrical, with many more smaller species than larger ones (Clauset and Erwin 2008; Rainford et al. 2016). Such skewed body size distributions can indicate size-dependent differences in diversification rates (Clauset and Erwin 2008) or simply increases in variance in the presence of a lower bound. The distribution of odonate wing length data shows such a typical right-skewed distribution (as shown in the histogram in Fig. 1C), as one would expect if species selection was operating against larger body size in odonates. Species selection is gaining increased interest as potential macroevolutionary force (Jablonski 2008; Rabosky and McCune 2010; Chevin 2016). Species selection has also been suggested as an explanation for the lack of phyletic size increases in many organismal groups, despite consistent positive directional selection and fitness benefits of being large (Kingsolver and Pfennig 2007). Species selection against large size (e.g., increased extinction rates of large taxa) could potentially counteract any evolutionary trend in body size that would be expected if individual selection was operating alone (Kingsolver and Pfennig 2007). A tension between species-level selection causing extinction of large taxa that is counteracted by within-species selection for large size could arise if large size is important in intraspecific competitive interactions. Such a tension could result in oscillating and frequency-dependent macroevolutionary dynamics of body size (Smith and Brown 1986; see also Frank and Slatkin 1992).

Differential rates of extinction and speciation might prevent niche filling of large lineages (Maurer et al. 1992), and could therefore counteract evolutionary increases in body size. In insects, Rainford et al. (2016) found little evidence for body size dependent diversification rates. In the case of odonates, we found no statistically significant support for a negative effect of body size on speciation rates in odonates (Fig. 7). Neither our study nor Rainford et al. (2016), however, estimated extinction rates. In our case, this was due to many missing tip taxa, which are needed to reliably estimate extinction rates (Nee et al. 1994; FitzJohn et al. 2009). Body size could of course still affect diversification rates through extinction (since diversification rate = extinction rate – speciation rate). It seems unlikely that differences in diversification rate alone, however, would be enough to balance the positive directional selection for large adults (Figs. 2 and 3) and lead to complete stasis in body size, although we cannot reject such a scenario completely based on our current data. We also note that speciation rates and extinction rates might be positively correlated with each other, resulting in different lineages having differential turnover rates (Rosenblum et al. 2012; Schlüter 2016). This might imply that estimated speciation rates could partly also reflect extinction rates. Explicit evolutionary models of the interactions between individual and species level selection are still few (for a recent exception see Chevin 2016).

Conclusion

Despite consistent positive selection in the adult stage (Figs. 2 and 3; Figs. S6 and S7), odonate body size evolution appears to be constrained at both the micro- and macroevolutionary level. Since the last mass extinction event 65 mya, body size evolution in odonates has not been positively biased. Hence, there is no evidence for Cope's Rule in this insect order. Rather multiple body size shifts have been detected in both upwards and downward directions (Fig. 6), and these shifts were presumably driven by local ecological conditions.

We have suggested that the positive selection on adult body size in odonates is counteracted by conflicting selection at the larval stage (Fig. 4). We stress that other explanations for the apparent stasis in this insect order are not mutually exclusive to our favored explanation and might still be plausible. Such explanations include the possibility of species selection favoring small body size. However, we argue that pleiotropic fitness costs of large body size that are paid in the form of a lengthened larval period are likely to at least partly counteract the pervasive positive selection for large size that we have documented for adult fitness components (Figs. 2 and 3). We are fully aware that conflicting selection due to such pleiotropic costs in the form of a longer juvenile period are unlikely to result in stasis in an absolute sense, except in the unlikely case where such conflicting selection pressures are completely equal in magnitude and opposite in sign (Hansen and Houle 2004). However, we suggest that conflicting selection is one of the most likely explanation as to why positive selection in the adult stage does not lead to the rapid increases body size that would otherwise be expected.

The data and analyses in this study are certainly not incompatible with other models for stasis, including population structure, coevolutionary interactions, and gene flow between contemporary populations (Eldredge et al. 2005), niche conservatism (Wiens and Graham 2005), parent-offspring conflict (Rollinson and Rowe 2015) or selection on the environmental, rather than the heritable component of phenotypic traits (Price et al. 1988). However, our results contradict some previously suggested explanations of stasis that have been widely discussed. For instance, we find no support for stabilizing selection within life-stages (see Charlesworth et al. 1982). Neither do we find any support for trade-offs between sexual and natural selection (Figs. 2 and 3). Finally, there is no evidence that intralocus sexual conflict (Bonduriansky and Chenoweth 2009; Cox and Calsbeek 2009) would constrain males and/or females from attaining their body size optima, as both sexes seem to benefit from large body size in odonates (Fig. 3). Thus, while our study cannot prove that conflicting selection is the sole explanation for stasis, our data are incompatible with several explanations for stasis that are based on

stabilizing, fluctuating, or conflicting selection within the adult life-stage.

AUTHOR CONTRIBUTIONS

Both authors contributed to the analysis, collection, planning, and writing of the manuscript.

ACKNOWLEDGMENTS

We are grateful to Nathalie Feiner, Nicolas Chazot, and five anonymous referees for constructive criticisms earlier drafts of this manuscript. We also wish to thank A. Nordén, S. Engel, K. Boorman, M. Gomez, and a number of other temporary students, field assistants, and coworkers who helped out during one or several field seasons of this study. Funding for this study has been provided by the Swedish Research Council (VR; grant no. 621-2012-3768) and Gyllenstiernska Kräpperupstiftelsen (grant no. 2014-0032) to E.I. S.

DATA ARCHIVING

Data are archived on <https://doi.org/datadryad.org/> and <https://doi.org/treebase.org/treebase-web/home>.

LITERATURE CITED

- Alroy, J. 1998. Cope's Rule and the dynamics of body mass evolution in North American fossil mammals. *Science* 280:731–734.
- Baird, I. R. C. 2014. Larval burrow morphology and groundwater dependence in a mire-dwelling dragonfly, *Petalura gigantea* (Odonata: Petaluridae). *Int. J. Odonatol.* 17:101–121.
- Baker, J., A. Meade, M. Pagel, and C. Venditti. 2015. Adaptive evolution toward larger size in mammals. *Proc. Natl. Acad. Sci. USA* 112:5093–5098.
- Barrett, R. D. H., S. M. Rogers, and D. Schlüter. 2008. Natural selection on a major armor gene in threespine stickleback. *Science* 322:255–257.
- Blanckenhorn, W. U. 2000. The evolution of body size: what keeps organisms small? *Q. R. Biol.* 75:385–407.
- Bokma, F., M. Godinot, O. Maridet, S. Ladeuze, L. Costeur, F. Sole, E. Gheerbrant, S. Peigne, F. Jacques, and M. Laurin. 2016. Testing for Deperet's Rule (body size increase) in mammals using combined extinct and extant data. *Syst. Biol.* 65:98–108.
- Bonduriansky, R. 2001. The evolution of mate choice in insects: a synthesis of ideas and evidence. *Biol. Rev.* 76:305–309.
- Bonduriansky, R., and S. F. Chenoweth. 2009. Intralocus sexual conflict. *Trends Ecol. Evol.* 24:280–288.
- Butler, M. A., and A. A. King. 2004. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *Am. Nat.* 164:683–695.
- Calsbeek, R., and B. Sinervo. 2004. Within-clutch variation in offspring sex determined by differences in sire body size: cryptic mate choice in the wild. *J. Evol. Biol.* 17:464–470.
- Cardillo, M., G. M. Mace, K. E. Jones, J. Bielby, O. R. P. Bininda-Emonds, W. Sechrest, C. D. L. Orme, and A. Purvis. 2005. Multiple causes of high extinction risk in large mammal species. *Science* 309:1239–1241.
- Charlesworth, B., R. Lande, and M. Slatkin. 1982. A Neo-Darwinian commentary on macroevolution. *Evolution* 36:474–498.
- Chevin, L. M. 2016. Species selection and random drift in macroevolution. *Evolution* 70:513–525.
- Clapham, M. E., and J. A. Karr. 2012. Environmental and biotic controls on the evolutionary history of insect body size. *Proc. Natl. Acad. Sci. USA* 109:10927–10930.
- Clauset, A., and D. H. Erwin. 2008. The evolution and distribution of species body size. *Science* 321:399–401.
- Corbet, P. S. 1999. Dragonflies: behavior and ecology of Odonata. Harley Books, Colchester, UK.
- Cox, R. M., and R. Calsbeek. 2009. Sexually antagonistic selection, sexual dimorphism, and the resolution of intralocus sexual conflict. *Am. Nat.* 173:176–187.
- Damuth, J. 1981. Population density and body size in mammals. *Nature* 290:699–700.
- . 1993. Cope's Rule, the Island Rule and the scaling of mammalian population density. *Nature* 365:748–750.
- De Block, M., S. Slos, F. Johansson, and R. Stoks. 2008. Integrating life history and physiology to understand latitudinal size variation in a damselfly. *Ecography* 31:115–123.
- De Block, M., and R. Stoks. 2008. Compensatory growth and oxidative stress in a damselfly. *Proc. Biol. Sci.* 275:781–785.
- Dijkstra K. D. B., and R. Lewington. 2006. Field guide to the dragonflies of Britain and Europe: including western Turkey and north-western Africa. British Wildlife Publishing, Dorset, England.
- Eldredge, N., J. N. Thompson, P. M. Brakefield, S. Gavrilets, D. Jablonski, J. B. C. Jackson, R. E. Lenski, B. S. Lieberman, M. A. McPeek, and W. Miller. 2005. The dynamics of evolutionary stasis. *Paleobiology* 31:133–145.
- Estes, S., and S. J. Arnold. 2007. Resolving the paradox of stasis: models with stabilizing selection explain evolutionary divergence on all timescales. *Am. Nat.* 169:227–244.
- FitzJohn, R. G., W. P. Maddison, and S. P. Otto. 2009. Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Syst. Biol.* 58:595–611.
- FitzJohn, R. G. 2010. Quantitative traits and diversification. *Syst. Biol.* 59:619–633.
- Frank S. A., and M. Slatkin. 1992. Fisher's fundamental theorem of natural selection. *Trends Ecol. Evol.* 7:92–95.
- Gosden, T. P., and E. I. Svensson. 2008. Spatial and temporal dynamics in a sexual selection mosaic. *Evolution* 62:845–856.
- Gotanda, K. M., C. Correa, M. M. Turcotte, G. Rolshaugen, and A. P. Hendry. 2015. Linking macrotrends and microtrends: re-evaluating microevolutionary support for Cope's rule. *Evolution* 69:1345–1354.
- Gould, S. J. 1988. Trends as changes in variance—a new slant on progress and directionalism in evolution. *J. Paleontol.* 62:319–329.
- . 1997. Cope's rule as psychological artefact. *Nature* 385:199–200.
- Grimaldi, D., and M. S. Engel. 2005. Evolution of the insects. Cambridge Univ. Press, New York.
- Hansen, T. F. 1997. Stabilizing selection and the comparative analysis of adaptation. *Evolution* 51:1341–1351.
- Hansen, T. F., C. Pélabon, and D. Houle. 2011. Heritability is not evolvability. *Evol. Biol.* 38.3:258–277.
- Hansen, T. F., J. Pienaar, and S. H. Orzack. 2008. A comparative method for studying adaptation to a randomly evolving environment. *Evolution* 62:1965–1977.
- Hansen, T. F., and D. Houle. 2004. Evolvability, stabilizing selection, and the problem of stasis. Pp. 130–150 in M. Pigliucci and K. Preston, eds. *Phenotypic integration*. Oxford Univ. Press, Oxford.
- Harrison, J. F., A. Kaiser, and J. M. VandenBrooks. 2010. Atmospheric oxygen level and the evolution of insect body size. *Proc. Biol. Sci.* 277:1937–1946.
- Hone, D. W. E., and M. J. Benton. 2005. The evolution of large size: how does Cope's Rule work? *Trends Ecol. Evol.* 20:4–6.
- Honek, A. 1999. Constraints on thermal requirements for insect development. *Entomol. Sci.* 4:615–621.
- Ingram, T., and D. L. Mahler. 2013. SURFACE: detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike Information Criterion. *Methods Ecol. Evol.* 4:416–425.

- Jablonski, D. 2008. Species selection: theory and data. *Ann. Rev. Ecol. Evol.* 39:501–524.
- Johansson, F. 2003. Latitudinal shifts in body size of *Enallagma cyathigerum* (Odonata). *J. Biog.* 30:29–34.
- Khabbazian, M., R. Kriebel, K. Rohe, and C. Ane. 2016. Fast and accurate detection of evolutionary shifts in Ornstein-Uhlenbeck models. *Methods Ecol. Evol.* 7:811–824.
- Kingsolver, J. G., H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri, C. E. Hill, A. Hoang, P. Gibert, and P. Beerli. 2001. The strength of phenotypic selection in natural populations. *Am. Nat.* 157:245–261.
- Kingsolver, J. G., and D. W. Pfennig. 2004. Individual-level selection as a cause of Cope's rule of phyletic size increase. *Evolution* 58:1608–1612.
- . 2007. Patterns and power of phenotypic selection in nature. *BioScience* 57:561–572.
- Kingsolver, J. G., and R. B. Huey. 2008. Size, temperature and fitness: three rules. *Evol. Ecol. Res.* 10:251–268.
- Kingsolver, J. G., and S. E. Diamond. 2011. Phenotypic selection in natural populations: what limits directional selection? *Am. Nat.* 177:346–357.
- Klok, C. J., and J. F. Harrison. 2009. Atmospheric hypoxia limits selection for large body size in insects. *PLoS One* 4:e3876.
- Kuchta, S. R., and E. I. Svensson. 2014. Predator-mediated natural selection on the wings of the damselfly *Calopteryx splendens*: differences in selection among trait types. *Am. Nat.* 184:91–109.
- Lande, R., and S. J. Arnold. 1983. The measurement of selection on correlated characters. *Evolution* 37:1210–1226.
- Mattila, T. M., and Bokma, F. 2008. Extant mammal body masses suggest punctuated equilibrium. *Proc. Biol. Sci.* 275:2195–2199.
- Maurer, B. A., J. H. Brown, and R. D. Rusler. 1992. The micro and macro in body size evolution. *Evolution* 46:939–953.
- Merilä, J., B. C. Sheldon, and L. E. B. Kruuk. 2001. Explaining stasis: microevolutionary studies in natural populations. *Genetica* 112:199–222.
- McPeek, M. A. 2004. The growth/predation risk trade-off: so what is the mechanism? *Am. Nat.* 163:E88–E111.
- McPeek, M. A., A. K. Schrot, and J. M. Brown. 1996. Adaptation to predators in a new community: swimming performance and predator avoidance in damselflies. *Ecology* 77:617–629.
- Millar, J. S., and R. M. Zammuto. 1983. Life histories of mammals—an analysis of life-tables. *Ecology* 64:631–635.
- Mousseau, T. A. 1987. Natural selection and the heritability of fitness components. *Heredity* 59:181–197.
- Nee, S., E. C. Holmes, R. M. May, and P. H. Harvey. 1994. Extinction rates can be estimated from molecular phylogenies. *Phil. Trans. R Soc. B.* 344:77–82.
- Nijhout, H. F., G. Davidowitz, and D. A. Roff. 2006. A quantitative analysis of the mechanism that controls body size in *Manduca sexta*. *J. Exp. Biol.* 5:16. <https://doi.org/10.1186/jbiol43>.
- Nijhout, H. F., D. A. Roff, and G. Davidowitz. 2010. Conflicting processes in the evolution of body size and development time. *Phil. Trans. R Soc. B.* 365:567–575.
- Nylin, S., and K. Gotthard. 1998. Plasticity in life-history traits. *Ann. Rev. Entomol.* 43:63–83.
- O'Meara, B. C., and J. M. Beaulieu. 2014. Modelling stabilizing selection: the attraction of Ornstein-Uhlenbeck models. Pp. 381–393 in L. Zsolt, ed. *Modern phylogenetic comparative methods and their application in evolutionary biology*. Springer, Berlin Heidelberg.
- Örtman, V. N., R. Stoks, M. De Block, and F. Johansson. 2012. Generalists and specialists along a latitudinal transect: patterns of thermal adaptation in six species of damselflies. *Ecology* 93:1340–1352.
- Paradis, E., J. Claude, and K. Strimmer. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289–290.
- Price, T., M. Kirkpatrick, and S. J. Arnold. 1988. Directional selection and the evolution of breeding date in birds. *Science* 240:798–799.
- R Development Core Team. 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing. <http://www.R-project.org>, Vienna, Austria.
- Rabosky, D. L. 2009. Ecological limits and diversification rate: alternative paradigms to explain the variation in species richness among clades and regions. *Ecol. Lett.* 12:735–743.
- Rabosky, D. L., and A. R. Mccune. 2010. Reinventing species selection with molecular phylogenies. *Trends Ecol. Evol.* 25:68–74.
- Rabosky, D. L., and E. E. Goldberg. 2015. Model inadequacy and mistaken inferences of trait-dependent speciation. *Syst. Biol.* 64:340–355.
- Rainford, J. L., M. Hofreiter, and P. J. Mayhew. 2016. Phylogenetic analyses suggest that diversification and body size evolution are independent in insects. *BMC Evol. Biol.* 16. <https://doi.org/10.1186/s12862-015-0570-3>.
- Rollinson, L., and Rowe, L. 2015. Persistent directional selection on body size and a resolution to the paradox of stasis. *Evolution* 69:2441–2451.
- Rosenblum, E. B., B. A. J. Sarver, J. W. Brown, S. D. Roches, K. M. Hardwick, T. D. Hether, J. M. Eastman, M. W. Pennell, and L. J. Harmon. 2012. Goldilocks meets Santa Rosalia: an ephemeral speciation model explains patterns of diversification across time scales. *Evol. Biol.* 39:255–261.
- Sandhall, A. 1987. *Trollsländor i Europa*. Stenströms Bokförlag (in Swedish).
- Schlüter, D. 1988. Estimating the form of natural selection on a quantitative trait. *Evolution* 5:849–861.
- . 2016. Speciation, ecological opportunity, and latitude. *Am. Nat.* 187:1–18.
- Schlüter, D., T. D. Price, and L. Rowe. 1991. Conflicting selection pressures and life-history trade-offs. *Proc. Biol. Sci.* 246:11–17.
- Schorr, M., and D. Paulson. 2015. World Odonata List. Electronic Database accessible at <http://www.pugetsound.edu/academics/academic-resources/slater-museum/biodiversity-resources/dragon-flies/world-odonata-list2/>. Accessed 11 (2015).
- Smith, J. M., and R. L. Brown. 1986. Competition and body size. *Theor. Pop. Biol.* 30:166–179.
- Sniegula, S., M. J. Golab, S. M. Drobniak, and F. Johansson. 2016. Seasonal time constraints reduce genetic variation in life-history traits along a latitudinal gradient. *J. Anim. Ecol.* 85:187–198.
- Sokolovska, N., L. Rowe, and F. Johansson. 2000. Fitness and body size in mature odonates. *Ecol. Entomol.* 25:239–248.
- Stinchcombe, J. R., A. F. Agrawal, P. A. Hohenlohe, S. J. Arnold, and M. W. Blows. 2008. Estimating nonlinear selection gradients using quadratic regression coefficients: double or nothing? *Evolution* 62:2435–2440.
- Stoks, R., and A. Cordoba-Aguilar. 2012. Evolutionary ecology of Odonata: a complex life cycle perspective. *Ann. Rev. Entomol.* 57:249–265.
- Svensson, E. I., and M. Friberg. 2007. Selective predation on wing morphology in sympatric damselflies. *Am. Nat.* 170:101–112.
- Svensson, E. I., L. Kristoffersen, K. Oskarsson, and S. Bensch. 2004. Molecular population divergence and sexual selection on morphology in the banded demoiselle (*Calopteryx splendens*). *Heredity* 93:423–433.
- Svensson, E. I., F. Erokuhamanoff, and M. Friberg. 2006. Effects of natural and sexual selection on adaptive population divergence and premating isolation in a damselfly. *Evolution* 60:1242–1253.
- Svensson, E. I., A. Runemark, M. N. Verzijden, and M. Wellenreuther. 2014. Sex differences in developmental plasticity and canalization shape population divergence in mate preferences. *Proc. Biol. Sci.* 281. <https://doi.org/10.1098/rspb.2014.1636>.
- Svensson, E. I., A. Norden, J. T. Waller, and A. Runemark. 2016. Linking intra- and interspecific assortative mating: consequences for asymmetric sexual isolation. *Evolution* 70:1165–1179.

- Thompson, D. J., and O. M. Fincke. 2002. Body size and fitness in Odonata, stabilising selection and a meta-analysis too far? *Ecol. Entomol.* 27:378–384.
- Uyeda, J. C., and L. J. Harmon. 2014. A novel Bayesian method for inferring and interpreting the dynamics of adaptive landscapes from phylogenetic comparative data. *Syst. Biol.* 63:902–918.
- Van Doorslaer, W., and R. Stoks. 2005. Growth rate plasticity to temperature in two damselfly species differing in latitude: contributions of behaviour and physiology. *Oikos* 111:599–605.
- Voje, K. L. 2016. Tempo does not correlate with mode in the fossil record. *Evolution* 70:10.1111/evo.13090
- Waller, J., and E. I. Svensson. 2016. The measurement of selection when detection is imperfect: how good are naive methods? *Methods Ecol. Evol.* 7:538–548.
- Walsh, B., and M. W. Blows. 2009. Abundant genetic variation plus strong selection = multivariate genetic constraints: a geometric view of adaptation. *Annu. Rev. Ecol. Evol. System* 40:41–59.
- Weber, M. G., and A. A. Agrawal. 2012. Phylogeny, ecology, and the coupling of comparative and experimental approaches. *Trends Ecol. Evol.* 27:394–403.
- Wiens, J. J., and C. H. Graham. 2005. Niche conservatism: integrating evolution, ecology, and conservation biology. *Ann. Rev. Ecol. Evol. Syst.* 36:519–539.

Associate Editor: M. Carling
Handling Editor: P. Tiffin

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Table S1. Mark-recapture (MR) analysis of our survival selection data, using male forewing length as a trait.

Table S2. Field guides and other sources of body size (wing lengths and body lengths) data from extant taxa.

Figure S1. Atmospheric oxygen (GEOCARBSULF model) and min and max wing length over the last 350 my.

Figure S2. Current state of odonate sequences on GenBank. Ordered plot of all downloaded sequences from GenBank for use in the construction of our phylogenetic tree.

Figure S3. Posterior node support output from BEAST for the suborder Anisoptera (dragonflies).

Figure S4. Posterior node support output from BEAST for the suborder Zygoptera (damselflies).

Figure S5. Effect of phylogenetic uncertainty for the comparative tests run in this study.

Figure S6. Survival and sexual selection (on wing length) in adult *Calopteryx splendens* males from 2011–2015.

Figure S7. Selection gradients using the first principal component (PC1) rather than forewing length as is used in the main text Figure 1.

Figure S8. The strong correlation ($r = 0.90$; $P < 0.001$) between wing length and body length in odonates (males shown). Each point represents measurements from one species.

Figure S9. Fitted QuaSSE-model (FitzJohn 2010) values for the dependence of speciation rate and male body length (mm).

Supporting Material

Paper IV



Is the blunderbuss a misleading visual metaphor for stasis and punctuated evolution?

Author

John T. Waller

Affiliation

Evolutionary Ecology Unit, Department of Biology, Lund University, SE-223 62 Lund, Sweden

Abstract

I discuss the usefulness of the so-called “blunderbuss pattern” of phenotypic evolution as a visual metaphor for stasis and punctuated evolution that was originally put forward in Uyeda et al. (2011) in their highly influential paper “The million-year wait for macroevolutionary bursts”. I argue the blunderbuss pattern is not surprising, and in some cases it is misleading. I review several publications that cite Uyeda et al. (2011) that seem to be confused about the meaning of the pattern and what it implies. I do not critique the original analysis within Uyeda et al (2011), but show the blunderbuss pattern itself would be produced even when assuming a Brownian motion (completely gradual) model of phenotypic divergence. Finally, I discuss how the interesting results of the paper have been overlooked in favor of the surprisingly powerful, but also misleading visual metaphor of the blunderbuss.

Introduction

In a highly influential and impressive macroevolutionary study, Uyeda et al. (2011) presented compelling evidence that body size evolution occurs in rare but substantial bursts, with bounded evolution on shorter time scales (Uyeda et al. 2011; their Table 1). These authors fitted several models to various sources of body size divergence data, compiled from fossil evidence, phylogenetic comparative methods, and studies of contemporary evolutionary rates. They found that the best-performing model was the multi-burst model of evolution. Furthermore, the parameter estimates from this model indicated that there is at least a 1 MYA wait for these bursts of body-size evolution. These significant and exciting results were presented in conjunction with a powerful visual metaphor that has also became highly influential: “the blunderbuss” (Uyeda et al. 2011, Fig. 1).

Uyeda et al. (2011) also tested an appropriate null model, Brownian Motion (BM), which makes very different predictions from their best performing model (the multi-burst). Under the BM-model, evolution is unbounded and occurs gradually (i.e. no paradoxical 1 MYA waits for bursts of evolution). Instead, under such a BM-model, phenotypic evolution accumulates continuously and gradually (Felsenstein 1985; Hansen 1997). These two models tell very different stories about the how body size diverges over time.

Here, I present the results from a simulation of divergence under a BM-model (Fig. 1). Interestingly, and somewhat problematically, a BM-model will also produce a blunderbuss pattern when plotted on a log scale. Thus, even though the BM-model and the multi-burst model (constrained evolution below a million-year bound) are qualitatively and quantitatively very different models of evolution, both types of models will produce a blunderbuss pattern on a log scale. One might therefore legitimately worry and wonder what is so special or instructive about a blunderbuss pattern? In other words, no fundamentally novel insights into any of these two models of evolution can be gained from the visual discovery a blunderbuss pattern in your data. Thus, the fact that the visual pattern that is observed happens to be resemble a blunderbuss when plotted on a log-scale is neither special, surprising, or provides any interesting insight.

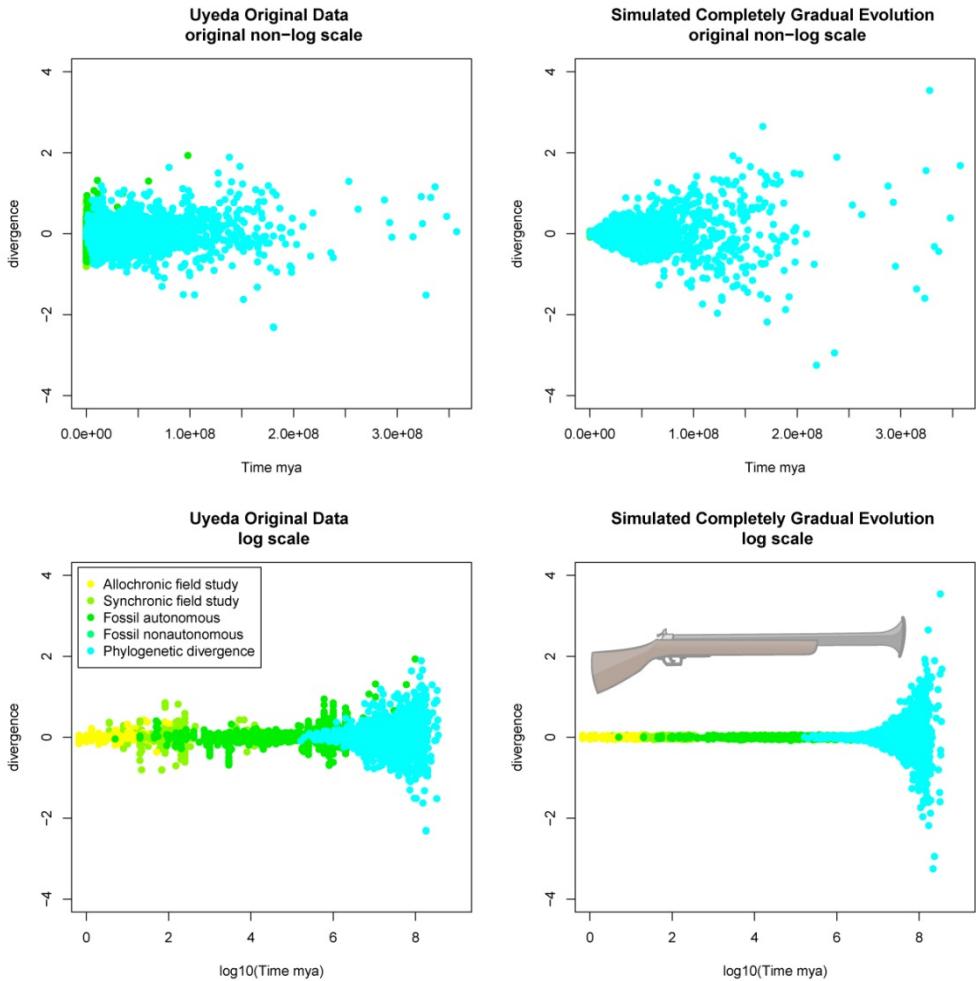


Figure 1.

Upper left panel, original data from Uyeda et al. (2011) plotted on original non-log scale. Here we see that much of the divergence data is not interpretable because divergence time is less than 100,000 years, explaining why Uyeda et al. would choose to plot it on a log scale in the lower left panel. In the upper right panel, I have used the time of divergence from the original Uyeda et al. (2011), but instead of using the empirical divergence, I have simulated divergence using a completely gradual model of evolution (Brownian motion). In the lower right panel, I have plotted this simulated gradual evolution on a log scale and a very distinctive blunderbuss pattern appears. From this plot it is clear that the logging of the y-axis causes all the data to be squished into the final third of the plot, resulting in the flared blunderbuss pattern even when phenotypic evolution is entirely gradual. An actual blunderbuss is also shown as a reference. Potentially, the blunderbuss illustrates the “long barrel of stasis,” but we see also that the BM-model can also produce a similar long barrel of stasis (Fig. 1; lower right panel) even though under this model, evolution occurs gradually and variance in phenotypic space is accumulating at a linear rate. Any appearance of bounded stasis in a Brownian motion model is therefore entirely due to it being plotted on a log scale. So if the visual metaphor of the blunderbuss as stasis cannot easily be distinguished from Brownian motion, is this metaphor even useful or is it simply misleading?

This commentary is not meant to be a criticism of the original paper of Uyeda et al (2011) nor of its methods. The intention of this commentary is only to draw attention to the fact that the blunderbuss pattern of body size divergences that was plotted in Uyeda et al. (2011, Fig. 1) is neither surprising or unique to their model or has any real biologically meaningful interpretive value. I therefore merely want to point out that the so-called blunderbuss pattern is almost guaranteed result when a class of closely related models of phenotypic evolution plotted on a log scale. The non-exhaustive list of interesting aspects of Uyeda et. al. (2011) are as follows:

1. The large dataset used.
2. The large temporal scale.
3. The diversity of taxonomic groups used.
4. The fact that the best-fitting model was a multi-burst model.
5. That the model parameters were consistent across several taxonomic groups.

The non-interesting aspects of Uyeda et al. (2011) is the fact that phenotypic divergence looks like an old-fashioned gun pattern when plotted on a log scale. However, this visual pattern seems to be the aspect that has captured the attention of other evolutionary biologists, and that has consistently and often been emphasized in many discussions and citations of the paper. For example, Voje (2016) cites the blunderbuss pattern as evidence that lineages evolve in narrow zones:

“That most traits show little net evolution but still undergo substantial change is consistent with the hypothesis that lineages mostly evolve within rather narrow adaptive zones, as envisioned by Simpson (1944, 1953) and as observed by Uyeda and colleagues in the blunderbuss pattern of body size evolution.” p. 2685 par. 2

Doebeli and Ispolatov (2016) go even further and refer to a “blunderbuss theory” of temporal patterns of macroevolutionary changes and diversification, but later more correctly characterize the results of the multi-burst model (Uyeda et al. 2011). But is the multi-burst model really best characterized as the “blunderbuss theory” when Brownian motion also produces a blunderbuss pattern? Fortelius et al. (2014) see the “the blunderbuss pattern” as compatible with the punctuated equilibria model of (Eldredge and Gould 1972), but never really discuss the multi-burst model.

Arnold (2014), who was one of the co-authors of Uyeda et al. (2011), recognizes that simple a simple BM-model can produce a blunderbuss pattern, but falls short of rejecting it as a useful visual metaphor as I do here (p. 738 para. 5). Later, Arnold (2014) also recognized that barrel of the blunderbuss does not represent absolute stasis:

“The bounded evolution revealed by the slender barrel of the blunderbuss should not be confused with stasis. The limits of that slender barrel lie 6 within-population standard deviations, or a 65% change in body size, on either side of literal stasis. Within those bounds, individual species can appreciably evolve.” p. 742 para. 5

Finally, even the first author of Uyeda et al. (2011) seems to be aware that the blunderbuss might have been taken the wrong way. In the comment section of Jerry Coyne’s blog “Why Evolution is True” (<https://whyevolutionisttrue.wordpress.com/2011/09/22/want-evolutionary-change-wait-a-million-years/>) about the article “Want evolutionary change? Wait a million years” user Uyeda writes:

“The blunderbuss pattern itself isn’t the product of punctuated equilibrium certainly. It is what you would see in either the case of a gradual Brownian motion or repeated punctuations (our Multiple-Burst model) on the log timescale (see Figure 3 in our paper) Consequently, it’s not surprising that it eventually looks like a blunderbuss, what was impressive to us was the consistency of pattern, timing and parameter estimates between independent data sources across a wide range of taxa.” (<https://whyevolutionisttrue.wordpress.com/2011/09/22/want-evolutionary-change-wait-a-million-years/#comment-135938>)

How useful is the blunderbuss as a visual metaphor for stasis or punctuated evolution? Here, I have argued that it varies from being misleading to technically correct, but for the wrong reasons. Often a creative use of new terms and metaphors can be a good way to get a concepts to stick in the minds of the reader, but I think in this case the mental imagery has been too strong. The metaphor of the blunderbuss pattern has caused many readers to believe it is mainly the visual pattern that is interesting, thereby drawing attention away from the other results of the paper. The fact that completely gradual evolution can produce a blunderbuss pattern, casts doubt on the metaphor as a useful mascot for stasis and punctuated evolution.

Methods

All analysis was done in R. R-code can be found in the supplementary material. I downloaded the dryad package of Uyeda et al. (2011) ([doi:10.5061/dryad.7d580](https://doi.org/10.5061/dryad.7d580)). Using both the “Dryad7.csv” and “Phylogeniesbynode.csv”, I extracted divergence times and phenotypic divergence data. Using this data, I re-plotted the original Figure 1 in Uyeda et al. (2011). I did not, however, include the phylogenetic time-series data, so the two figures will appear slightly different. Using the divergence times from the original study, I

simulated simple phenotypic evolution by Brownian motion in R ($\sigma = 0.001$). I simulated 5000 independent runs of phenotypic divergence under a Brownian motion model. In each model a discrete amount of time is drawn from a 8 million year time span and the lineages are allowed to evolve. At the end of the randomly chosen time period the divergence from the original population is measured. I then plotted these 5000 data points on both a log 10 scale and on the original scale of years.

Acknowledgement

I am grateful to Erik Svensson for comments on earlier drafts of this manuscript.

References

- Arnold, S. J. 2014. Phenotypic Evolution: The Ongoing Synthesis: (American Society of Naturalists Address). 729-746.
- Doebeli, M., and Ispolatov, I. 2017. Diversity and coevolutionary dynamics in high-dimensional phenotype spaces. *The American Naturalist*. 189.2: 105-120.
- Felsenstein, J. 1985. Phylogenies and the comparative method. *The American Naturalist*. 125.1: 1-15.
- Fortelius, M., Eronen, J. T., Kaya, F., Tang, H., Raia, P., and K. Puolamäki. 2014. Evolution of Neogene mammals in Eurasia: environmental forcing and biotic interactions. *Annual Review of Earth and Planetary Sciences*, 42: 579-604.
- Gould, S. J. 1972. Punctuated equilibria: an alternative to phyletic gradualism. pp 82-115 in "Models in paleobiology", edited by Schopf, TJM Freeman, Cooper & Co, San Francisco.
- Hansen, T. F. 1997. Stabilizing selection and the comparative analysis of adaptation. *Evolution*. 51.5: 1341-1351.
- Uyeda, J. C., Hansen T. F., Arnold S. J., and J. Pienaar. 2011. The million-year wait for macroevolutionary bursts. *Proceedings of the National Academy of Sciences*. 108:15908-15913.
- Voje, K. L. 2016. Tempo does not correlate with mode in the fossil record. *Evolution*. 70.12:2678-2689.

Paper V



Biotic and abiotic drivers of global body size distributions and macroevolutionary diversification in an old insect order

John T. Waller and Erik I. Svensson

Abstract

Global body size distributions in ectotherms are shaped by multiple ecological and evolutionary processes, including abiotic factors like temperature, natural and sexual selection on adults and biotic factors like natural enemies such as predators and parasites. However, little is known how these factors interact and shape global interspecific body size distributions, and latitudinal gradients in body size, such as “Bergmann’s Rule” (larger body sizes at higher latitudes). Here, we investigated the macroecological and evolutionary factors shaping latitudinal gradients and global body size distributions in an old insect order (Odonata; dragonflies and damselflies), using a global database of odonate phenotypic traits, fossils and phylogenetic comparative methods. Controlling for phylogeny, we find that extant species of both odonate suborders are larger in the temperate region than in the tropics. Next, using fossil data, we show that the relationship between body size and latitude has shifted over macroevolutionary time scales from having negative slopes in both suborders to become more shallow or even positive following the evolutionary emergence of birds 150 MYA. As birds are important predators on extant species of these insects, we argue that these macroevolutionary changing temporal pattern in the geographic body size-latitude trend is likely to be driven by avian predation in combination with high dispersal ability of large dragonflies. In further support of this, we show that interspecific body size variation of extant odonate taxa is negatively influenced by both regional avian diversity and temperature with a surprisingly large effect of the former. Finally, evolutionary shifts to larger body size were more likely to occur in the tropics,

followed by dispersal of large clades in to the temperature region. Once lineages invaded the temperate region, temperature had a limited direct effect on body size, but indirectly influences body size via prolonged larval development time. Ambient temperatures are therefore not the primary force shaping latitudinal body size variation in odonates, and instead predator diversity and dispersal ability might play more important roles in shaping interspecific geographic body size patterns. Thus, latitudinal body size gradients are shaped by both evolutionary processes leading to the emergence of large taxa in the tropics, followed by ecological processes such as increased dispersal ability of large clades.

Introduction

Body size is a fundamental phenotypic trait that is important to fitness and which is often the target of multiple and opposing selection pressures (Blanckenhorn 2000). Consequently, body size is one of the most popular phenotypic traits used by evolutionary biologists who employ phylogenetic comparative methods to study evolutionary processes and outcomes (Baker et al. 2015; Waller and Svensson 2017). For instance, in vertebrates such as birds, mammals, reptiles, and amphibians, body size has been investigated in relation to “Cope’s Rule”, the tendency for lineages to become larger over evolutionary time (Baker et al. 2015) and to evaluate the relative role of cladogenetic (“speciational”) vs. anagenetic (“gradual”) evolution (Mattila and Bokma 2008). In addition, researchers have tried to understand the temporal macroevolutionary pattern of body size diversification in adaptive radiations, looking for so-called “early bursts” of body size evolution in comparative data (Harmon et al. 2010). In addition to these evolutionary studies, body size is also of interest for researchers interested in physiological ecology, development and biogeography, and the intersection between these fields. For instance, the classical “Bergman’s Rule” states that body size in endotherms should increase with latitude, as larger bodies conserve more heat due to higher thermal inertia, which might be beneficial in colder environments (Ashton et al. 2000; Shelomi 2012).

While geographic and latitudinal body size patterns of mammals and birds can at least partly be attributed to Bergmann’s rule (Ashton et al. 2000), which predicts an increase in body size in colder environments at high latitudes, strong support for this rule is lacking in insects (Shelomi 2012). It is also difficult to make straightforward predictions from first principles about the expected relationship of body size and latitude in ectotherms (Mousseau 1997), and studies of environmental influences on body size in insects often lack a general theoretical framework (Shelomi 2012). Although there are some documented cases of Bergmann’s rule in ectotherms (VanVoorhies 1997), the interpretation of such

latitudinal patterns is subject to controversy (Mousseau 1997; Partridge and Coyne 1997). There is currently no consensus on the evolutionary and ecological causes behind latitudinal size gradients in ectotherms, or even the consistency of Bergmann's rule (Mousseau 1997; Shelomi 2012). Different species show either no latitudinal trends, positive trends or even negative body size trends with latitude (Mousseau 1997; Shelomi 2012).

In the case of temperature, the body size of ectothermic animals typically increases with decreasing developmental temperatures, a phenomenon known as the temperature-size rule (Kingsolver and Huey 2008). The explanation that has been put forward for this rule is accelerated maturation at higher temperatures, and it is supported by the fact that most ectothermic organisms mature at smaller body sizes when reared in warmer conditions (Atkinson 1996; Angilletta and Dunham 2003; Kingsolver and Huey 2008). However, warm temperatures could also lead to the opposite effect, as metabolic rate and season length increase with temperature, allowing higher growth rates under warmer climatic conditions. Additionally, differences in moisture and precipitation may increase or decrease desiccation resistance, since larger individuals may be more resistant than smaller to desiccation (Marron et al. 2003; Harrison et al. 2013; Bujan et al 2016). Starvation resistance can also increase with increasing size, and starvation resistance has been associated with latitude in some insects (McCue 2010). For both endotherms and ectotherms like insects, the balance between heat gain and heat loss is also important, which might cause associations between body size and latitude (Olalla-Tárraga et al. 2006). Finally, past geological and climatological events (such as glacial advances and retreats; Bhagwat 2008) can lead to different size assemblages at different latitudes, as dispersal, extinction, and speciation may all correlate with body size (Clauset and Erwin 2008; Rainford et al. 2016; Waller and Svensson 2017). Many abiotic and biotic variables are therefore associated with latitude and could potentially produce latitudinal differences in body size for different insect groups, even if clear predictions are not always possible for insects in general (Blackburn et al. 1999; Ashton et al. 2000; Stillwell 2008, 2010).

Body size evolution in odonates has a dramatic and interesting history (Waller and Svensson 2017; their Fig. 1). Over the last 350 MYA, maximum odonate wing length has varied by more than 60 mm (Waller and Svensson 2017), from crow-sized animals to those smaller than the width of a human hand (Baird 2014). Oxygen has been suggested to be a major driver of body size variation through time in this lineage (Clapham and Karr 2012). In support of this, larger extant dragonflies have been shown to adjust their flight behaviour in experimentally reduced oxygen (Henry and Harrison 2014). Moreover, dragonflies, unlike other insects, lack the ability to ventilate via abdominal pumping, but they must rely on actual wing beats in order to ventilate their bodies (Henry and Harrison 2014), so they might be more constrained by atmospheric oxygen than other insects.

Over the last 65 MYA, however, atmospheric oxygen levels have remained constant, and oxygen levels are uniform in the atmosphere (Glasspool and Scott 2010), (although lower oxygen levels are observed in colder water (Verberk and Atkinson 2013)). This has led some authors to argue that the evolution of flying competitors, starting around 150 MYA, is now the primary force keeping odonates and other flying insects small (Damuth 1981; Blackburn and Gaston 1994; Clapham and Karr 2012). Some field studies of predator-mediated selection by birds on damselflies have shown that avian predation favors smaller individuals Svensson and Friberg 2007; Outomoro and Johansson 2015). So far, however, no study have investigated whether the presence of birds is predictive of geographic body size patterns in extant dragonflies and damselflies.

Here we investigate the environmental and global macroecological factors involved in the biogeographic patterns of body size distributions in Odonata. In a recent study (Waller and Svensson 2017), we found evidence for multiple non-neutral body size shifts within this insect order. These non-neutral body size shifts are likely to be associated with the invasions in to new habitats with novel selective environments, for instance with novel predator faunas or differences in overall extent of predation pressure (McPeek et al. 1996). Sokolovska et al. (2000) and Waller and Svensson (2017) found that selection on adult body size is largely positive, but this positive selection has not lead to a directional body size trend in the fossil record, at least not since the last mass extinction about 65 MYA (Waller and Svensson 2017). Waller and Svensson (2017) argued that long development time is the main selective counterforce opposing the consistent positive selection in the adult stage, leading to weak net selection due to conflicting selection in most lineages. However, although such conflicting selection is likely to operate on body size in odonates, there are nevertheless several documented bursts of body size evolution within this order (Waller and Svensson 2017). This observed pattern of multiple body-size changes in Odonata fits with a recently proposed multiple burst model of phenotypic evolution, where bursts are rare and are separated by long periods of phenotypic stasis (Uyeda et al. 2011).

Therefore, present-day variation in body size in odonates could be due to multiple and sometimes conflicting environmental factors and associated selection pressures including differences in development times, environmental temperatures, habitat choice, or predation. Based on the results in this study, we suggest that ambient temperatures have indirectly influenced body size distributions, with dispersal ability, predation, and niche filling by birds playing more important roles in shaping latitudinal body size gradients in this insect order.

Methods

Phylogeny used

The phylogenetic tree that we use in this study was constructed from odonate DNA-sequences downloaded GenBank, using a traditional morphological taxonomy as our backbone (Waller and Svensson 2017). This phylogenetic tree has been used in a previous study on odonate body size (Waller and Svensson 2017). Our phylogeny comprised 1322 taxa, which is about 21% of all known extant odonate species (about 6400 in total; Schorr and Paulson 2015). See our Supplementary Material for a detailed description of the construction of the phylogenetic trees used in this study, and how we accounted for phylogenetic uncertainty.

Geographic variation in body size in relation to latitude

We quantified the relationship between latitude and two measures of body size (wing length and total body length) in both fossils and extant taxa. For the fossil data, we used wing length as our size measure, which is justified by the fact that overall body length and wing length are highly correlated ($r = 0.90$) with each other (Waller and Svensson 2017). For the extant taxa, we used both total body length (in mm; when available from the literature) and hindwing length (also in mm) in different analyses. We divided the fossils into different time periods between extinction events and other major evolutionary events: 1) Around 210 MYA marks the beginning of the dragonfly radiation after the Triassic–Jurassic extinction mass extinction; 2) 150 MYA marks the beginning of the radiation of birds; 3) 80 MYA right before the Cretaceous–Tertiary (K–T) extinction event (there is a large gap in the fossil record between 80–65 MYA). For the fossils, paleolatitude was extracted from the Paleobiology Database (Alroy et al. 2012). Mean latitude from the Global Biodiversity Information Facility (GBIF: <https://www.gbif.org/>) occurrence records was used to obtain distribution data from extant taxa. Absolute latitude (the distance in degrees from 0°) was used in all analyses.

Odonate size data

Odonate size data were taken from the Odonate Phenotypic Database (opdb) (www.odonatephenotypicdatabase.org). This open database is a general resource for comparative studies on phenotypic trait evolution using odonates. These size

data has also been used in a previous publication, and a description of how it was collected can be found there (Waller and Svensson 2017). Size data from fossil odonate taxa were obtained from Clapham and Karr (2012).

Geographical occurrences and tropical and temperate classifications

We considered 23°N and 23°S to be the boundaries for temperate latitudes, in accordance with the Köppen-Geiger climate classification system (Peel et al. 2007). Any species with a mean distribution of occurrence records above or below these latitudinal boundaries was classified as temperate. Conversely, all species with a mean distribution between these two points was classified as tropical. Occurrence records were downloaded from the Global Biodiversity Information Facility (GBIF: <https://www.gbif.org/>). All occurrences for each odonate species under 500 records were downloaded using the R-package rgbif (Chamberlain 2017). When a species had more than 500 occurrence records, we used a randomly chosen sample of 5000 records. This procedure was done to avoid downloading species with prohibitively large occurrence histories, while still accurately capture their geographic distributions.

Spatial environmental data

Spatial data was linked to quantify and investigate the environmental influences on these odonate species in terms of ambient temperatures, tree cover, precipitation, and bird diversity. These spatial environmental datasets were processed in the R packages raster (Hijmans et al. 2016) and sp (Bivand et al. 2013). Temperature and precipitation data were downloaded from the worldclim datasets (Fick and Hijmans 2017). WorldClim is a set of global climate layers (gridded climate data) with a spatial resolution of about 1 km². These data can be used for mapping and spatial modeling. These WorldClim datasets were downloaded using the R-package raster at 5m accuracy. Mean values for temperature (BIO1 = Annual Mean Temperature) and precipitation (BIO12 = Annual Precipitation) were computed for each species in our occurrences dataset.

Tree cover data was downloaded from the global land cover facility (glcf: <http://glcf.umd.edu/data/landsatTreecover/>) (Sexton et al. 2013). Tree cover is the proportional, vertical area of vegetation (including leaves, stems, and branches) of woody plants above a given height. The continuous classification scheme of glcf product enables a better depiction of land cover gradients than traditional discrete classification schemes (Sexton et al. 2013).

Avian species diversity spatial data was downloaded from biodiversitymapping.org

(<http://biodiversitymapping.org/wordpress/index.php/birds/> Jenkins et al. 2013). These maps were generated based on the Bird Species Distribution Maps of the World (BirdLife International 2011). We used total bird diversity in our analyses. To control for a possible effect of a biodiversity in general we also used mammal biodiversity data from the same source (Jenkins et al. 2013). The rationale here is that while birds are known predators on odonates, mammals are not, and hence we would expect an ecological effect of the former on odonate size distributions, but not the latter.

Phylogenetic path analysis

To investigate whether temperature had a direct effect on our body size in European odonates, we performed a phylogenetic path analysis using the R-package *phylopath* (Van der Bijl 2017). We also collected data on larval development from the literature of European odonates (Sandhall 1987). This is the same larval development time data that has been used in a previous study (Waller and Svensson 2017). We compared five nested models of the relationships between larval development time, temperature (annual mean temperature).

These path models differed in their final causal structure, but they all contained three variables: adult body size (a quantitative trait), environmental temperature at the locality of the species (at the midpoint of the species range; another quantitative trait) and development time (an integer trait that ranges from one to several years; Sandhall 1987; Waller and Svensson 2017). In these models, body size was always the dependent variable and it could be affected directly by development (i.e. longer development leads to larger adult body size) and both directly and indirectly by ambient temperature. For instance, adult body size could be directly influenced (either negatively or positively) by ambient temperatures, through physiological effects or competitive interactions with con- or heterospecific individuals leading to selection for larger or smaller body size. Alternatively, body size could be indirectly influenced by environmental temperatures through changes in larval development time, e.g. in warmer climates odonates would develop faster but at a smaller size, consistent with many previous studies on larval ecology and thermal adaptation in this group (e.g. Johansson 2003; Stoks and Aguilar 2012; Örtman et al. 2012).

Optima shifts and invasion of the temperate zone

We tested the relationship between positive shifts in size optima with invasion in to the temperate zone, using the *corDISC* function in the R-package *corHMM* (Beaulieu et al. 2017). Optima shifts were taken from the multi-optimum OU-

model inferred in Waller and Svensson (2017), using the 11ou R-package (Khabbazian et al. 2016). A positive shift in the body size optimum shift was defined as an increase from the background inferred neutral process. A negative body size optimum shift was classified the same as neutral background optima for this analysis. It is important to realize that a positive shift in optima does not necessarily imply that species within the positively shifted group will be largest odonates species, but they will be larger than closely related species that did not evolve toward that new inferred optimum. These shifts in optima can also roughly be interpreted as the signature of past selective regimes favouring larger or smaller size (Khabbazian et al. 2016).

Models linking hindwing length to spatial environmental data

To quantify the relationships between body size and environmental variables, we conducted a phylogenetic generalized least square (PGLS) regression using the corBrownian function in ape (Paradis et al. 2004). We used a model comparison procedure based on AIC. The best performing models coefficients were then interpreted, using standardized coefficients. Five spatial environmental variables were considered: 1) temperature (annual mean temperature), 2) precipitation (annual precipitation), 3) tree cover, 4) avian diversity, and 5) mammal diversity. The last of these variables was included to control for a possible “diversity hotspot”-effect rather than avian diversity.

Results

Global variation in wing length and geographic pattern in wing length and body size

We first compared three different models of body size evolution, using our entire global dataset, our phylogeny, and hindwing length as our measure of body size (Table 1). We found the highest support for an Ohrnstein-Uhlenbeck (OU) model, followed by a Brownian Motion (BM) model and the lowest support for an Early Burst (EB) model (Table 1).

Tabel 1. Models of continuous body size (hind wing length) evolution.

Models were fit using the `fitContinuous` function in the R package `geiger` (Harmon et al. 2008). Three nested models were compared: 1) Ornstein–Uhlenbeck (OU), 2) Brownian motion (BM) and 3) Early Burst (EB).

Model	LnLik	Parameter	Estimate	AIC	Rank
OU	-2371.0	Alpha	0.0089	4748.007	1
		sigma sq.	1.364		
BM	-2389.306	sigma sq.	1.085	4782.612	2
EB	-2389.308	Alpha	-1.00e-06	4784.616	3
		sigma sq.	1.085		

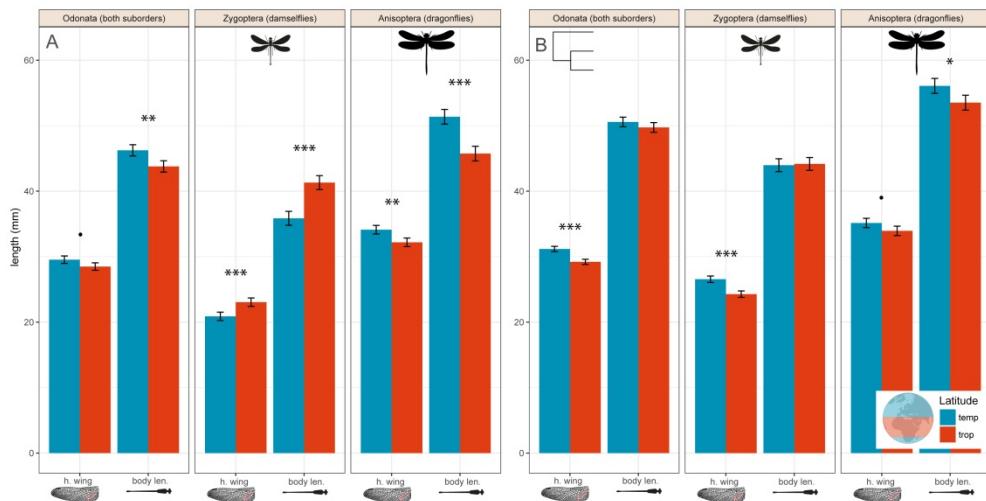


Figure 1. Body size in temperate and tropical regions.

Mean body size (hind wing length and total body length) of dragonflies and damselflies in temperate and tropical regions, when using individual species as independent datapoints (A) and controlling for phylogenetic relationships (B). We considered the tropical region to occur between 23°N and 23.0°S , whereas areas north and south of these regions were considered to be temperate, following the Köppen–Geiger geographical classification scheme. Across both suborders, odonates are larger in body size in temperate regions (controlling for phylogenetic non-independence; B).

We proceed by comparing wing and body sizes of dragonflies and damselflies of the tropical and the temperate regions (Fig. 1). We found that dragonflies were significantly larger in the temperate region whereas damselflies were significantly larger in the tropical regions, i.e. opposite geographic patterns in these two different suborders (Fig. 1A). After controlling for phylogeny using a phylogenetic generalized linear model (PGLM), however, we found that this latitudinal size-pattern was reversed in damselflies, and now both suborders were larger in the temperate region (Fig. 1B). This indicates that the geographic pattern in the non-corrected analysis for damselflies was partly confounded by phylogeny. This might suggest that some large damselfly clades are overrepresented in the tropics and these clades have not yet dispersed in to the temperate region (Fig. 1A,B).

Macroevolutionary trends in the fossil record between wing size and latitude

In the fossil record, damselflies showed a consistent pattern of decreasing body sizes with increasing reconstructed latitude (Paleo-Latitude) across all geological time periods, although the slope became shallower closer to the present (Fig. 2; Tables 2-3). Dragonflies showed a more varied temporal pattern with decreasing body size with increasing latitude prior to 65-80 MYA (large gap in the fossil record), followed by a reversal and more recently increasing size with latitude Fig. 2; Tables 2-3). The slopes of the relationship between wing size and latitude differed significantly between the time periods for both suborders, as shown by significant interaction terms between paleo-latitude and time period (Table 2). The general conclusion is that in the past, both dragonflies and damselflies were larger in the tropics than in the temperate region, but this strong negative relationship became less pronounced and even reversed as we move towards the present (Fig. 2; Table 3).

Evolutionary shifts in body size followed by invasions in the temperate regions

A binary test of correlated evolution suggest that evolutionary shifts to larger body sizes occurred in the tropics followed by range shift in to the temperate zone (Fig. 3). We tested a series of nested models in order to identify the best fitting model using the corDISC function in the R-package corHMM (Fig. 3). We compared three different models: the “all rates different” (ARD), “equal rates”(ER), and “symmetric rates” (SYM) models in corHMM. We found that the ARD had the lowest AIC score, implying that transition rates were not equal (AIC: ARD = 662.62, SYM = 699.39, ER = 804.09)(Fig. 4).

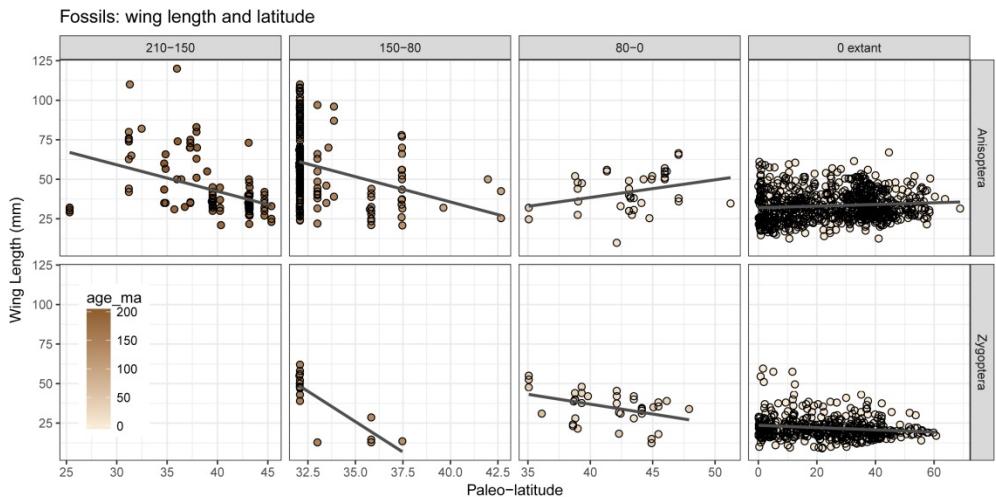


Figure 2. The continuous relationship between absolute latitude and hind wing length for fossils and extant dragonflies.

The relationship between latitude and body size has been negative for the major part of the evolutionary radiation of Odonata, including both dragonflies (Anisoptera) and damselflies (Zygoptera). However, immediately before and after the last mass extinction (65 MYA) event, dragonflies have shifted towards a positive relationship with latitude. Plot have been divided according to major geological events, namely major extinction events and following the beginning of the radiation of birds (150 MYA). See Tables 2-3 for statistical tests of these relationships, as well as slope differences between the time periods for both fossil taxa. For all plots we used absolute latitude, which is the distance in decimal degrees from the equator. Latitude for the fossils is paleo-latitude, which takes into account effects of continental drift.

Table 2. General linear model (glm) of the relationship between size and latitude in Zygoptera and Anisoptera fossils.

Period is a discrete categorical variable with 4 levels corresponding to the panels in Fig. 2 (extant, 80-0 mya, 150-80 mya, 210-150 mya). Latitude is absolute latitude.

Suborder	Term	Df	N	P-value	Significance
Zygoptera	Period	2	529	< 0.0001	***
	Latitude	1		< 0.0001	***
	Period x Latitude	2		< 0.0001	***
Anisoptera	Period	3	1219	< 0.0001	***
	Latitude	1		< 0.0001	***
	Period x Latitude	3		< 0.0001	***

Table 3. Linear models of the relationship between size and latitude in Zygoptera and Anisoptera extant taxa and fossils.

Analyses in this table are not corrected for phylogeny, but use species as single datapoints. Note that in Zygoptera, the phylogenetically un-corrected analysis shows that species are larger in the tropical region, whereas when we correct for phylogeny, we find that the reverse, i. e. both damselflies and dragonflies are larger in the temperate region. Period is a discrete categorical variable with four levels, corresponding to the panels in Fig. 2 (extant, 80-0 mya, 150-80 mya, 210-150 mya). Latitude is absolute latitude.

Suborder	Period	Variable	Estimate	N	P-value	Significance
Zygoptera	extant	Latitude	-0.071	473	< 0.0001	***
	80-0 mya	Paleo-Latitude	-1.248	39	0.007	**
	150-80 mya	Paleo-Latitude	-7.689	17	< 0.0001	***
Anisoptera	extant	Latitude	0.052	798	0.008	**
	80-0 mya	Paleo-Latitude	1.117	39	0.058	.
	150-80 mya	Paleo-Latitude	-3.198	252	< 0.0001	***
	210-150 mya	Paleo-Latitude	-1.676	130	< 0.0001	***

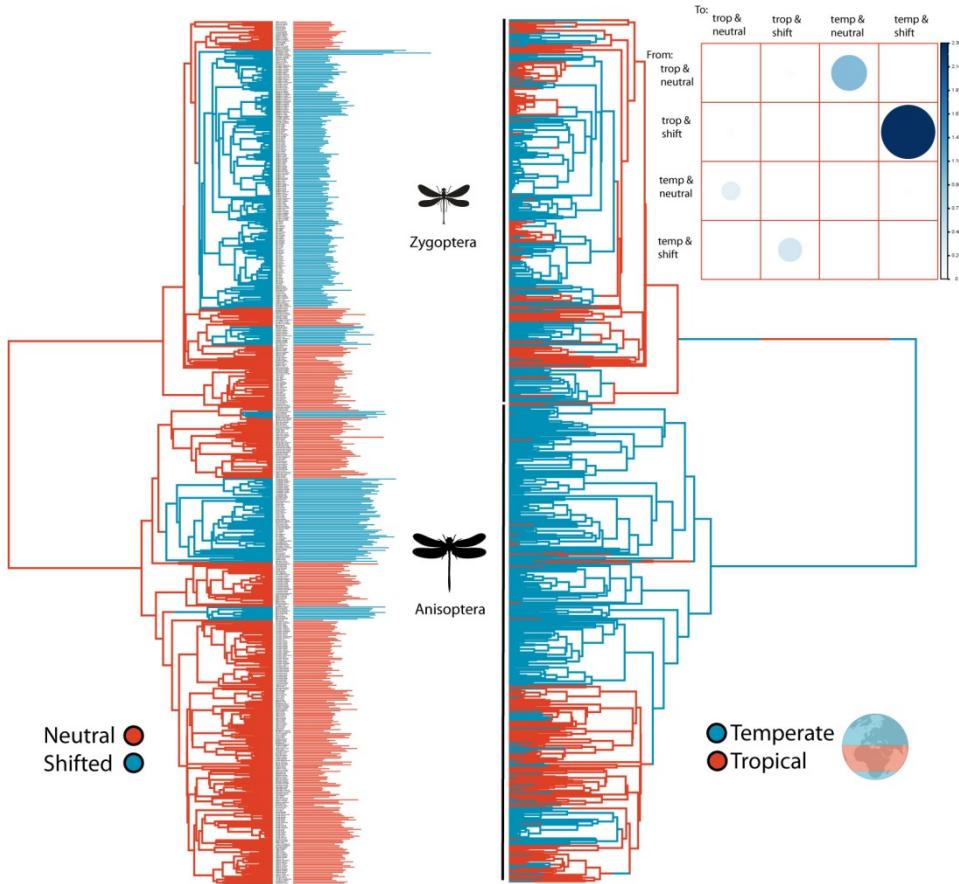


Figure 3. The continuous relationship between absolute latitude and hind wing length for fossils and extant dragonflies.

The evolution of temperate range plotted against the evolution of positive shifts in body size. The matrix diagram in the upper right visualizes the transition matrix from our corDISC model (all rates different) (Beaulieu et al. 2017). In this matrix we see that the largest transition rate is from lineages with tropical ranges and positively-shifted body size optima to temperate ranges and positively-shifted optima. This matrix indicates that larger body sizes are primarily not evolving at temperate latitudes, but instead emerge first in the tropics, followed by dispersal by such shifted clades in to temperate regions. This suggests that temperature is not the primary factor driving geographic patterns in body size, but instead that increased dispersal ability of large clades after glacial events might have been more important in establishing the current body size latitudinal gradient in odonates.

Direct and indirect relationship between temperature and body size

Our phylogenetic path analysis revealed that temperature has only an indirect influence on body size distribution in European Odonata (Fig 4). Our best model (model 1) was one in which development time has a direct negative effect on total body length, and temperature has a direct negative effect on development time. Model averaged coefficient estimates (over all five models tested), showed that temperature only had a weak and non-significant direct negative effect on total body length compared to development time (Fig. 4), which means that temperature may not play a major role in structuring the distributions of larger and smaller species in Europe.

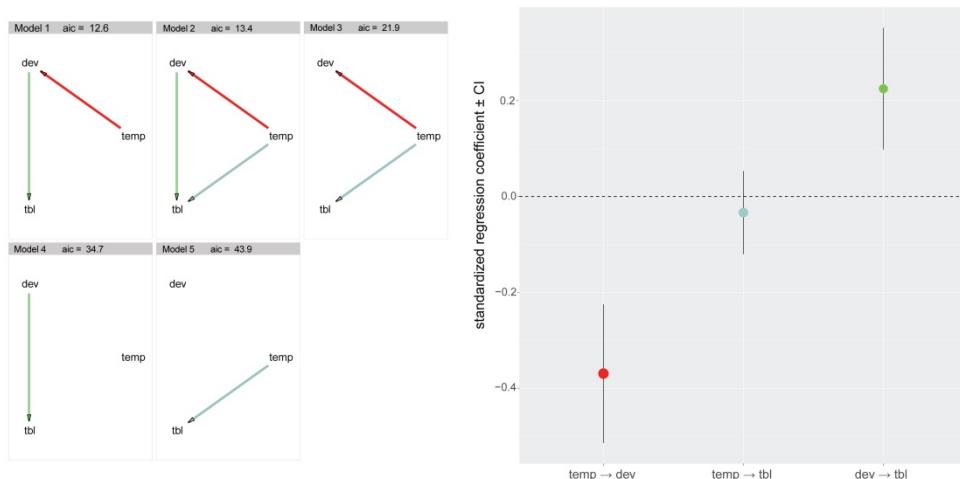


Figure 4. Diagrams of five nested phylogenetic path models.

These models show the statistical support for different cause-effect relationships between larval development time, temperature (annual mean temperature), and body size (total body length). Models are plotted in order of their AIC score (lowest first). The best model (model 1) was one in which development time has a direct positive effect on body size (tbl: total body length), and temperature has a direct negative effect on development time. However, we found no support for a direct effect of temperature on body size, only an indirect negative effect through the effect of temperature on development time. Model averaged coefficient estimates (over all the 5 models) illustrates this weak non-significant effect of temperature on total body length, compared to development time (right panel). This implies that temperature does not directly influence the geographic distribution of body size in Europe, but that temperature might only indirectly influence body sizes through its effect on development time. All models were fit using the R-package phylopath (Van der Bijl 2017).

Geographic variation in hind wing length in relation to spatial environmental variation

We quantified how different environmental variables (Fig. 5) affected hindwing length in a series of nested phylogenetic generalized least squares (PGLS) models (Table 4). In Model 1, we explored how temperature affected wing length. Model 3 suggested that when considered alone, temperature has a significant negative effect on wing length (Table 4). However, after accounting for other potentially important ecological variables such as tree cover, precipitation, and avian diversity (species richness), the effect of temperature on interspecific size variation was reduced, whereas avian diversity was found to have a strong and large negative effect (Table 4). To investigate whether avian diversity was simply measuring a general species diversity effect, mammal diversity was included as a covariate, since mammals are unlikely predate on dragonflies or to occupy the same habitat niches as odonates. After accounting for mammal diversity, avian diversity still had a strong and significant negative effect on hind wing length (Table 4).

Table 4. Phylogenetic Generalized Least Squares (PGLS).

Phylogenetic Generalized Least Squares (PGLS) models of global geographic variation in odonate hind wing lengths. Five models of the relationship between hind wing length and geographic position were compared. All models corrected for phylogenetic non-independence using the *corBrownian* function in ape (Paradis et al. 2004), using the same phylogenetic tree ($N = 674$) and species set for structuring the covariance matrix. The best performing model (Model 4) is shown in bold. The intercept term is not presented for simplicity. All spatial variables were scaled to have mean 0 and variance of 1, so as to be comparable to each other.

Model	Variable	Est	SE	P value	N	AIC	Rank
1	Temperature	-0.964	0.159	< 0.001	674	4157.17	3
2	Temperature	-0.814	0.179	< 0.001	674	4157.21	4
2	Precipitation	-0.366	0.200	0.0677			
3	Temperature	-0.789	0.181	< 0.001	674	4159.82	5
3	Precipitation	-0.394	0.203	0.0525			
3	Tree Cover	-0.174	0.204	0.3933			
4	Temperature	-0.437	0.191	0.0223	674	4136.57	1
4	Precipitation	-0.253	0.201	0.2083			
4	Tree Cover	0.187	0.212	0.3776			
4	Avian Diversity.	-1.421	0.276	< 0.001			
5	Temperature	-0.451	0.192	0.0192	674	4138.09	2
5	Precipitation	-0.237	0.203	0.2433			
5	Tree Cover	0.166	0.215	0.4403			
5	Avian Diversity	-1.597	0.411	< 0.001			
5	Mammal Diversity	0.248	0.428	0.5625			

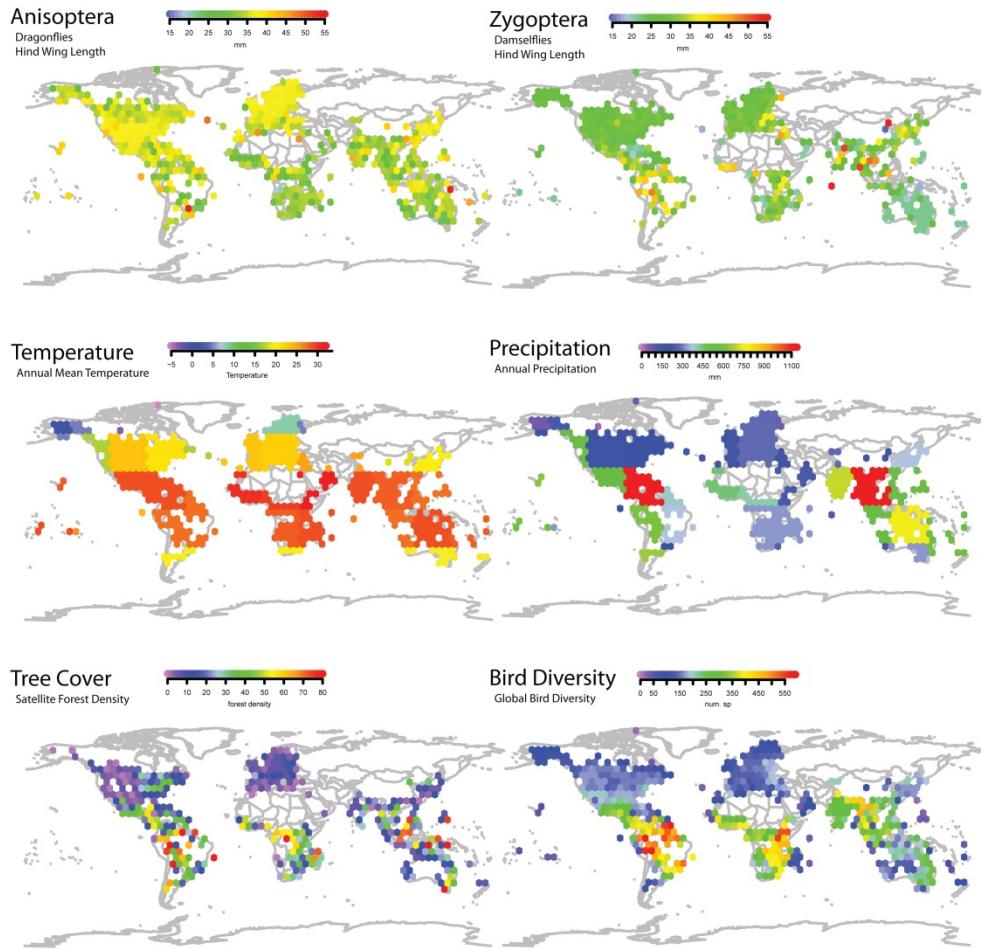


Figure 5. Spatial hexagon plots.

Spatial hexagon plots of the five main geographic datasets used in our PGLS-models of factors influencing odonate body sizes across the globe. Each hexagon represents the mean value of the variable over which the hexagon cover, but not over the entire space but only from points drawn from the odonate occurrence dataset. Thus, the value from each hexagon represents only the values of that spatial variable that are relevant to odonates. Hexagons are spaced by around 5° between each hexagon. Plots are shown to illustrate the global distribution of the data. Mean values for each species were used in all analyses, which limits the effect of spurious spatial auto-correlation. See Materials and Methods for a description of the spatial variables that we used.

Discussion

Overall, body size evolution and interspecific diversification in odonates show signs of constrained divergence (Waller and Svensson 2017), where accumulation of body size diversity is clearly non-neutral (Table 1), but there is no signature of a so-called “early burst”, which is one expected pattern in adaptive radiations (Schluter 2000; Harmon et al. 2010). This prediction is, however, often discussed in the context of adaptive radiations driven by ecological resource competition and niche divergence, which might not apply to odonates, which are characterized by only weak ecological niche differentiation, as these are generalist predators with no or little differentiation in thermal or climatic niches (McPeek and Brown 2000; Siepielski et al. 2010; Wellenreuther et al. 2012; Svensson 2012; Svensson et al. 2018). Based on the ecological features and the natural history of odonates, we would argue that we should not even a priori expect such an early burst pattern, neither for body size nor for other traits, as phenotypic evolution in these ectothermic insects is more likely to be driven by environmental tolerance to temperature and other abiotic factors, dispersal and natural enemies, like predators and parasites, rather than competition for resources or feeding niches. Results in this study supports this view, as further discussed below.

We found some interesting differences in geographic body size distributions in dragonflies and damselflies, with dragonflies being larger in temperate regions and damselflies being larger at tropical latitudes (Fig. 1A; Table 3). This difference between the two suborders disappeared, however, when we corrected for phylogeny (Fig. 1B). This difference between the two types of analyses can be explained by the fact that some large damselfly clades in the tropics that have not invaded the temperate regions (e.g. the large helicopter damselflies in the genus *Megaloprepus*; Waller and Svensson 2017), and such clades might partly drive this latitudinal pattern, which becomes reversed when phylogeny is taken in to account (Fig. 1). The general picture that emerges is that both dragonflies and damselflies are on average larger in the temperate regions, compared to the tropics, when phylogeny is taken in to account. Interestingly, this latitudinal trend might have appeared more recently, judged by significant negative trends between latitude and body size in the fossil record (Fig. 2; Tables 2-3).

On macroevolutionary time scales, the relationship between latitude and wing size in the fossil record revealed changing temporal patterns in both dragonflies and damselflies (Fig. 2; Tables 2-3). Prior to, and up to the start of the major evolutionary radiation of birds (210-80 MYA) (Padian and Chiappe 1998), dragonflies showed a pattern of decreasing body size with increasing paleolatitude, i.e. a reversed Bergmann’s Rule. Then, right before and after the last major extinction event (65-80 MYA), this latitudinal trend became reversed (Fig. 2).

These results indicate that temperature is only one of several factors driving the dynamically changing latitudinal-size trend, among both extant and fossil taxa. Temperature might even play a secondary role compared to many other important ecological factors, both abiotic and abiotic (Figs. 2–5). The macroevolutionary temporal difference between the two suborders might also imply that they follow different temperature-size rules in terms how temperature structures the geographic organization of body size variation (Fig. 2). For instance, the generally larger bodied dragonflies might be more sensitive to the temperature-size rule than damselflies (Forster et al. 2012). However, the fact that both suborders have changed the slope sign (dragonflies) or slope magnitude (damselflies) of their latitudinal relationship over geological time (Fig. 2) indicates that a simple version of the temperature-size rule is not sufficient to explain the latitudinal size gradients. As this latitudinal relationship is not stable over geological time (Fig. 2), it is unclear exactly to what extent and how the classical temperature-size rule at the intraspecific level (Kingsolver and Huey 2008) contributes to structure the interspecific geographic size assemblages over the entire globe (Figs. 2 and 5).

Our phylogenetic path analysis revealed that, at least for the odonates that inhabit temperate latitudes in Europe, temperature does not directly influence on body size, but do instead indirectly influence body size indirectly via development time. A direct effect, if present, would have been consistent with temperature as the primary variable structuring geographic interspecific variation of adult body size in Europe. However, we instead found evidence for an indirect negative effect of temperature on body size through its effect on development time (Fig. 3). The lack of any direct effect of temperature on adult size can be interpreted as there are no habitat preference for cooler habitats among large-bodied odonates, such as dragonflies.

When we linked our previously demonstrated evolutionary shifts towards larger body size optima (Waller and Svensson 2017) to temperate or tropical ranges, we found that relatively large-sized odonate lineages in the temperate zone did not evolve there *in situ* but rather invaded those high latitude areas after they had evolved larger body size in the tropics (Fig. 4). This suggests that the pattern of larger dragonfly species in the temperate region compared to the tropics is not a simple consequence of the classical temperature-size rule documented in many intraspecific experimental studies (Kingsolver and Huey 2008). Instead, this latitudinal pattern is likely to partly result from higher dispersal rates and/or colonization ability of large-bodied dragonflies from the tropics, where they originated, in to the temperate zone. Thus, lower temperatures do not necessarily lead to the evolution of larger body size in dragonflies or odonates at the macroevolutionary level, but it is rather so that lineages that have evolved larger body sizes are more likely to invade the temperate region (Fig. 4).

Across our global dataset, we found that temperature had a negative effect on hind wing size in both dragonflies and damselflies, after controlling for

phylogenetic non-independence (Table 4). Accounting for other confounding variables, such as tree cover, precipitation, and bird and mammal diversity (Fig. 5), this effect of temperature was reduced, although still significantly negative (Table 4) Interestingly, regional avian diversity (species richness) was found to be the strongest predictor of hind wing length (Table 4). This negative effect of bird diversity on odonate size remained even after accounting for other confounding variables, such as temperature, tree cover, and precipitation (Table 1).

There is empirical evidence from ecological field studies and suggestive indications from paleontological time series that predators are a major selective pressure against large body size in odonates and other insects (Gotthard et al. 2007; Clapham and Karr 2012). More specifically for odonates, it has been shown that avian predators (white wagtails, *Motacilla alba*) selected against longer wings in two different extant damselfly populations of the damselfly *Calopteryx splendens* in Sweden (Svensson and Friberg 2007; Outomuro and Johansson 2015). In a third study of *C. splendens* in Finland, the researchers found no evidence for significant selection on neither fore- nor hindwing length (Rantala et al. 2011), whereas in a fourth independent study in Sweden, there was evidence for stabilizing selection on forewing length (Kuchta and Svensson 2014).

Based on these data, we suggest that the macroevolutionary changing body size-latitude pattern (Fig. 2; Tables 2-3) might reflect that prior to the evolutionary adaptive radiation of the major bird groups, dragonflies might have occupied aerial niches in lower latitudes that gradually and subsequently became occupied by birds. Alternatively, although not mutually exclusive, predation from birds and other biotic interactions gradually increased at the lower latitudes in the tropics as bird diversity accumulated, compared to higher latitudes in the temperate region. There is indeed a small but growing body of empirical evidence suggesting that predation rates on insects are higher in the tropics compared to the temperate region (Schemske et al. 2009; Freestone et al. 2011; Roslin et al. 2017). Whatever the exact mechanism behind these geographic patterns of body size in relation to latitude, these relationships might suggest that the larger dragonflies could be more successful in colonizing temperate latitudes after glacial retreats than are damselflies, given that extant dragonfly taxa show a stronger positive relationship with latitude (Fig. 2; Tables 2-3). This stronger relationship in dragonflies could be due to a positive relationship between size, flight performance and by extension colonization ability. In general, damselflies are smaller and less mobile than dragonflies and, not surprisingly, all the known long-distance migrants among odonates are dragonfly species, not damselflies (Wikelski et al. 2006; Anderson 2009; Chapman et al. 2015; Troast et al. 2016).

To investigate whether the effects of avian diversity on geographic size distribution of odonates (Fig. 5) was due to some confounding and unknown role of high biodiversity in general, we used mammal diversity as a control variable (Fig. S1). It is unlikely that mammals would compete for same niches as odonates

and mammals are not known to be significant predators of odonates. Mammals still show high species diversity in the tropics, just like birds (Fig S1). After including mammalian diversity in a model with precipitation, tree cover, temperature, and bird diversity, we found that avian diversity still had a large and negative effect on body size in odonates (Table 4; model 5). We therefore conclude that the effect of avian diversity on odonate body size geographic distributions cannot simply be explained away as a simple general biodiversity effect. Even given this control, however, we cannot of course exclude that some third variable that is correlated with avian diversity is the causal factor that structures geographic body size variation in odonates. Whether the effects of avian diversity is reflecting niche competition, predation, or a combination of the two is not possible to test with the current data and using phylogenetic comparative analyses alone. These results do, however, fit with suggestions that biotic factors might be as important in explaining macroevolution in insects and other organisms (Voje et al. 2015) as are environmental abiotic factors such as atmospheric oxygen levels (Clapham and Karr 2012).

Conclusions

We conclude, based on the results of the phylogenetic comparative analyses in this study, that environmental factors (temperature, precipitation, and forest density) may play a secondary role when compared to ecological factors (dispersal, predation, or niche filling) in shaping geographic distributions of body size in odonates. We have argued that neither the temperature-size rule nor a simple version Bergman's rule for insect are sufficient in themselves to fully explain the observed geographic pattern in body size among odonate lineages. Rather surprisingly we found that bird diversity had strong influence on odonate body size distributions even when we controlled for a possible diversity hotspot effect by comparing with mammalian diversity.

The exact mechanisms by which bird diversity shape interspecific geographic variation in odonate body sizes is not yet known. Potential likely explanations are niche filling and avian predation. With more studies on predation and selection in the tropics, it might be possible to better understand why bird diversity can so strongly predict geographic body size patterns in odonates. Combined with the changing latitudinal patterns of body size distributions on the macroevolutionary time scale, the findings in this study are consistent with a scenario whereby odonate taxa evolve large body sizes at lower latitudes, followed by dispersal out of the tropics, possibly partly driven by the emergence of modern birds and associated competition and predation. Large dragonfly clades presumably had a higher dispersal propensity or capacity and were the first to invade the temperate

regions followed by glacial retreats, whereas many of the large damselfly clades have remained in the tropics, resulting in a geographical size-pattern that changes depending on if phylogeny is considered (Fig. 1B) or ignored (Fig. 1A).

References

- Alroy, J., Marshall, C., & Miller, A. (2012). The paleobiology database. <http://paleodb.org>.
- Van der Bijl, W. (2017). phylopath: Perform phylogenetic path analysis.R package version 0.3.1. <https://github.com/Ax3man/phylopath>.
- BirdLife International (2011) <http://www.birdlife.org/>
- Anderson, R. C. (2009). Do dragonflies migrate across the western Indian Ocean? *J. Trop. Ecol.* 25: 347-358.
- Angilletta, Jr, M. J., & Dunham, A. E. (2003). The temperature-size rule in ectotherms: simple evolutionary explanations may not be general. *Am. Nat.* 162: 332-342.
- Ashton, K. G., Tracy, M. C., & Queiroz, A. D. (2000). Is Bergmann's rule valid for mammals? *Am. Nat.* 156: 390-415.
- Kuchta, S. R., & Svensson, E. I. (2014). Predator-mediated natural selection on the wings of the damselfly *Calopteryx splendens*: differences in selection among trait types. *Am. Nat.* 184: 91-109.
- Atkinson, D. (1996). Ectotherm life-history responses to developmental temperature. *Animals and temperature: Phenotypic and evolutionary adaptation.*
- Baird, I. R. (2014). Larval burrow morphology and groundwater dependence in a mire-dwelling dragonfly, *Petalura gigantea* (Odonata: Petaluridae). *Int. J. Odonatol.* 17: 101-121.
- Beaulieu, J. M., J. Oliver, C., & O'Meara, B. (2017) corHMM: Analysis of Binary Character Evolution. R package version 1.22 .<https://CRAN.R-project.org/package=corHMM>.
- Baker, J., Meade, A., Pagel, M., & Venditti, C. (2015). Adaptive evolution toward larger size in mammals. *Proc. Natl. Acad. Sci., USA* 112: 5093-5098.
- Bhagwat, S. A., & Willis, K. J. (2008). Species persistence in northerly glacial refugia of Europe: a matter of chance or biogeographical traits? *J. Biogeogr.* 35: 464-482.
- Bivand, R. S., Pebesma, E. J., Gomez-Rubio, V., & Pebesma, E. J. (2008). *Applied spatial data analysis with R* (Vol. 747248717). New York: Springer.
- Blackburn, T. M., & Gaston, K. J. (1994). Animal body size distributions: patterns, mechanisms and implications. *Trends Ecol. Evol.* 9: 471-474.
- Blackburn, T. M., Gaston, K. J., & Loder, N. (1999). Geographic gradients in body size: a clarification of Bergmann's rule. *Div. Distrib.* 5: 165-174.
- Blanckenhorn, W. U. (2000). The evolution of body size: what keeps organisms small? *Q. Rev. Biol.* 75: 385-407.

- Bujan, J., Yanoviak, S. P., & Kaspari, M. (2016). Desiccation resistance in tropical insects: causes and mechanisms underlying variability in a Panama ant community. *Ecol. Evol.* 6: 6282-6291.
- Chapman, J. W., Reynolds, D. R., & Wilson, K. (2015). Long-range seasonal migration in insects: mechanisms, evolutionary drivers and ecological consequences. *Ecol. Lett.* 18: 287-302.
- Clapham, M. E., & Karr, J. A. (2012). Environmental and biotic controls on the evolutionary history of insect body size. *Proc. Natl. Acad. Sci., USA* 109: 10927-10930.
- Clauset, A., & Erwin, D. H. (2008). The evolution and distribution of species body size. *Science* 321: 399-401.
- Damuth, J. (1981). Population density and body size in mammals. *Nature* 290: 699-700.
- Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.*
- Freestone, A. L., Ruiz, G. M., & Torchin, M. E. (2013). Stronger biotic resistance in tropics relative to temperate zone: effects of predation on marine invasion dynamics. *Ecology* 94: 1370-1377.
- Glasspool, I. J., & Scott, A. C. (2010). Phanerozoic concentrations of atmospheric oxygen reconstructed from sedimentary charcoal. *Nature Geosci.* 3: 627-630.
- Gotthard, K., Berger, D., & Walters, R. (2007). What keeps insects small? Time limitation during oviposition reduces the fecundity benefit of female size in a butterfly. *Am. Nat.* 169: 768-779.
- Harmon, L. J., Losos, J. B., Jonathan Davies, T., Gillespie, R. G., Gittleman, J. L., Bryan Jennings, W., Kozak, K. H., McPeek, M.A., Moreno-Roark, F., Near, T.J., & Purvis, A. (2010). Early bursts of body size and shape evolution are rare in comparative data. *Evolution* 64: 2385-2396.
- Harrison, J. P., Gheeraert, N., Tsigelnitskiy, D., & Cockell, C. S. (2013). The limits for life under multiple extremes. *Trends Microbiol.* 21: 204-212.
- Henry, J. R., & Harrison, J. F. (2014). Effects of body size on the oxygen sensitivity of dragonfly flight. *J. Exp. Biol.* 217: 3447-3456.
- Hijmans, R. J., van Etten, J., Cheng, J., Mattiuzzi, M., Sumner, M., Greenberg, J. A., ... (2016). Package ‘raster’. *R package. https://cran.r-project.org/web/packages/raster/index.html* (accessed 1 October 2016).
- Jenkins, C. N., Pimm, S. L., & Joppa, L. N. (2013). Global patterns of terrestrial vertebrate diversity and conservation. *Proc. Natl. Acad. Sci., USA* 110: E2602-E2610.
- Johansson, F. (2003). Latitudinal shifts in body size of *Enallagma cyathigerum* (Odonata). *J. Biogeogr.* 30: 29-34.
- Khabbazian, M., Kriebel, R., Rohe, K., & Ané, C. (2016). Fast and accurate detection of evolutionary shifts in Ornstein–Uhlenbeck models. *Methods Ecol. Evol.* 7: 811-824.
- Kingsolver, J. G., & Huey, R. B. (2008). Size, temperature, and fitness: three rules. *Evol. Ecol. Res.* 10: 251-268.

- Marron, M. T., Markow, T. A., Kain, K. J., & Gibbs, A. G. (2003). Effects of starvation and desiccation on energy metabolism in desert and mesic *Drosophila*. *J. Insect Physiol.* 49: 261-270.
- Mattila, T. M., & Bokma, F. (2008). Extant mammal body masses suggest punctuated equilibrium. *Proc. R. Soc. Lond. B.* 275: 2195-2199.
- McCue, M. D. (2010). Starvation physiology: reviewing the different strategies animals use to survive a common challenge. *Comp. Biochem. Physiol. A: Mol. Integr. Physiol.* 156: 1-18.
- McPeek, M. A. and J. M. Brown. 2000. Building a regional species pool: diversification of the *Enallagma* damselflies in Eastern North America. *Ecology* 421: 904-920.
- McPeek, M. A., Schrot, A. K., & Brown, J. M. (1996). Adaptation to predators in a new community: swimming performance and predator avoidance in damselflies. *Ecology* 77: 617-629.
- Mousseau, T. A. (1997). Ectotherms follow the converse to Bergmann's rule. *Evolution* 51: 630-632.
- Nilsson-Örtman, V., Stoks, R., De Block, M. & Johansson, F. (2012). Generalists and specialists along a latitudinal transect: patterns of thermal adaptation in six species of damselflies. *Ecology* 93: 1340-1352.
- Olalla-Tárraga, M. Á., Rodríguez, M. Á. & Hawkins, B. A. (2006). Broad-scale patterns of body size in squamate reptiles of Europe and North America. *J. Biogeogr.* 33: 781-793.
- Outomuro, D. & Johansson, F. (2015). Bird predation selects for wing shape and coloration in a damselfly. *J. Evol. Biol.* 28: 791-799.
- Padian, K. & Chiappe, L. M. (1998). The origin and early evolution of birds. *Biol. Rev.* 73: 1-42.
- Paradis, E., Claude, J., & Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20: 289-290.
- Partridge, L., & Coyne, J. A. (1997). Bergmann's rule in ectotherms: is it adaptive? *Evolution* 51: 632-635.
- Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci. Discuss.* 4: 439-473.
- Rainford, J. L., Hofreiter, M., & Mayhew, P. J. (2016). Phylogenetic analyses suggest that diversification and body size evolution are independent in insects. *BMC Evol. Biol.* 16: 8-30.
- Rantala, M. J., Honkavaara, J., Dunn, D. W., & Suhonen, J. (2011). Predation selects for increased immune function in male damselflies, *Calopteryx splendens*. *Proc. R. Soc. Lond. B.* 278: 1231-1238.
- Roslin, T., Hardwick, B., Novotny, V., Petry, W. K., Andrew, N. R., Asmus, A., Barrio I. C., Basset Y., Boesing A. L., Bonebrake, T. C., & Cameron, E. K. (2017). Higher predation risk for insect prey at low latitudes and elevations. *Science* 356: 742-744.
- Sandhall, A. (1987). Trollsländor i Europa. *Stenströms Bokförlag (in Swedish)*.

- Schemske, D. W., Mittelbach, G. G., Cornell, H. V., Sobel, J. M., & Roy, K. (2009). Is there a latitudinal gradient in the importance of biotic interactions? *Annu. Rev. Ecol. Evol. Syst.* 40: 245-269.
- Schlüter, D. (2000). The Ecology of Adaptive Radiation. Oxford University Press, Oxford.
- Schorr, M., and D. Paulson. (2015). World Odonata List. Electronic Database accessible at <http://www.pugetsound.edu/academics/academic-resources/slater-museum/biodiversity-resouces/dragon flies/world-odonata-list2/>. Accessed 11 (2015)
- Sexton, J. O., Urban, D. L., Donohue, M. J., & Song, C. (2013). Long-term land cover dynamics by multi-temporal classification across the Landsat-5 record. *Remote Sens. Environ.* 128: 246-258.
- Shelomi, M. (2012). Where are we now? Bergmann's rule sensu lato in insects. *Am. Nat.* 180: 511-519.
- Siepielski, A. M., K. L. Hung, E. E. B. Bein & M. A. McPeek. 2010. Experimental evidence for neutral community dynamics governing an insect assemblage. *Ecology* 91: 847-857.
- Sokolovska, N., Rowe, L., & Johansson, F. (2000). Fitness and body size in mature odonates. *Ecol. Entomol.* 25: 239-248.
- Stillwell, R. C. (2010). Are latitudinal clines in body size adaptive? *Oikos* 119: 1387-1390.
- Stillwell, R. C., Moya-Laraño, J., & Fox, C. W. (2008). Selection does not favor larger body size at lower temperature in a seed-feeding beetle. *Evolution* 62: 2534-2544.
- Stoks, R., & Córdoba-Aguilar, A. (2012). Evolutionary ecology of Odonata: a complex life cycle perspective. *Ann. Rev. Entomol.* 57: 249-265.
- Svensson, E. I. (2012). Non-ecological speciation, niche conservatism and thermal adaptation: how are they connected? *Org. Div. Evol.* 12: 229-240.
- Svensson, E. I., & Friberg, M. (2007). Selective predation on wing morphology in sympatric damselflies. *Am. Nat.* 170: 101-112.
- Svensson, E. I., Gomez-Llano, M., Rivas Torres, A. & Bensch, H. M. (2018). Frequency-dependence and ecological drift shape coexistence of species with similar niches. *Am. Nat.* (in press).
- Troast, D., Suhling, F., Jinguiji, H., Sahlén, G., & Ware, J. (2016). A global population genetic study of *Pantala flavescens*. *PLOS ONE*: 11(3): e0148949.
- Uyeda, J. C., Hansen, T. F., Arnold, S. J., & Pienaar, J. (2011). The million-year wait for macroevolutionary bursts. *Proc. Natl. Acad. Sci., USA* 108: 15908-15913.
- Van Voorhies, W. A. (1997). On the adaptive nature of Bergmann size clines: a reply to Mousseau, Partridge and Coyne. *Evolution* 51: 635-640.
- Verberk, W. C., & Atkinson, D. (2013). Why polar gigantism and Palaeozoic gigantism are not equivalent: effects of oxygen and temperature on the body size of ectotherms. *Funct. Ecol.* 27: 1275-1285.
- Voje, K. L., Holen, Ø. H., Liow, L. H. & Stenseth, N. C. (2015). The role of biotic forces in driving macroevolution: beyond the Red Queen. *Proc. R. Soc. Lond. B.* 282: 20150186; DOI: 10.1098/rspb.2015.0186

- Waller, J. T. & Svensson, E. I. (2017). Body size evolution in an old insect order: No evidence for Cope's Rule in spite of fitness benefits of large size. *Evolution*. 71:2178-93.
- Wellenreuther, M., K. W. Larson & E. I. Svensson. (2012). Climatic niche divergence or conservatism? Environmental niches and range limits in ecologically similar damselflies. *Ecology* 93: 1353-1366.
- Wikelski, M., Moskowitz, D., Adelman, J. S., Cochran, J., Wilcove, D. S., & May, M. L. (2006). Simple rules guide dragonfly migration. *Biol. Lett.* 2: 325-329.

Paper VI



1 **Limited genetic variation for male mating success reveals low
2 evolutionary potential for thermal plasticity in *Drosophila*
3 *melanogaster***

4

5 John T. Waller[†] (corresponding author) john.waller@biol.lu.se +46 070 488 7577

6 Anna Kell[†]

7 Mireia Ballesta[†]

8 Aude Giraud[†]

9 Jessica K. Abbott[†]

10 Erik I. Svensson[†]

11 [†]Evolutionary Ecology Unit, Department of Biology, Lund University, SE-223 62 Lund,
12 Sweden

13 Short Title: Limited genetic variation in thermal plasticity

14 Acknowledgements: We are grateful to anonymous referees for constructive criticisms earlier
15 drafts of this manuscript. Funding for this study has been provided by the Swedish Research
16 Council (VR; grant no. 621-2012-3768) and Gyllenstiernska Kräpparupstiftelsen (grant no.
17 2014-0032) to E.I. S.

18

19 Data is archived on <https://datadryad.org/>

20

21

22

23

24

25

26

27

28 **Abstract**

29 Populations respond to novel environmental challenges either through genetic changes,
30 through adaptive phenotypic plasticity for the traits in question, or by a combination of these
31 factors. Here, we investigated the evolutionary potential of phenotypic plasticity for male
32 mating success, locomotory ability, and heating rate (a physiological performance trait) in the
33 fruitfly *Drosophila melanogaster*, using isogenic male lines from the Drosophila Reference
34 Genome Panel (DGRP) and hemi-clonal males. We quantified thermal reaction norms of how
35 male mating success changed in relation to a temperate gradient, ranging from cold (18 °C)
36 via optimal (24 °C) to hot and stressful environments (either 30 °C or 36 °C). We found
37 significant differences in male mating success and locomotory performance between different
38 lines, as well as significant main effects of temperature, but no significant genotype-by-
39 environment interactions (GEI:s). A statistical power analysis revealed that the variance
40 explained by GEI:s for thermal plasticity using this sample size is likely to be modest or very
41 small, and represent only 4% of the total variation in male mating success. The lack of strong
42 GEI:s for these two behavioral traits contrast with the presence of significant GEI:s for male
43 heating rate, as measured by thermal imaging (infrared camera technology). These results
44 suggest that sexual selection through male mating success is not likely to be efficient in
45 mediating evolutionary rescue through changed plasticity in response to changing
46 temperatures.

47 **Keywords:**

48 additive genetic variation, fruit fly, infrared camera, sexual selection, insects, thermal
49 imaging, thermal performance, thermal plasticity

50

51

52

53

54 **Introduction**

55 Populations can respond to environmental challenges either evolutionarily through changes in
56 allele frequencies or through adjustments in phenotypic plasticity (Schlichting and Pigliucci
57 1998). Phenotypic plasticity is the capacity of a single genotype to change its phenotype under
58 different environmental conditions (Bradshaw 1965; Roff 1997; Pigliucci 2001). Phenotypic
59 plasticity can increase a population's mean fitness across several environments, and plasticity
60 might increase niche space and geographical range (Ayrinhac et al. 2004; Manenti *et al.* 2015;
61 Mather and Schmidt 2017). Responding plastically to changing environmental conditions is,
62 however, a short-term survival strategy for a population, as there are costs and limits to
63 plasticity (Lande 2014; Murren et al. 2015; Sgro et al. 2016), which can limit the potential to
64 respond to sustained environmental change (Gonzalez et al. 2013).

65

66 Adaptive evolution of phenotypic plasticity requires not only that a population responds to
67 changing environmental conditions, but the different genotypes must also differ in how they
68 respond to these changing environmental conditions (Lande 2009; 2014; Chevin et al. 2010).
69 A population must thus harbor enough standing genetic variation in environmental reaction
70 norms to respond to sustained selection pressures driven by environmental change and
71 thereby evolve adaptive phenotypic plasticity (Schlichting and Pigliucci 1998; Gonzalez et al.
72 2013). The presence of genetic variation in phenotypic plasticity is recorded by the presence
73 of significant genotype-by-environment interactions (GEI:s)(Schlichting and Pigliucci 1998;
74 Lande 2009). If a population lacks such genetic variation in reaction norm slopes,
75 microevolutionary changes in plasticity will be prevented (Sisodia & Singh 2010; Husby *et al.*
76 2010).

77

78 Hansen and Houle (2004) argued that the vast majority phenotypic traits have large amounts
79 additive genetic variation, even when such traits show evidence of long term stasis in the
80 fossil record or in extant populations. However, and in contrast to this view, there are some
81 documented empirical examples where genetic variation for physiological traits has been
82 demonstrated to be low enough to act as an evolutionary limit (Blows & Hoffmann 2005). If a
83 population is invariant in its evolutionary response to sustained environmental change, any
84 change outside of a critical range should rapidly lead to population decline and ultimately
85 extinction (Charmantier *et al.* 2008; Visser 2008; Chevin *et al.* 2010).

86

87 Thermal plasticity is a form of phenotypic plasticity that is particularly important for
88 ectotherms, which have limited ability to buffer themselves against external temperature
89 changes (Angilletta *et al.* 2002; Angilletta 2009). Many ectotherm species might already be
90 close to their upper physiological thermal limits (Addo-Bediako *et al.* 2000; Deutsch *et al.*
91 2008; Kellermann *et al.* 2012). In particular, small insects and other ectotherms may lack the
92 ability to buffer themselves against external temperatures altogether (Stevenson 1985). This
93 might be reflected as canalization (i.e. low genetic variance in thermal reaction norms)
94 resulting in low thermal plasticity (Angilletta *et al.* 2002; Charmantier *et al.* 2008). In general,
95 we know relatively little about the amount of genetic variation in thermal plasticity in natural
96 populations. Moreover, most previous studies on thermal adaptation and thermal plasticity in
97 insects and other ectotherms focus on how temperature affects survival and hence the
98 implications for natural selection. In contrast, the consequences of temperature challenges for
99 sexual selection (e.g. how mating rates and mating success is affected by temperature and
100 thermal plasticity) has seldom been a focus for empirical investigations (see Olsson *et al.*
101 2011 and Taylor *et al.* 2015 for exceptions). Only recently has temperature also been linked
102 to several aspects of sexual selection and speciation in ectotherms. Examples of such recent

103 studies exploring the link between thermal adaptation and sexual selection include how
104 melanin-based dark colouration affects body temperatures within local populations and across
105 latitudinal clines (Punzalan et al. 2008; Svensson and Waller 2013), how different
106 microclimatic environments reduces immigrant male viability (Gosden et al. 2015), how
107 mating rates might be temperature-dependent (Olsson et al. 2011; Taylor et al. 2015) and a
108 recent finding that postzygotic isolation evolves more rapidly between various species of
109 *Drosophila* in hot tropical areas, compared to cooler temperate areas (Yukilevich 2013).

110

111 Isogenic *Drosophila melanogaster* lines offer an excellent opportunity to investigate genetic
112 variation in thermal reaction norms. The Drosophila Genetic Reference Panel (DGRP) is a
113 public resource consisting of more than 200 inbred lines derived from a population in Raleigh,
114 North Carolina (Mackay et al. 2012). Since the lines are isogenic, any phenotypic differences
115 between these lines can be attributed to genetic effects, provided that they are raised and kept
116 under identical conditions. Here we investigate and quantify the amount of genetic variation
117 in thermal reaction norms of 30 DGRP lines, using a hemiclonal experimental approach
118 (Abbott & Morrow 2011), and also by comparing these different DGRP-lines directly with
119 each other. These DGRP lines should be representative sample and a snapshot of naturally
120 segregating genetic variation in the local populations from which they were derived (Mackay
121 et al. 2012) and have also been previously used them in a study on sexual selection on wing
122 interference patterns (WIPs)(Katayama et al. 2014).

123

124 A necessary condition for the evolution of adaptive phenotypic plasticity in a novel thermal
125 environment is that the population harbors enough standing genetic variation in the trait of
126 interest (Chevin et al. 2010). A relatively low amount of genetic variation would indicate that

127 the population is unlikely to respond evolutionarily to these novel thermal conditions. Here
128 we investigate how two behavioral traits important to male fitness – male locomotion and
129 mating success – are influenced by varying thermal conditions and whether these two fitness-
130 related traits show any evidence for genetic variation in plasticity. We used male locomotor
131 activity as a measure of performance as this trait is likely to be associated with fitness because
132 of its links with reproductive success, dispersal, predator avoidance, and foraging (Gilchrist,
133 1996; Roberts et al., 2003; Long & Rice, 2007; Latimer et al. 2011). We complemented these
134 analyses of behavioral performance traits with an analysis of a physiological trait – heating
135 rate – using the technique of thermal imaging (“infrared camera”) on a subset of these DGRP-
136 lines. Heating rate is also likely to covary with physiological performance and mating success,
137 especially in ectotherms. Thermal imaging is a technique by which body temperatures of both
138 endotherms and ectotherms can be quantified, under laboratory, semi-natural, and natural
139 field conditions (Tattersall et al 2009; Tattersall and Cadena 2010; Symonds and Tattersall
140 2010; Svensson and Waller 2013).

141

142 Methods

143 *Drosophila melanogaster* sources

144 Isogenic lines used in this experiment were obtained from the *Drosophila melanogaster*
145 Genetic Reference Panel (DGRP) of the Bloomington Stock Centre (Mackay et al. 2012),
146 which were created after 20 generations of full sibling inbreeding of the stock inbred fly
147 populations (Mackay et al. 2012). Wild type (LHm) flies were originally obtained from
148 Edward H. Morrow (EHM), University of Sussex, Falmer, UK, and maintained in Lund since
149 2012 in the laboratory of J. Abbott. These LHm flies originated from 400 flies collected by L

150 Harshman in central California in 1991 and they have been maintained since that time by L
151 Harshman (1991–1995), WR Rice (1995–2004) and EHM (2004–present) (Carter et al. 2009).

152

153 **Producing hemiclonal males from DGRP-lines**

154 A total of 16 male DGRP flies for each isogenic line were crossed with 16 wildtype virgin
155 LHm females (in a total of 992 vials). The outbred male offspring were then used for the
156 mating and locomotion assays. We refer to these outbred male flies as hemi-clones, following
157 previous terminology (Abbott and Morrow 2011). This outbreeding procedure was conducted
158 to reduce any inbreeding effects on mating behaviour that could potentially remain among the
159 DGRP-lines (Huang et al. 2012). Moreover, by comparing male mating success in the DGRP
160 background vs. mating success in a hemi-clonal background, we were also able to reduce the
161 effects of non-heritable genetic variation across the different genetic backgrounds. Additive
162 genetic variance in male mating success is expected to produce a significant correlation in
163 male mating success between the DGRP-lines and the corresponding male genotype in the
164 hemi-clonal background. Conversely, a weak correlation between male mating success in
165 these different genetic backgrounds would imply low additive genetic variance for this trait
166 and might also indicate a large non-additive effects, arising from e.g. epistatic genetic
167 variance (Meffert et al. 2002) or dominance variation (Merilä & Sheldon 1999). All flies were
168 cultured in a 20mm medium of cornmeal, yeast, and molasses and kept at 25°C in an
169 incubator, on a light-dark cycle of 12hrs:12hrs.

170 *DGRP-lines and hemiclonal experiments*

171 Two complementary groups of flies were used: 1) 32 pure isogenic DGRP lines, and 2) 30
172 hemi-clonal DGRP lines (Mackay et al. 2012). Throughout this paper, we refer to these
173 separate experimental fly categories as DGRP-lines and hemiclonal lines, respectively. For

174 the 32 pure DGRP-lines, the mating assay was performed in three test temperatures: 18°C
175 (cold), 24°C (optimal) and 30°C (hot). We performed three replicates for each line and
176 temperature treatments for these pure DGRP-lines for a total of 265 vials.

177 In a first pilot study using hemiclonal males, we selected 10 DGRP lines for outbreeding to
178 produce 10 groups of hemiclonal males. These 10 DGRP lines were specifically selected for
179 outbreeding because they showed variable patterns in their mating rates to different
180 temperatures in the initial mating assay with the 32 pure DGRP-lines. In general, however, the
181 individual patterns we observed in the pure DGRP lines were only weakly related to the
182 patterns in the outbred lines (Fig. 4) (Huang et al. 2012). Hemi-clones were produced by
183 mating 992 vials of 16 DGRP males with 16 virgin females from the outbred LH_M population
184 (Chippindale et al. 2001). Male offspring from these crosses will therefore have one set of
185 DGRP autosomes and the Y in a random LH_M background. For three test temperatures, 18°C
186 (cold), 24°C (optimal) and 30°C (hot), we performed both mating and locomotion assays.

187

188 We followed up the first pilot hemiclonal study with a second study, where we used 30
189 isogenic lines (DGRP) to produce a new set of 30 hemiclonal groups of males. In this follow-
190 up study, we performed mating and locomotion assays at 18°C (cold), 24°C (optimal), 30°C
191 (hot) and we also added one additional temperature treatment at 36°C (extremely hot) (Trotta
192 *et al.* 2006). This temperature range is similar to what has been used in other studies (Latimer
193 *et al.* 2011). We performed six replicates for each line and temperature treatment in this assay,
194 although the total number of replicates will be slightly greater for those lines in which the
195 pilot study was also included. We pooled the results from first hemiclonal study with the
196 second to increase statistical power, while accounting for the effect of experimental sessions
197 as a block in our statistical analyses.

198 **Mating assays**

199 Each line was anaesthetised with CO₂ gas. Seven males per vial were collected and placed in
200 separate 25 x 95 mm vials (with fly medium). Seven LHm virgin females were also collected
201 per vial, and placed in separate vials. Vials of males and females were kept at 25°C overnight,
202 to allow recovery from anaesthesia. For the experiment, vials containing hemiclonal males
203 (one vial per line), and vials of LHm females were placed in an incubator at the test
204 temperature and allowed to acclimatise for 30 minutes. After this time, the flies from one
205 female vial were transferred without anaesthesia (“flipped”) into a male vial. Vials with males
206 and females were shaken lightly to avoid early mating while the other vials were being
207 combined. The vials, now with 14 flies in total, were placed back in the incubator at the test
208 temperature for 1 hour as the mating assay was conducted. One vial for each of our clonal
209 lines (hemi or pure) was placed in the incubator at a time. For each of the test temperatures,
210 the number of copulating pairs was used a measure of male mating success. The number of
211 mating males in each vial could thus vary from 0 to 7, and it was recorded every 10 minutes
212 (or 15 minutes for the pure lines) over a one hour period. Here, we are measuring a mating
213 rate, which also captures any variation in the latency to mate. All observations were recorded
214 each day between 9.00 and 17.00hrs. We randomized the time of day for mating observations
215 within each line.

216 **Locomotion assay**

217 After 30 minute period of incubation, and before the flies from the male and female vials were
218 combined, the male vials were tapped to cause all the males to fall to the bottom of the food
219 vial. We then recorded the time required for the fastest male fly to walk up the side of the vial
220 from the bottom of the food to the top of the vial plug (95 mm). This was repeated three times
221 for each line and we took the average value per vial (Gibert *et al.* 2001).

222 **Heating rate assay**

223 Using a thermal imaging camera (NEC Avio Infrared Technologies H2640) and macro lens
224 (NEC Avio Infrared Technologies TH92-486), we recorded the heating rate of 10 DGRP lines
225 (see supplement for information about the specific DGRP lines used). Fly individuals from
226 each line were cooled in a climate chamber at 5°C for 3 minutes before being removed from
227 the chamber and allowed to heat up to room temperature (approximately 23°C) in a petri dish.
228 Thermal images were taken every 5 seconds for around 30 seconds or when all the flies had
229 left the petri dish. Two experimental blocks of each line were performed. We recorded how
230 body temperature changed for each line over this time period by analyzing these thermal
231 images using the software provided by NEC Avio (see Svensson and Waller 2013 for more
232 methodological details).

233 **Statistical analyses**

234 All statistical analysis in this paper were conducted in R (R Development Core Team 2008).
235 R-code for all the analyses are uploaded to Dryad, and details are provided in the
236 Supplementary Material. To analyze the heating rates of 10 chosen DGRP lines, we
237 performed an analysis of variance with temperature at each time point as the dependent
238 variable. The following model was used: Temp = Time + Line + Block + Time*Line +
239 Time*Block + Temp*Line*Block.

240

241 For the hemi-clones, we performed a two-way analysis of variance (ANOVA) with the
242 number of matings at 10 minutes as the dependent variable, with line, temperature, and their
243 interactions as dependent factors. In this analysis, line and temperature were both treated as
244 categorical factors. In a follow-up analysis of these hemi-clonal males, we instead treated
245 temperature as continuous variable, both as a simple linear term and as a quadratic term and

246 their two-way interactions with line (i.e. mating rate = Temp + Line + Temp*Line + Temp²
247 + Line*Temp²). We chose to analyse temperature as both a continuous and categorical
248 variable because both analyses have useful interpretative value. In the analysis of the pure
249 DGRP-lines, we performed a similar two-way analysis of variance (ANOVA) with the
250 number of matings at 15 minutes as the dependent variable and line, temperature and their
251 interaction as independent dependent factors. In this analysis, line and temperature were also
252 both treated as categorical factors.

253

254 We additionally performed a repeated measures analysis on the hemiclonal lines using the R-
255 package ‘nlme’ (Pinheiro et al. 2016). We used the number of matings as the dependent
256 variable, and experimental temperature category (18, 24, 30, 36 °C as different levels), line
257 (the 30 hemi-clonal lines as different levels), and time (10, 20, 30, 40, 50, 60 minutes) were
258 treated as fixed effects in this model. Vial and experimental block were treated as nested a
259 random effects. This allowed us to control for the non-independence of the repeated mating
260 counts over each vial.

261 The two-way interaction between line and temperature treatment in these tests, should reflect
262 the magnitude and possible statistical significance of genotype-by-environment interaction
263 with respect to thermal plasticity for male mating rate. Significant line-by-temperature
264 interactions would thus indicate the presence of genetic variation in the thermal reaction
265 norms of males belonging to different DGRP- and hemi-clones respond in terms of their
266 mating success at different temperatures.

267 Finally, we performed a power analysis simulation to quantify the minimum amount of
268 genetic variation in thermal reaction norms that we would be able to detect an effect in the 30
269 hemi-clonal lines, and using a given sample size (R-code for this simulation will be provided

270 on Dryad). To analysis variation in heating rate, we performed a two-way analysis of variance
271 (ANOVA) with the average temperature at each time point as the dependent variable and line,
272 time, and their interactions as dependent factors, while also controlling for a block effect. An
273 analysis modelling mating rate as a proportion is presented in the Supplementary Material
274 (Table S1).

275

276 **Results**

277 We found significant variation in mating rates and locomotory performance among lines,
278 using both the hemi-clonal and the pure DGRP lines (Tables 1-4, S1, S2, S3; Fig. 1). As
279 expected, all lines of both hemi-clonal and pure DGRP experimental categories responded
280 plastically to the different temperature treatments (Fig. 2). Male mating rates and locomotory
281 performance were significantly affected by temperature (Fig 2).

282

283 For all of our statistical analyses, and for neither the hemi-clonal nor the pure DGRP assays,
284 we did not find any statistically significant interaction between line and temperature, that
285 would be indicative of GEI:s (Fig. 1; Tables 1-3, S1, S2, S3). Moreover, we were not able to
286 find any evidence for a statistically significant interaction between line and the squared
287 temperature component (Table 3), i.e. neither the slopes nor the curvatures of the thermal
288 response curves differed significantly between lines (Fig. 1).

289

290 Visual inspection of the thermal performance curves for male mating success in Fig. 1
291 revealed that 24 out of the 30 hemi-clonal lines peaked at intermediate temperatures (24 °C or
292 30 °C), whereas only 6 lines had maximal mating success in the cold (18 °C) or extremely hot

293 treatments (36 °C). For the DGRP-lines, 17 peaked in male mating success at intermediate
294 temperature (24 °C), 9 at cold temperature (18 °C) and the remaining four at 30 °C (note that
295 the DGRP-lines were not evaluated under the extreme treatment (36 °C; see Fig. 1). However,
296 note that the relationship between male mating success and temperature was often quite flat in
297 the range between 18 and 24 °C, after which it dropped (Fig. 1). Taken together, these results
298 suggest a genetically quite invariant intermediate temperature optimum and more or less flat
299 fitness peak around 24 °C (Fig. 2). Thus, we found no evidence for any statistically significant
300 difference between genotypes in the location of this fitness optimum and neither any evidence
301 for different slopes of the thermal response curves or their curvatures (Table 3).

302

303 Using our statistical power simulation we were able to put a minimum bound on the genetic
304 variation in thermal plasticity (variation in the effect of our Line x Temp interaction) (Fig. 3).
305 The main conclusion from these simulations is that our statistical power is high enough under
306 realistic parameter values, meaning that we should have detected a large GEI:s if they had
307 existed (Fig. 3). The magnitude of the GEI we recorded in these experiments can explain at
308 most 4% of variation in mating rate. For instance, our statistical power approaches one with
309 an effect size of 0.03 (treating temperature as a continuous variable) and 0.15 (treating
310 temperature as a categorical variable)(Fig. 3). Finally, at 18° and 30° C there was no
311 detectable correlations between the mating rate of the hemi-clones and our DGRP lines. At
312 24°C, there was a slightly stronger and significant correlation between the mating rates of the
313 DGRP and hemi-clonal males (Fig. 4).

314

315 In the analysis of heating rates using thermal imaging, we found significant variation among
316 the 10 DGRP-lines we investigated, and a significant Line x Time interaction (Fig. 5; Table

317 5). This shows that for these 10 DGRP-lines, the slopes of the thermal reaction norms
318 differed, although the effect was significant only in the first experimental block (Fig. 5; Table
319 5).

320

321 **Discussion**

322 Evolutionary change requires genetic variation in the traits under selection (Blows &
323 Hoffmann 2005). This basic requirement for evolutionary change also applies to a trait like
324 thermal plasticity, and the evolution of thermal plasticity will require genetic variation in the
325 slopes of thermal reaction norms. In this study, we have found significant genetic variation in
326 male mating rates, using two different approaches: a pure-clonal approach (using a subset of
327 un-manipulated DGRP-lines bred by Mackay et al. 2012) and a hemi-clonal approach (Abbott
328 and Morrow 2011), where these DGRP-lines were introduced into an outbred LH_M
329 background (Fig. 1). Using these two complementary experimental approaches, we found no
330 statistically significant variation in the reaction norm of male mating rate to temperature in the
331 different lines (Figs. 1, 3). This shows that although all these lines altered their mating rates in
332 relation to temperature (i.e. phenotypic plasticity; Fig. 1, Table 1-4) the changes were parallel
333 and all lines responded in a similar manner. This implies that genetic variation in thermal
334 plasticity for male mating success is low (Fig. 1), explaining at most 4 % of the variation (Fig.
335 3). Similar conclusions apply to male locomotory performance (Fig. 2), where we also did not
336 find any evidence for significant GEI (Table 4).

337

338 The lack of strong GEI:s for these two behavioural traits contrasts our findings of a significant
339 GEI for male heating rate, a physiological performance trait, measured using thermal imaging
340 (Fig. 5), where we did find evidence for GEI (Table 5). Heating rate is an important fitness

341 trait for ectotherms because it might covary with ability to adjust to natural ambient
342 temperatures. It also measures the capacity of an individual to buffer itself against the
343 ambient environmental temperatures. The use of thermal imaging is a very powerful tool to
344 quantify genetic and phenotypic variation of a physiological trait like heating rate, as done in
345 previous studies of non-model organisms (Tattersall et al 2009; Tattersall and Cadena 2010;
346 Symonds and Tattersall 2010; Svensson and Waller 2013). The fact that the significant GEI
347 for thermal reaction norm slopes in these heating rates have no counterpart in the behavioural
348 assays of mating rates and locomotory assays (Figs. 1-2; Tables 1-4) might be biologically
349 important. Behavioural traits and life-history traits are further downstream than the
350 physiological traits are from the genes that govern phenotypic traits (Price and Schluter 1991),
351 hence genetic variation on these grounds expected to be lower for such higher-level traits,
352 such as mating rate and locomotory performance.

353

354 Overall, and across all lines, male mating success was maximal at 24 °C (DGRP-lines) or at
355 either 24 °C or 30 °C (Fig. 2), with a few exceptions (Fig. 1). This suggests that the thermal
356 optimum for male mating success falls well within the normal temperature range *Drosophila*
357 *melanogaster* will experience in North Carolina in the wild (Annual high temperature: 21.5°,
358 Annual low temperature: 9.3°, Average temperature: 14.9° C (Daly 2000)), where these
359 DGRP-lines originated (Mackay et al. 2012). Thus, although our results suggests limited
360 evolutionary potential for thermal plasticity with respect to male mating success, they are
361 consistent with males being locally adapted with respect to their local temperature regime,
362 consistent with a previous study on *Drosophila melanogaster* (Dolgin et al. 2006, Latimer et
363 al. 2011).

364 These DGRP lines were derived from field-caught flies and variation among these lines
365 should reflect naturally segregating genetic variation in the source population (Mackay *et al.*
366 2012). Using a statistical power simulation (Fig. 3), we were able to put a minimum bound on
367 the amount of variation in reaction norms in our clonal lines. The line effect had a standard
368 deviation of < 0.5, which implies that variation in mating rate was at a maximum less than
369 0.01 matings for any given temperature. Comparing the slopes of the reaction norms from
370 optimal temperature condition (24° C) to extreme temperature condition (36° C), we conclude
371 that the standard deviation in slopes is likely < 0.01. This effect covers only about 4% of the
372 total variation in male mating success (Line = 11%, Temp = 58%, Block = 21%, Line:Temp =
373 4%, Residual = 5%). This means that lines varied less than around ±0.008 matings (in 10 min)
374 per 1° C (Fig. 3). Whether this low amount of variation in thermal reaction norms would allow
375 for evolution of adaptive phenotypic plasticity is an open question and depends on several
376 other ecological and evolutionary factors, including population size, the strength of selection,
377 the rate of environmental change, generation time, and intrinsic rate of increase (Hoffmann
378 2010; Chevin *et al.* 2010). However, we note that the effect sizes in these power calculations
379 are minimum effect sizes, and the true amount of genetic variation in thermal plasticity might
380 be considerably lower.

381
382 One concern is that mating rate is the product of the behaviour of multiple individuals
383 interacting, so it is not only the male's behaviour that matters, but also the female's
384 preference. For instance, LHM females prefer males from some of the DGRP-lines more than
385 males from other DGRP-lines. A second concern is that it is the additive genetic variation that
386 determines the evolutionary potential of a population, yet the hemi-clones still share the
387 same half-genome, so any epistatic effects arising from interactions between chromosomes
388 within the DGRP half will also be included. Our experimental design is for these reasons

389 conservative with respect to our ability to detect significant GEI:s, since the line-effects will
390 partly also include non-additive effects. Additionally, there might exist variation in latency to
391 mate after a disturbance between lines, and this might be of some concern. However, any
392 differences in willingness to mate after disturbance will be captured by the line effect in our
393 statistical analyses. This means that lines might differ not only mating rate per se but also in
394 their willingness to mate at each temperature. This variation in willingness to mate after a
395 disturbance would not interfere with detecting GEI:s in mating rate, unless there was an
396 interaction between the latency to mate and temperature treatment.

397

398 For the hemi-clonal lines, epistatic effects would not be included by the line factor, since our
399 starting iso-genetic lines (30) were homozygous at all loci. Thus, epistatic interactions with
400 the outbred (LHm) genetic background would not be included in the line effect, and would
401 become part of the error variance. Such epistatic genetic variance would represent hidden
402 genetic variation that we were not able to detect with our hemi-clonal lines (Huang *et al.*
403 2012; Mackay 2014). Epistatic variance of traits related to mating success, such as courtship,
404 have been demonstrated for other species of flies (Meffert *et al.* 2002). In these previous
405 studies it has been shown that such epistatic variance can be converted to additive genetic
406 variance following population bottlenecks (Meffert *et al.* 2002). From a theoretical viewpoint,
407 traits that are closely related to fitness such as male mating success (a major fitness
408 component) are expected to show low additive genetic variance, due to the depleting effects
409 of strong directional sexual selection (Rowe & Houle 1996). Directional sexual selection
410 should therefore expect reduce the additive genetic variance fraction for male mating success,
411 resulting in a relatively higher fraction of the remaining genetic variation being non-additive,
412 reflecting either epistatic (Meffert *et al.* 2002) or dominance variance (Merilä & Sheldon
413 2002). Such non-additive genetic variance for male mating success could potentially explain

414 the low concordance between male mating success in the DGRP-lines and the hemi-clonal
415 lines (Fig. 4). This interpretation of high epistatic variance for male mating success in these
416 DGRP-lines would be consistent with previous studies of these DGRP-lines, where high
417 epistatic variance was found for a number of other fitness-related traits, including cold
418 tolerance (measured as chill coma recovery) (Huang et al. 2012).

419

420 Our results have some implications for the prospects of evolutionary rescue through the
421 evolution of adaptive phenotypic plasticity (Chevin et al. 2013) and by sexual selection
422 (Candolin & Heuschele 2008). Previous laboratory experiments on several species of
423 *Drosophila* (Holland 2002; Rundle et al. 2006) and seed beetles *Callosobruchus maculatus*
424 (Martinossi-Allibert et al. 2016) have found at most weak or at best mixed support for sexual
425 selection improving the rate of evolution of local adaptation to novel stressful environments,
426 such as thermally challenging environments. The results in the present study add to this small
427 but growing body of literature, and indicate that thermal plasticity for male mating success is
428 unlikely to evolve, as the genetic variation in thermal reaction norms is limited (Fig. 3). Some
429 theoretical models suggest that sexual selection could improve environmental adaptation at
430 low demographic costs, due to purging of deleterious alleles in males (Agrawal 2001; Siller
431 2001; Whitlock & Agrawal 2009). However, the results in this and other experimental studies
432 (cited above) give only weak support to these models. Thus, our results provide only limited
433 support to the hypothesis that sexual selection could act as an evolutionary rescuer of
434 populations experiencing rapid environmental change (Candolin & Heuschele 2008).

435

436 Our results also agree with other research in this area showing that evolutionary responses to
437 novel and challenging thermal conditions may be constrained (Bennett & Lenski 1993;

438 Kellermann *et al.* 2009; Mitchell & Hoffmann 2010; Hoffmann 2010; Kelly *et al.* 2012; Kelly
439 *et al.* 2013; Kristensen *et al.* 2015). In particular, ectotherms, particularly those living in
440 tropical areas that already experience temperatures close to their upper thermal tolerance
441 limits might have a reduced capacity to adapt to higher temperatures via evolutionary means
442 (Araujo *et al.* 2013; Kristensen *et al.* 2015) (Fig 2). Insights from studies of niche
443 conservatism also suggests that it might be difficult for many species to evolve new
444 physiological limits (Parsons 1982; Kimura & Beppu 1993; Wiens & Graham 2005;
445 Kellermann *et al.* 2009; Hoffmann 2010). For example, the invasive species *Drosophila*
446 *subobscura* has a range in the Americas that is restricted to climates that are similar to those
447 found in its native range in Europe (Prevosti *et al.* 1988; Gilchrist *et al.* 2008; Huey & Pascual
448 2009; Hoffmann 2010). However, some natural *Drosophila* populations have adapted to their
449 local climates (Trotta *et al.* 2006). For instance, warm-adapted populations in India have
450 higher levels of desiccation resistance and melanism compared with cold-adapted populations
451 (Parkash *et al.* 2008). Similarly, populations in Africa are more viable at warm temperatures
452 than temperate populations (Bouletrau-Merle *et al.* 2003; David *et al.* 2004; David *et al.*
453 2005; Hoffmann 2010). Other laboratory studies have shown that populations might have the
454 capacity to adapt to temperature via evolutionary adaptation (Hoffmann 2010; Latimer *et al.*
455 2011; Mukuka *et al.* 2010; Hoffmann *et al.* 2013; van Heerwaarden and Sgro 2013;
456 Blackburn *et al.* 2014; van Heerwaarden *et al.* 2016). For example, in some laboratory
457 experiments, *Drosophila* have been successfully selected for increased survival after cold and
458 heat shocks (Bubliy & Loeschke 2005). Finally, it might be that the expression of additive
459 genetic variation for heat tolerance is contingent on developmental temperature (van
460 Heerwaarden *et al.* 2016). For instance, two rainforest *Drosophila* species exhibited
461 significant levels of additive genetic variation when raised in warmer environments, but not
462 when raised at 25°C (van Heerwaarden *et al.* 2016).

463 **Conclusions**

464 The capacity of natural populations of *Drosophila melanogaster* to adapt evolutionarily (i.e.
465 via changing the genetic composition of the population) to novel thermal conditions through
466 the evolution of adaptive phenotypic plasticity remains an open question. In this study, we
467 were not able to detect any significant genetic variation in thermal reaction norms to different
468 thermal environments in either the DGRP-lines or the hemi-clonal lines. However, we were
469 able to put a lower bound on the amount of genetic variance in thermal plasticity. Isogenic
470 lines such as these DGRP-lines in combination with hemi-clonal approaches offer
471 opportunities to quantify genetic variation in phenotypes where the underlying genotypes are
472 known and line-differences can attributed to genetic differences without the need for complex
473 breeding designs (Abbott & Morrow 2011). Such isogenic lines could also be raised in
474 different temperatures in the future, which might reveal that significant genetic variation
475 exists, but this variation might depend on temperatures experienced during the larval
476 development stage (van Heerwaarden et al. 2016). It is possible that GEI:s might be more
477 pronounced among juveniles, and that strong selection during these earlier life stages might
478 have reduced additive genetic variance in thermal reaction norms that could be detected
479 during the adult stage (cf. Martinossi-Allibert et al. 2016). In the present study, we were not
480 able to detect any significant genetic variation in thermal plasticity in adult male mating
481 success or locomotion.

482 **Figure Legends**

483

484 **Fig 1.** Thermal reaction norms of male mating success estimated for each pure DGRP
485 (black) and hemi-clonal (empty) DGRP-line in relation to experimental temperature
486 treatments. Numbers above each sub-panel refer to the DGRP-lines described in Mackay et al.

487 (2012). Although temperature affects mating rates in all lines and often peaks at intermediate
488 temperatures, there is no evidence for any genotype-by-environment interaction with respect
489 to temperature, i.e. the mating success of the different lines are similarly affected across
490 environmental treatments. Mating rate is the number of copulations that were recorded in a
491 vial after 10 minutes (for the hemiclones) and after 15 minutes (for the pure clones). The
492 mating assay was only performed for some lines with either only the hemi-clones (303, 486)
493 or only the pure lines (305, 315). Confidence intervals (95%) are shown as vertical lines.

494

495 **Fig 2.** Temperature-related male performance for the hemi-clonal males showed as male
496 locomotory performance (left panel) and male mating rate (right panel). Males from different
497 lines vary in mean mating rates and locomotory performance (intercepts), but variation
498 between lines in thermal reaction norms (slopes) is low (Fig. 1). Hence, there is no evidence
499 for significant genetic variation in the thermal reaction norms (Fig. 1). Each data point
500 represents the mean performance of each line at each temperature treatment.

501

502 **Fig 3.** Power plots to detect significant GEI:s (reflected as Line*Temp interactions) for male
503 mating success. Temperature is either treated as continuous variable (left plot) or as a discrete
504 experimental factor (right plot). Comparing the slopes of the reaction norm (left panel) from
505 optimal temperature to extreme (hemiclonal lines: 24° to 36° C; pure DGRP-lines: 24° to 30 °
506 C), we found that the standard deviation in slope is likely to be < 0.01 for the hemi-clonal
507 lines and < 0.025 for the pure DGRP-lines. This means that lines varied in less ±0.008
508 matings (per 10 min) per 1° C. Treating temperature as discrete (right panel), the standard
509 deviation of this Line*Temp effect is most likely < 0.07 for the hemi-clonal lines and < 0.12
510 for the pure clonal lines. This means that the variation in mating rate is maximally < 0.01

511 matings for a given temperature. These are minimum effect sizes, and the true variation in
512 thermal plasticity is likely to be lower. R-code for these power simulations are provided on
513 Dryad. Each data point represents 1000 simulations (hemi-clonal: n = 7 and pure-clonal: n = 3
514 replicates per line n = 30).

515

516 **Fig. 4.** Correlations between mating rates (at 10 min and 15 between our hemi-clonal and
517 DGRP lines, measured at three different temperatures (measured at 18 °C (“cold” = white), 24
518 °C (“optimal” = gray) and 30 °C (“hot” = black). We studied a total of 28 Lines per
519 temperature treatment, i.e. the total number of datapoints in this graph is 84). The “random”
520 Lhm genetic background in our hemi-clonal lines likely has a large effect on mating rate,
521 since the mating rates the DGRP-lines, put either in to hemi-clones or measured as pure
522 clones show only weak non-significant relationships (All: r=0.12, P=0.24, N=84, 18C: r
523 =0.08, P = 0.68, N=29; 24C: r= 0.37, P=0.05, N=29; 30C: r= 0.10, P=0.61, N=29). The lack
524 of strong significant correlations between the mating rates of the different lines in different
525 genetic backgrounds indicates low additive genetic variance in male mating success and
526 implies strong epistatic effects on mating rate (Huang et al. 2012), as our hemi-clonal and
527 pure-clonal lines will only share the additive effects of their genotypes.

528

529 **Fig 5. A)** Heating rates of 10 DGRP lines after exposer to a cool shock 5°C for 3 minutes and
530 then allowed to recover at room temperature for 30 seconds. Two experimental blocks were
531 performed of these same 10 DGRP lines. Thermal images were taken every 5 seconds.
532 Heating rate was significantly variable between lines in block 1 but not in block 2 (Table 5).

533

534

535

536

537

538 **Figures and Tables**

539 **Table 1.** Analysis of variance (ANOVA) for the mating rate in hemi-clonal males. Line
540 (N=30) and temperature (N=4; 18,24,30,36 °C) were both treated as categorical factors. We
541 used the first mating observation at 10 minutes to avoid double counting matings at the next
542 observation time 20 minutes. Here, a significant Line x Temp interaction would be indicative
543 of a GEI and reveal significant genetic variation (greater than zero) in plasticity in our hemi-
544 clonal males. Block is a categorical factor (N=2), which controls for differences in the two
545 experimental runs. See table S1 in the supplementary material and a model wherein mating
546 rate is treated as a proportion.

Term	DF	SSE	MSE	F-value	P-value	Significance
Line	29	4.54	0.16	2.37	<0.0001	****
Temp	3	2.48	0.83	12.51	<0.0001	****
Block	1	0.3	0.3	4.52	0.034	*
Line x Temp	87	4.88	0.06	0.85	0.83	ns
Residuals	870	57.38	0.07			

547

548

549

550

551

552

553

554

555

556 **Table 2.** Analysis of variance (ANOVA) for the mating rate of the DGRP males. Line (N=32)
557 and temperature (N=3; 18,24,30 °C) were both treated as categorical factors. We used the first
558 observation time at 15 minutes to avoid double counting matings at the next observation time
559 30 minutes. A significant Line x Temp interaction would be indicative of a GEI.

Term	DF	SSE	MSE	F-value	P-value	Significance
Line	31	5.22	0.17	3.55	<0.0001	****
Temp	2	1.31	0.66	13.82	<0.0001	****
Line x Temp	58	2.35	0.04	0.86	0.752	ns
Residuals	172	8.16	0.05			

560

561

562

563

564

565

566

567

568

569

570

571

572

573 **Table 3.** Analysis of covariance (ANOVA) for the mating rate hemi-clonal males. Here Line
574 (N=30) is treated as a categorical factor and temperature is treated as a continuous variable.
575 The quadratic effect of temperature (Temp^2) allows the model to detect any curvature in the
576 reaction of mating rate to temperature. We used the first observation time at 10 minutes to
577 avoid double counting matings at the next observation time 20 minutes. Here, a significant
578 Line x Temp interaction would be indicative of a GEI, and indicate significant genetic
579 variation (greater than zero) in plasticity in our hemi-clonal males. Block is a categorical
580 factor (N=2), which controls for differences in the two experimental runs.

Term	DF	SSE	MSE	F-value	P-value	Significance
Line	29	4.54	0.16	2.41	<0.0001	****
Temp	1	0.02	0.02	0.34	0.557	ns
Temp ²	1	2.39	2.39	36.81	<0.0001	****
Block	1	0.31	0.31	4.82	0.028	*
Line x Temp	29	2.27	0.08	1.2	0.212	ns
Line x Temp ²	29	1.67	0.06	0.89	0.641	ns
Residuals	900	58.38	0.06			

581

582

583

584

585

586

587

588

589

590

591 **Table 4.** Analysis of locomotion performance in the hemi-clonal lines (N=30). Locomotion
592 is the speed (mm/seconds) of the fastest hemiclonal male (7 per vial) to walk up the side of
593 the vial (repeated 3 times for each line) at the different experimental temperatures. We
594 performed a two-way analysis of variance (ANOVA) with locomotion time the as the
595 dependent variable and line, temperature and their interaction as independent dependent
596 factors.

Term	DF	SSE	MSE	F-value	P-value	Significance
Line	29	4197.69	144.75	1.76	0.009	**
Temp	3	7449.08	2483.03	30.22	<0.0001	****
Line x Temp	87	5691.16	65.42	0.8	0.907	ns
Residuals	592	48635.14	82.15			

597

598

599

600

601

602

603

604

605

606

607

608

609

610

611 **Table 5.** Analysis of variance (ANOVA) for the heating rate of DGRP males. 2 replicates of
612 Lines (N=10). Time is measured in seconds. A thermal image was taken at each time point.
613 Here, a significant Line x Time interaction would be indicative of significant variation in
614 heating rate between lines. However, a significant Time x Line x Block effect complicates
615 this interpretation. Block is a categorical factor (N=2), which controls for differences in the
616 two experimental runs.

Term	DF	SSE	MSE	F-value	P-value	Significance
Time	1	135.14	135.14	290.3	<0.0001	***
Line	9	42.62	4.74	10.17	<0.0001	***
Block	1	35.44	35.44	76.13	<0.0001	***
Time x Line	9	9.46	1.05	2.26	0.024	*
Time x Block	1	3.94	3.94	8.47	0.004	**
Line x Block	9	16.48	1.83	3.93	0.0002	***
Time x Line x Block	9	8.47	0.94	2.02	0.04	*
Residuals	95	44.22	0.47			

617

618

619

620

621

622

623

624

625

626

627

628

629

630

631

632

633

634

635

636

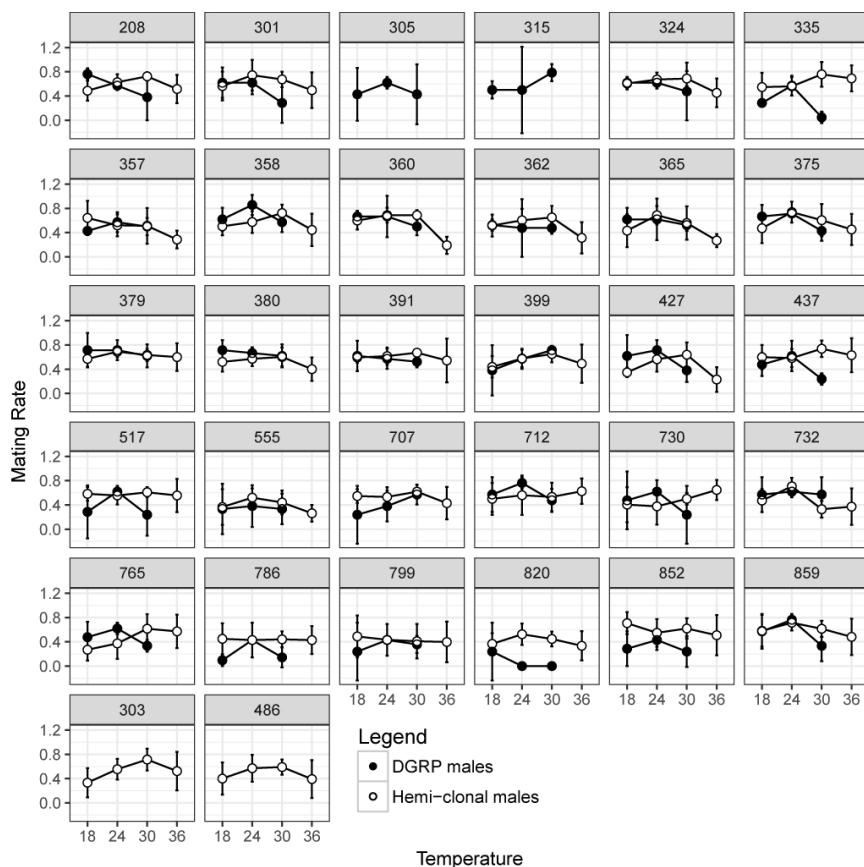
637

638

639

640

641



642

643

644

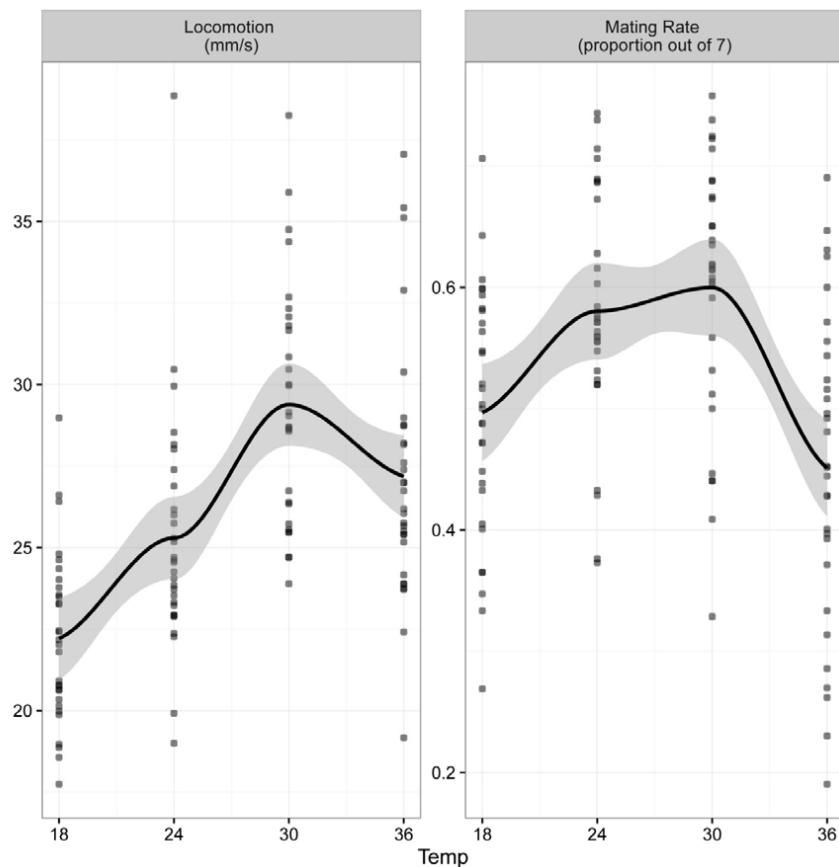
645

646

647

648

649



650

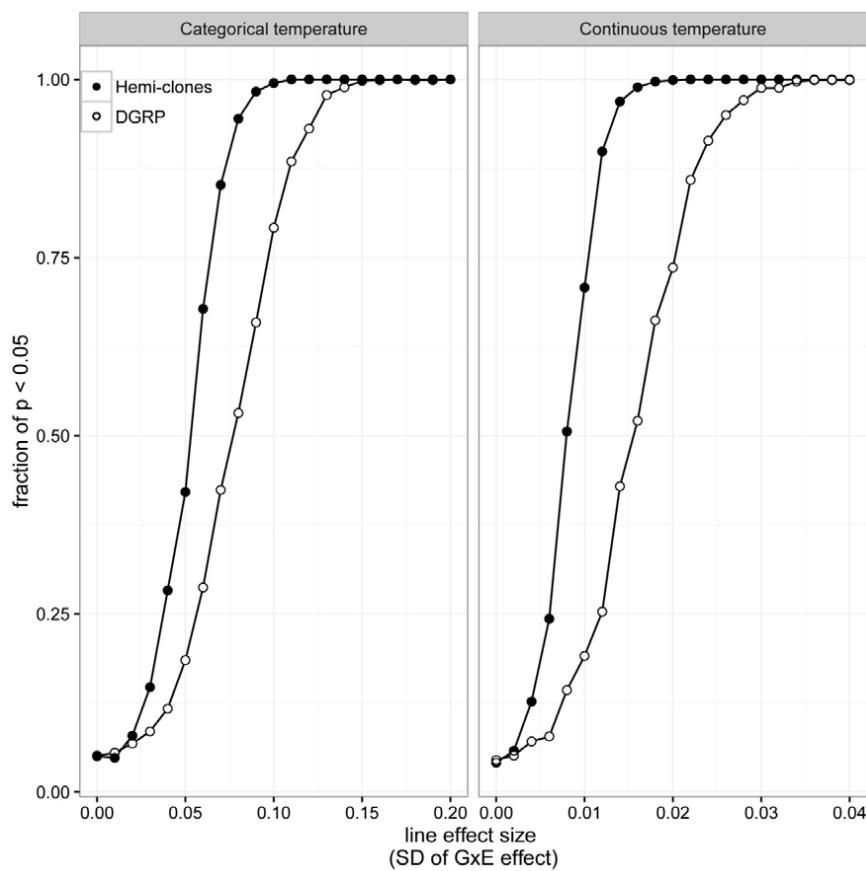
651

652

653

654

655



656

657

658

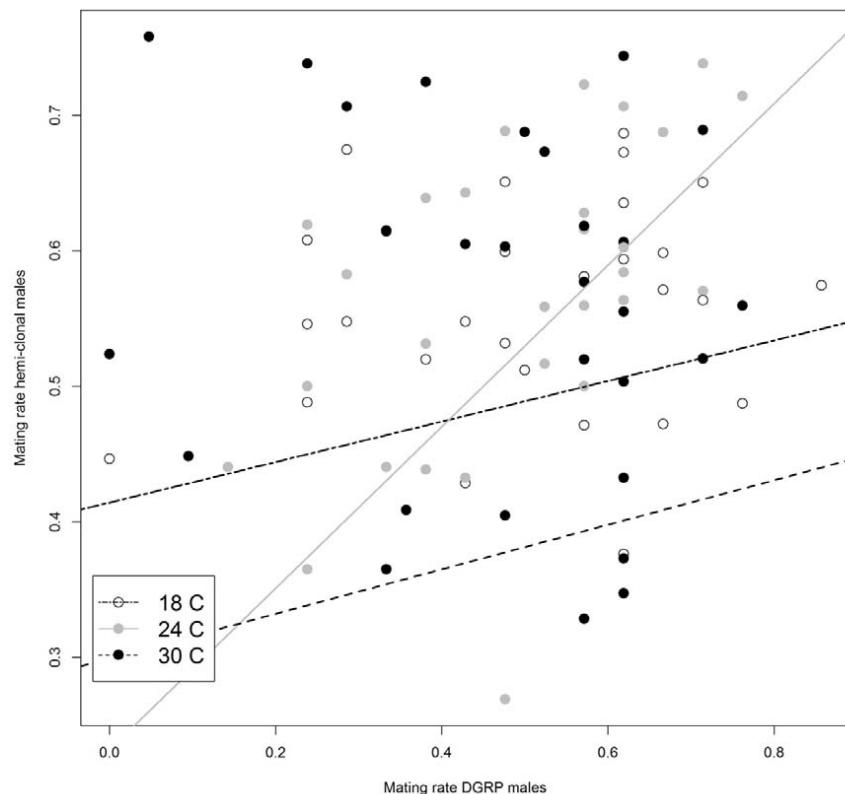
659

660

661

662

663



664

665

666

667

668

669

670

671

672

673

674

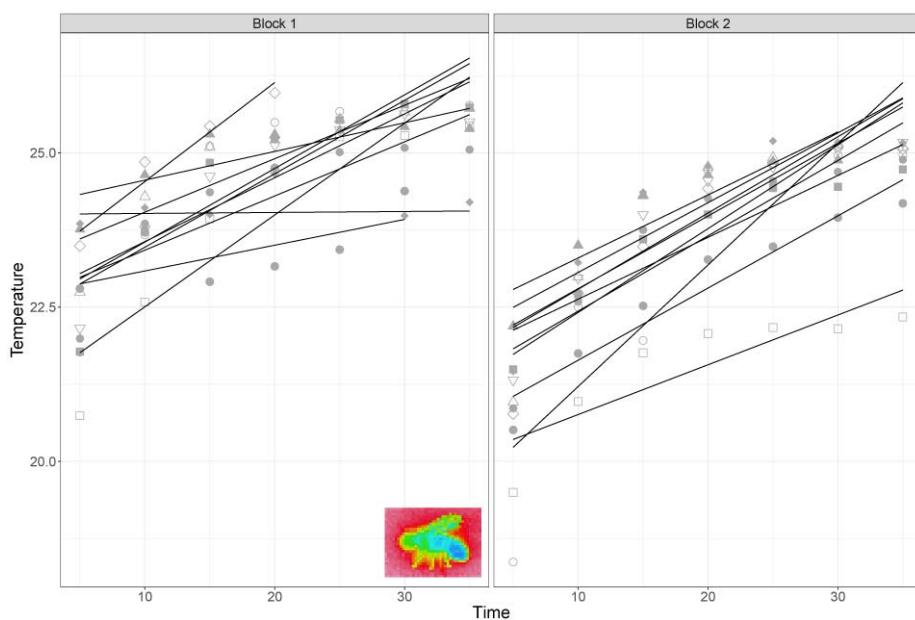
675

676

677

678

679



680

681

682

683

684

685

686

687

688

689 **References**

- 690 Abbott, J. K. and E. H. Morrow. 2011. Obtaining snapshots of genetic variation using hemiconal
691 analysis. *Trends Ecol. Evol.* **26**: 359-368.
- 692 Addo-Bediako, A., S. L. Chown and K. J. Gaston. 2000. Thermal tolerance, climatic variability and
693 latitude. *Proc. R. Soc. Lond., B, Biol. Sci.* **267**: 739-745.
- 694 Agrawal, A.F. 2001. Sexual selection and the maintenance of sexual reproduction. *Nature* **411**: 692-
695 695.
- 696 Angilletta, M. J. 2009. Thermal adaptation: a theoretical and empirical synthesis. Oxford University
697 Press, Oxford.
- 698 Angilletta, M. J., P. H. Niewiarowski and C. A. Navas. 2002. The evolution of thermal physiology in
699 ectotherms. *J. Therm. Biol.* **27**: 249-268.
- 700 Araujo, M. B., F. Ferri-Yanez, F. Bozinovic, P. A. Marquet, F. Valladares and S. L. Chown. 2013.
701 Heat freezes niche evolution. *Ecol. Lett.* **16**: 1206-1219.
- 702 Ayrinhac, A., V. Debat, P. Gibert, A. G. Kister, H. Legout, B. Moreteau, R. Vergilino, and J. R.
703 David. 2004. Cold adaptation in geographical populations of *Drosophila melanogaster*: phenotypic
704 plasticity is more important than genetic variability. *Funct. Ecol.* **5**: 700-706.
- 705 Bennett, A. F. and R. E. Lenski. 1993. Evolutionary adaptation to temperature II. Thermal niches of
706 experimental lines of *Escherichia coli*. *Evolution* **47**: 1-12.
- 707 Blackburn, S., B. van Heerwaarden, V. Kellermann and C. M. Sgro. 2014. Evolutionary capacity of
708 upper thermal limits: beyond single trait assessments. *J. Exp. Biol.* **217**: 1918-1924.
- 709 Blows, M. W. and A. A. Hoffmann. 2005. A reassessment of genetic limits to evolutionary change.
710 *Ecology*, **86**: 1371-1384.
- 711 Bouletreau-Merle, J., P. Fouillet and J. Varaldi. 2003. Divergent strategies in low temperature
712 environment for the sibling species *Drosophila melanogaster* and *D. simulans*: overwintering in
713 extension border areas of France and comparison with African populations. *Evol. Ecol.* **17**: 523-548.
- 714 Bradshaw, A. D. 1965 Evolutionary significance of phenotypic plasticity in plants. *Adv.
715 Genet.* **13**:115-155.
- 716 Bubliy, O. A. and V. Loeschke. 2005. Correlated responses to selection for stress resistance and
717 longevity in a laboratory population of *Drosophila melanogaster*. *J. Evol. Biol.* **18**: 789-803.
- 718 Candolin, U. and J. Heuschele. 2008. Is sexual selection beneficial during adaptation to environmental
719 change? *Trends Ecol. Evol.* **23**: 446-452.

- 720 Carter, A. J. R., T. M. Weier and D. Houle. 2009. The effect of inbreeding on fluctuating asymmetry
721 of wing veins in two laboratory strains of *Drosophila melanogaster*. *Heredity* **102.6**: 563-572.
- 722 Charmantier, A., R. H. McCleery, L. R. Cole, C. Perrins, L. E. B. Kruuk and B. C. Sheldon. 2008.
723 Adaptive phenotypic plasticity in response to climate change in a wild bird population. *Science*, 320:
724 800-803.
- 725 Chevin, L. M., R. Gallet, R. Gomulkiewicz, R.D. Holt, and S. Fellous. 2013. Phenotypic plasticity in
726 evolutionary rescue experiments. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **368**: 89-99.
- 727 Chevin, L. M., R. Lande and G. M. Mace. 2010. Adaptation, plasticity, and extinction in a changing
728 environment: Towards a predictive theory. *PLoS Biol.* **8**: e1000357.
- 729 Chippindale, A. K., R. G. Jonathan and R. R. William. 2001. Negative genetic correlation for adult
730 fitness between sexes reveals ontogenetic conflict in *Drosophila*. *Proc. Natl. Acad. Sci.* **98.4**: 1671-
731 1675.
- 732 Daly, C., G. H. Taylor, W. P. Gibson, T. W. Parzybok, G. L. Johnson and P. A. Pasteris. 2000. High-
733 quality spatial climate data sets for the United States and beyond. *Trans. ASAE* **43**: 1957-1967.
- 734 David, J. R., L. O. Araripe, M. Chakir, H. Legout, B. Lemos, G. Petavy, C. Rohmer, D. Joly and B.
735 Moreteau. 2005. Male sterility at extreme temperatures: a significant but neglected phenomenon for
736 understanding *Drosophila* climatic adaptations. *J. Evol. Biol.* **18**: 838-846.
- 737 David, J. R., R. Allemand, P. Capy, M. Chakir, P. Gibert, G. Petavy and B. Moreteau. 2004.
738 Comparative life histories and ecophysiology of *Drosophila melanogaster* and *D. simulans*. *Genetica*
739 **120**: 151-163.
- 740 Deutsch, C. A., J. J. Tewksbury, R. B. Huey, K. S. Sheldon, C. K. Ghilambor, D. C. Haak and P. R.
741 Martin. 2008. Impacts of climate warming on terrestrial ectotherms across latitude. *Proc. Natl. Acad.
742 Sci.* **105**: 6668-6672.
- 743 Fox J. and S. Weisberg. 2011. An R Companion to Applied Regression. Sage, Thousand Oaks CA.
- 744 Gibert, P., B. Moreteau, G. Petavy, D. Karan and J. R. David. 2001. Chill-coma tolerance, a major
745 climatic adaptation among *Drosophila* species. *Evolution* **55**: 1063-1068.
- 746 Gilchrist, G.W. 1996. A quantitative genetic analysis of thermal sensitivity in the locomotor
747 performance curve of *Aphidius ervi*. *Evolution* **50**:1560–1572.
- 748 Gilchrist, G. W., L. M. Jeffers, B. West, D. G. Folk, J. Suess and R. B. Huey. 2008. Clinal patterns of
749 desiccation and starvation resistance in ancestral and invading populations of *Drosophila subobscura*.
750 *Evol. Appl.* **1**: 513-523.
- 751 Gonzalez, A., O. Ronce, R. Ferriere and M. E. Hochberg. 2013. Evolutionary rescue: an emerging
752 focus at the intersection between ecology and evolution. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **368**:
753 1610-1631.
- 754 Gosden, T.P., J. T. Waller and E. I. Svensson. 2015. Asymmetric isolating barriers between different
755 microclimatic environments caused by low immigrant survival. *Proc. R. Soc. Lond., B, Biol. Sci.* **282**:
756 2410-2459.
- 757 Hansen, T. F. and D. Houle. 2004. Evolvability, stabilizing selection, and the problem of
758 stasis. Phenotypic integration: studying the ecology and evolution of complex phenotypes. Oxford
759 University Press, Oxford, 130-150.

- 760 Hoffmann, A. A. 2010. Physiological climatic limits in *Drosophila*: patterns and implications. *J. Exp. Biol.* **213**: 870-880.
- 762 Hoffmann, A. A., S. L. Chown and S. Clusella-Trullas. 2013. Upper thermal limits in terrestrial
763 ectotherms: how constrained are they? *Funct. Ecol.* **27**: 934-949.
- 764 Holland, B. 2002. Sexual selection fails to promote adaptation to a new environment. *Evolution* **56**:
765 721-730
- 766 Huang, W., S. Richards, M. A. Carbone, D. H. Zhu, R. R. H. Anholt, J. F. Ayroles, et al. 2012.
767 Epistasis dominates the genetic architecture of *Drosophila* quantitative traits. *Proc. Natl. Acad. Sci.*
768 **109**: 15553-15559.
- 769 Huey, R. B. and M. Pascual. 2009. Partial thermoregulatory compensation by a rapidly evolving
770 invasive species along a latitudinal cline. *Ecology* **90**: 1715-1720.
- 771 Husby, A., D. H. Nussey, M. E. Visser, A. J. Wilson, B. C. Sheldon and L. E. B. Kruuk. 2010.
772 Contrasting patterns of phenotypic plasticity in reproductive traits in two great tit (*Parus major*)
773 populations. *Evolution* **64**: 2221-2237.
- 774 Katayama, N., J.K. Abbott, J. Kjaerandsen, Y. Takahashi and E.I. Svensson. 2014. Sexual selection
775 on wing interference patterns in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci.* **111**: 15144-15148.
- 776 Kellermann, V., B. van Heerwaarden, C. M. Sgro and A. A. Hoffmann. 2009. Fundamental
777 evolutionary limits in ecological traits drive *Drosophila* species distributions. *Science* **325**: 1244-1246.
- 778 Kellermann, V., J. Overgaard, A. A. Hoffmann, C. Flojgaard, J. C. Svennning and V. Loeschcke. 2012.
779 Upper thermal limits of *Drosophila* are linked to species distributions and strongly constrained
780 phylogenetically. *Proc. Natl. Acad. Sci.* **109**: 16228-16233.
- 781 Kelly, M. W., E. Sanford and R. K. Grosberg. 2012. Limited potential for adaptation to climate change
782 in a broadly distributed marine crustacean. *Proc. R. Soc. Lond., B, Biol. Sci.*, **279**: 349-356.
- 783 Kelly, M. W., R. K. Grosberg and E. Sanford. 2013. Trade-Offs, geography, and limits to thermal
784 adaptation in a tide pool copepod. *Am. Nat.* **181**: 846-854.
- 785 Kimura, M. T. and K. Beppu. 1993. Climatic adaptations in the *Drosophila immigrans* species group -
786 seasonal migration and thermal tolerance. *Ecol. Entomol.* **18**: 141-149.
- 787 Kristensen, T. N., J. Overgaard, J. Lassen, A. A. Hoffmann and C. Sgro. 2015. Low evolutionary
788 potential for egg-to-adult viability in *Drosophila melanogaster* at high temperatures. *Evolution* **69**:
789 803-814.
- 790 Lande, R. 2009. Adaptation to an extraordinary environment by the evolution of phenotypic plasticity
791 and genetic assimilation. *J. Evol. Biol.* **22**: 1435-1446.
- 792 Lande, R. 2014. Evolution of phenotypic plasticity and environmental tolerance of a labile quantitative
793 character in a fluctuating environment. *J. Evol. Biol.* **27**: 866-875.
- 794 Latimer, C.A.L., Wilson, R.S., and Chenoweth, S.F., 2011. Quantitative genetic variation for thermal
795 performance curves within and among natural populations of *Drosophila serrata*. *J. Evol.*
796 *Biol.*, **24**: 965-975.
- 797 Long, T.A.F., and Rice, W.R. 2007. Adult locomotory activity mediates intralocus sexual conflict in a
798 laboratory-adapted population of *Drosophila melanogaster*. *Proc. Biol. Sci.* **274**: 3105–3112.

- 799 Mackay, T. F. C. 2014. Epistasis and quantitative traits: using model organisms to study gene-gene
800 interactions. *Nat. Rev. Genet.* **15**: 22-33.
- 801 Mackay, T. F. C., S. Richards, E. A. Stone, A. Barbadilla, J. F. Ayroles, D. H. Zhu, et al.. 2012. The
802 *Drosophila melanogaster* Genetic Reference Panel. *Nature* **482**: 173-178.
- 803 Manenti, T., V. Loeschke, N. N. Moghadam and J. G. Sorensen. 2015. Phenotypic plasticity is not
804 affected by experimental evolution in constant, predictable or unpredictable fluctuating thermal
805 environments. *J. Evol. Biol.* **28**: 2078-2087.
- 806 Martinossi-Allibert I., G. Arnqvist and D. Berger. 2016. Sex-specific selection under environmental
807 stress in seed beetles. *J. Evol. Biol.* DOI: 10.1111/jeb.12996.
- 808 Mathur, V. and P. S. Schmidt. 2017. Adaptive patterns of phenotypic plasticity in laboratory and field
809 environments in *Drosophila melanogaster*. *Evolution* **71**: 2: 465-474.
- 810 Meffert, L.M., S.K. Hicks and J.L Regan. 2002. Nonadditive genetic effects in animal behavior. *Am.*
811 *Nat.* **160**:198-213.
- 812 Merilä, J. and B.C. Sheldon. 1999. Genetic architecture of fitness and nonfitness traits: empirical
813 patterns and development of ideas. *Heredity* **83**: 103-109.
- 814 Mitchell, K. A. and A. A. Hoffmann. 2010. Thermal ramping rate influences evolutionary potential
815 and species differences for upper thermal limits in *Drosophila*. *Funct. Ecol.* **24**: 694-700.
- 816 Mukuka, J., O. Strauch, L. Waeyenberge, N. Viaene, M. Moens and R. U. Ehlers. 2010. Heat tolerance
817 among different strains of the entomopathogenic nematode *Heterorhabditis bacteriophora*. *Biocontrol*
818 **55**: 423-434.
- 819 Murren, C.J., Auld, J.R., Callahan, H., Ghalambor, C.K., Handelsman, C.A., Heskel, M.A., et al.
820 2015. Constraints on the evolution of phenotypic plasticity: limits and costs of phenotype and
821 plasticity. *Heredity*, 115:293-301.
- 822 Olsson, M., E. Wapstra, T. Schwartz, T. Madsen, B. Ujvari and T. Uller. 2011. In hot pursuit:
823 fluctuating mating system and sexual selection in sand lizards. *Evolution* **65**:574-583.
- 824 Parkash, R., S. Rajpurohit and S. Ramniwas. 2008. Changes in body melanisation and desiccation
825 resistance in highland vs. lowland populations of *D. melanogaster*. *J. Insect Physiol.* **54**: 1050-1056.
- 826 Parsons, P. A. 1982. Adaptive strategies of colonizing animal species. *Biol. Rev.* **57**: 117-148.
- 827 Pigliucci, M. 2001. Phenotypic plasticity: beyond nature and nurture. John Hopkins University Press,
828 Baltimore.
- 829 Prevosti, A., G. Ribo, L. Serra, M. Aguade, J. Balana, M. Monclús and F. Mestres. 1988. Colonization
830 of America by *Drosophila subobscura* - experiment in natural populations that supports the adaptive
831 role of chromosomal inversion polymorphism. *Proc. Natl. Acad. Sci.* **85**: 5597-5600.
- 832 Price, T. and D. Schlüter. 1991. On the low heritability of life-history traits. *Evolution* **45**: 853-861.
- 833 Punzalan, D., F. H. Rodd and L. Rowe. 2008. Sexual selection mediated by the thermoregulatory
834 effects of male colour pattern in the ambush bug *Phymata americana*. *Proc. R. Soc. Lond., B, Biol.*
835 *Sci.* **275**:483-492.

- 836 R Development Core Team. 2008. R: A language and environment for statistical computing. R
837 Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- 839 Roberts, S.P., Marden, J.H., and Feder, M.E. 2003. Dropping like flies: environmentally induced
840 impairment and protection of locomotor performance in adult *Drosophila melanogaster*. *Physiol.*
841 *Biochem. Zool.* **76**:615–621.
- 842 Roff, D.A. 1997. Evolutionary quantitative genetics. Chapman and Hall, New York.
- 843 Rowe, L. and D. Houle. 1996. The lek paradox and the capture of genetic variance by condition
844 dependent traits. *Proc. R. Soc. Lond., B, Biol. Sci.* **263**: 1415-1421.
- 845 Rundle, H.D., S.F. Chenoweth, and M.W. Blows. 2006. The roles of natural and sexual selection
846 during adaptation to a novel environment. *Evolution* **60**: 2218-2225
- 847 Schlichting, C. D. and M. Pigliucci. 1998. Phenotypic evolution: a reaction norm perspective. Sinauer
848 Associates Incorporated, Sunderland.
- 849 Sgrò, C.M., Terblanche, J.S., and Hoffmann, A.A., 2016. What can plasticity contribute to insect
850 responses to climate change? *Annu. Rev. Entomol.* **61**:433-451.
- 851 Siller, S. 2001. Sexual selection and the maintenance of sex. *Nature* **411**:689-695.
- 852 Sisodia, S. and B. N. Singh. 2010. Resistance to environmental stress in *Drosophila ananassae*:
853 latitudinal variation and adaptation among populations. *J. Evol. Biol.* **23**: 1979-1988.
- 854 Stevenson, R. D. 1985. The relative importance of behavioral and physiological adjustments
855 controlling body temperature in terrestrial ectotherms. *Am. Nat.* **126**: 362-386.
- 856 Svensson, E. I. and J. T. Waller. 2013. Ecology and sexual selection: evolution of wing pigmentation
857 in Calopterygid damselflies in relation to latitude, sexual dimorphism and speciation. *Am. Nat.*
858 **182**:174-195.
- 859 Symonds, M. R. R. and G. J. Tattersall. 2010. Geographic variation in bill size across bird species
860 provides evidence for Allen's rule. *Am. Nat.* **176**:188-197.
- 861 Tattersall, G. J., D. V. Andrade, and A. S. Abe. 2009. Heat exchange from the toucan bill reveals a
862 controllable vascular thermal radiator. *Science* **325**:468-470.
- 863 Tattersall, G. J. and V. Cadena. 2010. Insights into animal temperature adaptations revealed through
864 thermal imaging. *Imaging Sci. J.* **58**:261-268.
- 865 Taylor, M.L., T. A. R. Price, A. Skeats and N. Wedell. 2015. Temperature can shape a cline in
866 polyandry, but only genetic variation can sustain it over time. *Behav. Ecol.* **27**:462-469.
- 867 Trotta, V., F. C. F. Calboli, M. Ziosi, D. Guerra, M. C. Pezzoli, J. R. David and S. Cavicchi. 2006.
868 Thermal plasticity in *Drosophila melanogaster*: A comparison of geographic populations. *BMC Evol.*
869 *Biol.*, **6**: 67.
- 870 Visser, M. E. 2008. Keeping up with a warming world; assessing the rate of adaptation to climate
871 change. *Proc. R. Soc. Lond., B, Biol. Sci.* **275**: 649-659.
- 872 Whitlock, M.C. and A.F. Agrawal. 2009. Purging the genome with sexual selection: reducing mutation
873 load through selection on males. *Evolution* **63**: 569-582.

- 874 Wiens, J. J. and C. H. Graham. 2005. *Niche conservatism: Integrating evolution, ecology, and*
875 *conservation biology*. *Annu. Rev. Ecol. Evol. Syst.* **36**:519-39.
876
877 Yukilevich, R. 2013. Tropics accelerate the evolution of hybrid male sterility in *Drosophila*. *Evolution*
878 **67**:1805-1814.

879
880

881
882

Paper VII



Sexual and natural selection on thermal plasticity and thermal canalization in insects at high latitudes

John T. Waller and Erik I. Svensson

Abstract

The evolution of phenotypic plasticity is of much principal interest, but there are few empirical studies in natural populations demonstrating how natural or sexual selection favour either increased plasticity or its reverse (increased canalization). Insects and other ectotherms are highly influenced by ambient temperatures and thermal plasticity might be beneficial to maintain optimal body temperatures and thereby maximize survival and reproduction. Here, we estimated the strength of natural and sexual selection on thermal plasticity (heating rate) and thermal canalization in two congeneric and sympatric species of damselflies (genus *Caloperyx*). Thermal plasticity is an important trait insects since they must largely rely on sunlight and ambient temperature in order to reach and maintain optimal body temperatures for survival and reproduction, particularly after nights or cold periods. In natural populations of these damselflies, the temperature fitness optimum for survival and mating rate is between 23 °C and 26 °C. We hypothesized that selection on thermal plasticity would increase outside this temperature range. Consistent with this prediction, we found that thermal plasticity was selectively beneficial when temperature conditions fell outside this range. Our results further suggest that when ambient temperatures are close to optimal body temperatures, sexual selection favours thermal canalization and disfavours maladaptive thermal plasticity since individuals might already be at optimal body temperatures most of the time. These results also have important implications for the evolution of thermal traits in relation to climatic factors, as ectotherms in more stable environments, such as the tropics, might experience weak selection on

thermal plasticity, which might affect their ability to respond evolutionarily or adjust plastically to rapidly changing thermal environments.

Introduction

Evolutionary biologists are increasingly directing their focus towards the evolution of phenotypic plasticity in temporally and spatially heterogeneous environments, as exemplified by both empirical studies in natural populations (Charmantier et al. 2008; Vedder et al. 2013) and theoretical models of the evolution of reaction norm slopes and intercepts (Lande 2009; Chevin et al. 2010; Chevin and Lande 2011). Research on the thermal plasticity is not only of interest for increasing our basic understanding of how traits evolve and why, but are also of practical concern in conservation biology, in particular the question if, when and how plasticity can lead to evolutionary rescue of populations (Chevin et al. 2013; Ashander et al. 2016). One form of phenotypic plasticity that is currently subject to much interest is thermal plasticity, or how organisms respond plastically in relation to ambient temperatures (Angilletta 2009). Understanding how genetic adaptation in thermal adaptations evolve is crucial for our understanding of how organisms will respond to increasing global or regional temperatures (Hoffmann et al. 2003; Deutsch et al. 2008; Kellermann et al. 2009; Sinervo et al. 2010; Wiens 2016).

Insects and other small ectotherms are traditionally considered to be thermal conformers because they allow their body temperatures to closely match ambient temperatures (Sanborn 2004; Angilletta 2009). In ectotherms, important physiological processes are at the mercy of ambient temperature. From an enzymatic point of view, being an obligate thermal conformer might not always be beneficial (Huey and Berrigan 2001; Blouin-Demers et al. 2005; Martin et al. 2008). Enzymes are less efficient at any non-optimal temperatures (Martin et al. 2008), and even though ectotherms will tend to match ambient temperatures without immediate mortality, we might still expect better thermoregulatory ability to be selected for in certain contexts, especially in extreme thermal environments (Huey et al. 2009). While ectotherms have less control over their internal body temperatures than endotherms, they typically still have some ability to thermoregulate to some degree via behavioural and/or physiological adjustments (Huey et al. 2003). Variation in thermoregulatory ability typically differs between species as well as between individuals within single populations (Hoffmann et al. 2011; Somero 2010; Kearney et al. 2009; Buckley et al. 2015). Therefore, the dichotomy of being a true thermoregulator versus a thermal conformer is as false one, and even groups considered to be thermal conformers there are fitness

incentives to regulate their internal body temperatures by behavioral means (Huey and Slatkin 1976, Martin et al. 2008).

Ongoing and anticipated future climate change will most likely increase seasonal variation in temperature in many areas, making the ability to thermoregulate increasingly important for ectotherms (Deutsch et al. 2008). However, some temperate ectotherms appear to have maximum thermal tolerances that are 10–20 °C higher than required to withstand the average summer air temperatures where they currently live, which might reflect their evolutionary origin in tropical environments (Sanborn 2004; Sunday et al. 2011). There are thus many behavioral, morphological, and physiological adaptations that might allow ectotherms to survive unfavourable thermal environments outside their current thermal range (May 1979; Heinrich 2013), and in some more extreme cases the body temperatures of some insects have been recorded more than 35°C above ambient temperature (Sanborn 2004).

A large source of heat in flying insects comes from the activity of their flight muscles, and flying insects can increase heat generation through increased muscle activity (May 1979). Basking or changing body orientation to the sun is also another common thermoregulatory behavioral strategy that minimizes the exposed surface area and heat uptake (Porter 1982; Heinrich 2013). Body color such as melanism, provides yet another means by which the amount of radiation that gets turned into heat energy can be increased (True 2003; Trullas 2007; Svensson and Waller 2013). Finally, simply evolving a larger body size will increase the surface-to-volume ratio and thermal inertia (Kingsolver and Huey 2008; Angiletta 2009).

Thermal plasticity is a measure of an individual's ability to change its body temperature in relation to ambient temperature (Trotta et al. 2006; Waller et al. 2017). Thermal plasticity is a specific form of a reaction norm in relation to a specific environmental variable (temperature)(Angiletta 2009). Such reaction norms can be quantified in terms of both their slopes (plasticity) and their intercept (Angiletta 2009). Mechanistically, thermal plasticity can be quantified by estimating the slope parameter of how body temperature changes in relation to ambient temperature (Angiletta 2009). This measure of thermal plasticity captures any physiological differences between individuals, but also might capture any behavioral adjustments such as habitat selection, orientation, and muscle activity.

The rate of evolution in response to selection on thermal plasticity will depend the additive genetic variances and covariances underlying variation of this trait and the strength of selection on these traits (Lynch and Walsh 1998). Some previous studies have demonstrated that there is indeed genetic variation in the slope of thermal reaction norms (Gutteling et al. 2007; van Asch et al. 2007; Winterhalter and Mousseau 2007), whereas others have failed to find any such genetic variation (Waller et al. 2017). Under some environmental conditions, we might expect the evolution of thermal canalization towards an optimal body temperature that is associated with high fitness (Liefting et al. 2009). Conversely, classic theory for

the evolution of phenotypic plasticity suggests that plasticity should be high in variable environments, so that individuals could express their optimum phenotype even if environmental conditions change radically (Lande 2014; Chevin and Hoffmann 2017). How these classic theories for the evolution of phenotypic plasticity in general apply to the more specific case of thermal plasticity is yet largely unknown, as there are few empirical studies in natural populations in this area.

Here, we studied how natural and sexual selection operate on thermal reaction norms in two congeneric European species of damselflies (the banded demoiselle *Calopteryx splendens* (Harris 1780) and the beautiful demoiselle *Calopteryx virgo* Linnaeus 1758). The genus *Calopteryx* did not evolve in Europe, but has a Asiotropical origin, from which they expanded in to temperate regions at high latitudes (Misof et al. 2000; Svensson and Waller 2013). These two species of *Calopteryx* damselflies invaded southern Sweden approximately 12,000 years ago after the last glacial period (Misof et al. 2000; Clark et al. 2009). Their tropical origin and relatively recent establishment in the temperate areas at high latitudes make these damselflies interesting in terms of thermal adaptation and thermal plasticity, as they might potentially not yet be fully adapted to the temperature regimes in the north.

In *Calopteryx*, males are highly territorial (Marden and Waage et al. 1990; Plaistow et al. 1996; Córdoba-Aguilar 2002), and they rely on basking behaviour to reach optimal body temperatures after cool nights (Marden et al. 1996). These damselflies usually spend most of their active period on perches and take flights when feeding or defending mating territories, so maintaining optimal body temperatures is a challenge for them. *Calopteryx* males have large wing patches composed of melanin, and these dark wing patches might also help males to achieve higher body temperatures (Svensson and Waller 2013). Because mating success in these damselflies is largely dependent on a male's ability to hold a territory, males that are able to maintain optimal body temperatures over a wide range of ambient environmental temperatures would be expected to be more successful in obtaining matings.

Here, we define thermal plasticity as the slope of the thermal reaction norm for individuals that warm up after a cold period, such as after a night. If an individual has a high slope (i.e. a high heating rate), it means they are more affected by ambient temperature, but they might also more quickly reach their optimal body temperature after thermally stressful period. Therefore, a high heating rate might be either a fitness advantage or a fitness disadvantage to a male damselfly, depending on environmental conditions. For instance, if the thermal environment is below the thermal optimum of a male damselfly, it might be advantageous to quickly heat up. Conversely, if the ambient temperature is above the thermal optimum, it might be advantageous to heat up more slowly or not at all. In other words, under some conditions there could be selection for thermal canalization

that buffers the organism against the stressful thermal environment (cf. Huey et al. 2003). Finally, if ambient temperatures are on average close or near to the thermal fitness optimum, selection on the thermal reaction norms would be expected to become relaxed or at least weakened, which could potentially result in the accumulation of high phenotypic and genetic variation thermal reaction norms in such environments, such as in relatively stable tropical thermal environments. On the other hand, thermal plasticity could also be expected to increase in variable temperate environments for adaptive reasons, if such plasticity is more beneficial at higher compared to lower latitudes (Mitchell et al. 2011; Lancaster et al. 2015). The latitudinal hypothesis predicts increasing thermal plasticity as one moves from the equator to the poles as seasonality increases. This hypothesis would be supported by observations of either reduced plasticity in organisms in extreme environments or increased plasticity in less extreme environments. For a trait that is initially canalized, a major change in the environment can select for increased plasticity (Gavrillets and Scheiner 1993; Lande 2009; Lancaster et al. 2015).

Here, we show that sexual selection on thermal plasticity overall favors canalization and shallower thermal reaction norm slopes, suggesting that even at these high latitudes, insects can suffer from overheating, at least when temperatures are high. However, the strength of sexual selection for thermal canalization also depends on ambient temperatures and interacts with morphological variation, albeit in different ways in these two phenotypically very similar species.

Methods

Field site and general field work procedures

Our field studies were performed along the river Klingavälsån in the province of Skåne in southern Sweden at Sövdemölla Gård (Latitude: 55.601492, Longitude: 13.657340). These field studies were conducted each summer (June and July) during four seasons (2013 - 2016). In each year, we collected morphological data, fitness data (survival and mating success), and thermal performance measures on two sympatric damselfly species (*Calopteryx splendens* and *Calopteryx virgo*). Field work took place between 10.00 and 15.00 when activity levels were the highest. Damselflies at our field site have very low activity levels when temperature are below 15°C and during rainy and windy days (i.e. they do not fly or mate under these such conditions). Therefore, we only captured, processed, and re-sighted individuals on the relatively warm and calm days when they were

active, whereas on cold and rainy days, such data could not be collected, as animals were not active and could not be found.

Morphological measurements

Males and females of both *C. splendens* and *C. virgo* were captured with handnets and processed in the field. Individual damselflies were measured by placing their wings flat against two layers of transparency paper. Digital scans were taken with a known scale and the body and wings traits were measured in ImageJ (Schneider et al. 2012; Waller and Svensson 2016; 2017). We measured the following morphological traits: 1) forewing length and width; 2) total body length; 3) abdomen length; 4) thorax length; 5) fore- and hindwing melanin patch length (for *C. splendens* males only); 6) fore- and hindwing melanin patch width (for *C. splendens* males only); 7) mass (to the nearest ± 0.002 g) and 8) melanin wing patch darkness based on the mean intensity of a grayscale image of the patch (larger values meaning darker wing patches).

Thermal imaging

A series of thermal images (NEC Avio Infrared Technologies H2640) were taken of each field-caught individual (Fig. 1). Individuals were first allowed to cool for 3 minutes in a cooler box at approximately 5°C before being placed back into ambient temperature, each within an individual cup, under a thermal imaging camera. Thermal images were taken every 10 seconds for 3 minutes (i.e. up to 18 frames per individual, except for those that flew away within 3 minutes). This experiment aimed to simulate how these ectothermic insects heat up with increasing ambient temperature, e.g. after a cold night in the morning hours.

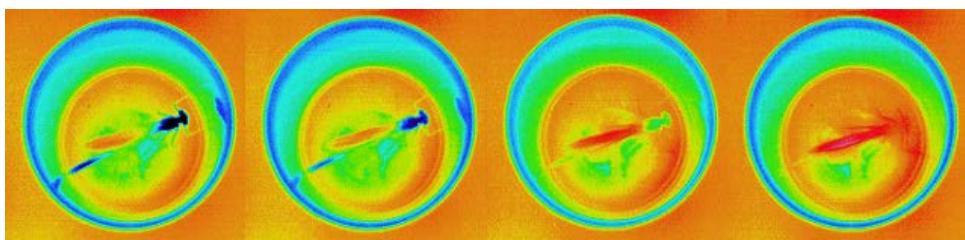


Figure 1. A series of thermal images.

Thermal images of a damselfly heating up over 3 minutes. Warm colors (red, orange) represent higher temperatures and cooler colors (blue, green) represent cooler colors.

Batches of 6 individuals were cooled and imaged simultaneously. After the thermal images were taken, individuals were released. Because individuals were exposed to ambient temperature and sunlight, variability in these two parameters were controlled for before analysis of heating rates (thermal plasticity) of each individual. Images were then analysed to retrieve the body temperature of each individual at each time point.

Estimating slopes for thermal reaction norms

All statistical analyses were done in R unless otherwise stated. We defined thermal plasticity as the slope of the heating rate of individuals after a cooling stress event. This slope was calculated by fitting a linear model of body temperature (obtained from thermal images) against ambient temperature. After this, we fit an additional linear model accounting for daily variation in solar radiation and temperature, and the residuals from this model were used in any downstream analysis of thermal performance. Solar radiation data was taken from the STRÅNG model of solar radiation from the Swedish Meteorological and Hydrological Institute (SMHI) (<http://strang.smhi.se/>). Ambient temperature was extracted from the iButton temperature loggers at various damselfly neighborhoods at our field site (described below).

Measuring thermal microenvironmental variation

We placed approximately 30 iButton thermologgers (1-Wire; DS1921G-F5) at different locations within our field site to measure thermal microenvironmental variation. Such temperature loggers are able to record temperature at set intervals for several months. We recorded ambient temperature every hour for the duration of our field study each year. The field site was divided into 10 neighborhoods (sections; three iButtons per neighborhood) along our study site. Neighborhoods were marked with colored tape, so that when individuals were recaptured, they could always be linked to a particular neighborhood. When ambient temperature is used in our statistical models, we are using temperature data from these temperature loggers.

Longevity and mating success

In conjunction with our recordings of morphological and thermal plasticity data, all individuals were marked with a unique code with three colors on the last three segments of the abdomen. These unique markings were used to identify

individuals and record their subsequent survival (longevity) in the field. Apart from catching and re-capturing single individuals, we also caught as many copulating pairs in the field as possible so that we could obtain estimates of sexual selection through male mating success. We tried to capture as many mating pairs as possible during days with high mating activity, however, we could of course not capture all observed pairs. Nevertheless, we recorded all observed copulations in the field (both those that were captured and those that were not) to quantify mating activity during all days.

Selection on thermal plasticity in relation to ambient temperatures

To examine the factors that influence thermal plasticity (heating rate) in these damselflies we fit a series of mixed-effect models using the R package *nlme*. We estimated the following fixed effects: ambient temperature, solar radiation, body mass, body size, and wing patch density (darkness). All these models contained year as a random factor.

We estimated linear selection gradients of thermal plasticity (slopes of the thermal reaction norms) for both sexual selection (male mating success) and survival selection and related this to natural variation in temperature in the environment. We divided individuals into groups based month and year. We did this in order to capture as much natural variation in temperature as possible while still having reasonable statistical power to detect selection. Abundance data at this field site reveals that there are two different cohorts each summer, due to so-called cohort splitting (Norling 1984), which is common in many species of damselflies. These two seasonal cohorts are roughly divided between June and July. We analyzed the data using generalized linear mixed effect models. We calculated linear selection differentials and quadratic selection gradients (stabilizing and disruptive selection) on slopes of thermal reaction norms and morphological traits following the general statistical methodology of Lande and Arnold (1983). The quadratic regression coefficients were multiplied by two to obtain correct estimates of the strength of stabilizing and disruptive selection (Stinchcombe et al. 2008). All of our phenotypic traits were standardized to mean zero and unit variance and our two fitness components (minimum lifespan or male mating success) were divided by the population average (mean fitness) before being used as the response variables in these regression analyses of selection (Lande and Arnold 1983).

In the sexual selection analyses, we used a binary response variable as our measure of male mating success (males that obtained at least one mating = 1; non-mated males = 0). From our daily field counts and captures at our study site, we found 311 males mating, out of a total of 2945 observed ($311/2945 = 0.105$). Based on these numbers, we therefore conservatively assumed that only 10% of

males in this population successfully mated at least once, which we thus used as our measure of average male mating fitness (i.e. 0.10) in our regression analyses of selection (cf. Lande and Arnold 1983). This quite low mating success estimate of 10% males obtaining at least one copulation at this site is consistent with similar estimates from our previous independent selection studies at two other sites (“Klingavälsåns Naturreservat” and “Höje Å, Värpinge”), where we collected similar data from 2001-2003 (Svensson et al. 2004; 2006). This low mating success reflects the reproductive biology of *Calopteryx* damselflies, which are characterized by high mating skew and strong sexual selection. We continued to monitor the population until all of our marked individuals had died. The average lifespan of an adult *C. splendens* males in the field is around 3-4 days, although some males can live for more than two weeks (Svensson et al. 2006; Waller and Svensson 2016).

Effect of ambient temperature on longevity and mating rate

To quantify the general effect of temperature on average lifespan of both species, we estimated the relationship between individual longevity variation and the average daytime temperature experienced by each male individual. The lifespan temperature of each individual male was calculated by taking the average temperature of experienced by each individual in its given microenvironment. We performed this analysis for males only, as the recapture rates for females were very low. We also used the average temperature and the number of mating pairs each day in the field to quantify the population-level effect of temperature on mating activity. Linear models with quadratic terms were then fit to the datasets of both *C. splendens* and *C. virgo*.

Results

Effect of temperature on lifespan and mating rate

To assess the general sensitivity of *C. splendens* and *C. virgo* individuals to temperature, we modeled survival and mating rate as a function of temperature (Fig. 2; Table 1). In both *C. splendens* and *C. virgo* mating activity peaked somewhere between 20 – 25°C, whereas higher or lower temperatures reduced the population level mating activity (Table 1; Fig. 2; A,B). Although plots of mating rate of the two species appear to be different (Fig. 2 A,B), there is no statistical difference between the peak mating rates of the two species (Table 1:

Temperature² x Species (ref. *C. virgo*), $P = 0.897$). Survivorship in *C. splendens* males peaked at temperatures between 20 – 25°C, whereas higher or lower temperatures reduced male lifespan in the field (Fig. 1; C). Similarly, *C. virgo* males had reduced lifespan at high and low temperatures, with a tendency for a somewhat increased mortality with increasing temperatures, compared to *C. splendens* which seemed less heat sensitive in terms of survival (Table 1: Temperature² x Species; $P = 0.087$).

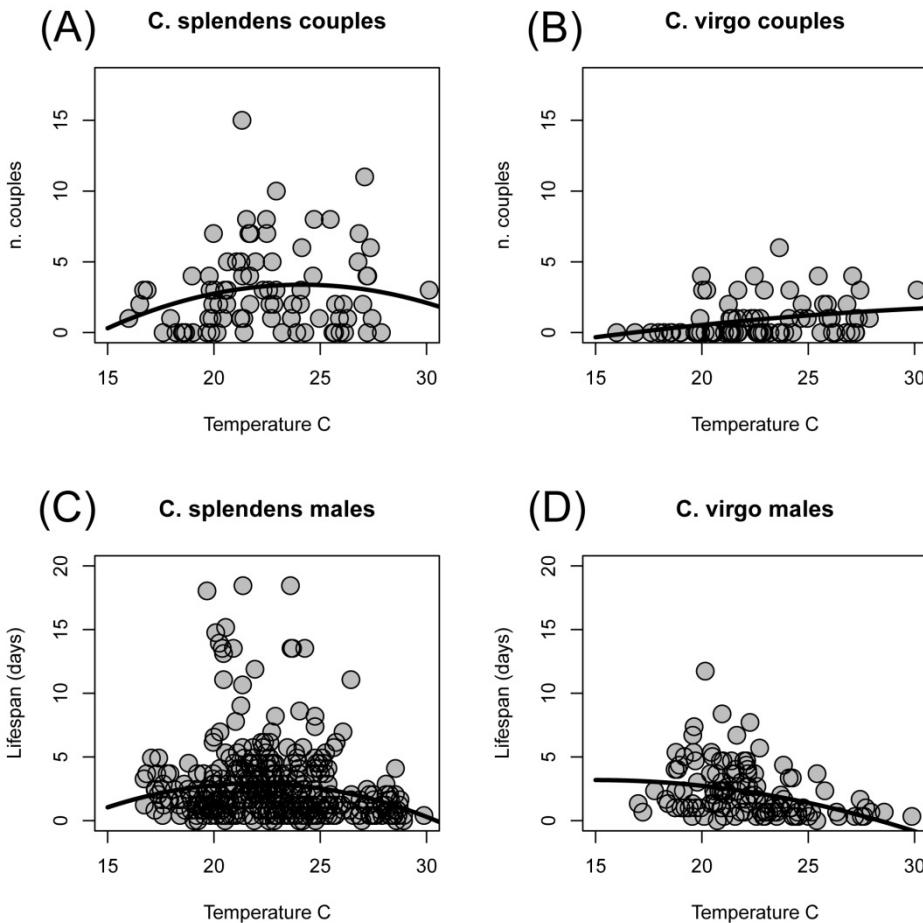


Figure 2. Effect of ambient temperature on survival and mating rate.

The effect of ambient temperature on mating activity and lifespan in *C. splendens* and *C. virgo* at the sympatric site at Klingavälsåns near Södremölla. In *C. splendens*, mating rate peaks between 20 – 25°C, but declines at higher or lower temperatures. Mating rate for *C. virgo* individuals, may increase at higher temperatures (B). Survival in *C. splendens* males peaked at temperatures between 20 – 25°C, whereas higher or lower temperatures reduces lifespan in the field (C). Similarly, *C. virgo* males had low lifespan at higher or lower temperatures, with the peak at lower temperature than *C. splendens* (D). See Table 1 for statistics.

Tabel 1. Modeling mating rate and survival with temperature.

Modeling mating rate and survival in our population of *C. virgo* and *C. splendens*. Both daily mating rate and survival were modeled as with a negative binomial generalized linear model assuming an over-dispersed count outcome. The intercept term was estimated, but is omitted from the table for brevity. Mating was taken as the total number of matings each day and Temperature in this model was mean ambient temperature of that day between 10:00-15:00 hours. Species is a two level factor of both species. Survival is the number of days observed in the field of males in the population. temperature in this model is the mean temperature over the lifespan of each individual male. Significant effects are shown in **bold**.

Fitness	Term	Estimate	SE	P-value	N
Mating	Species (ref. <i>C. virgo</i>)	-5.945	11.777	0.613	163
	Temperature	0.981	0.470	0.037*	
	Temperature ²	-0.020	0.010	0.049*	
	Temperature x Species (ref. <i>C. virgo</i>)	0.268	1.013	0.791	
	Temperature ² x Species (ref. <i>C. virgo</i>)	-0.002	0.021	0.897	
Survival	Species (ref. <i>C. virgo</i>)	9.987	4.333	0.021*	957
	Temperature	0.555	0.128	<0.001***	
	Temperature ²	-0.013	0.002	<0.001***	
	Temperature x Species (ref. <i>C. virgo</i>)	-0.768	0.379	0.043*	
	Temperature ² x Species (ref. <i>C. virgo</i>)	0.014	0.008	0.087	

Factors influencing thermal plasticity

To examine what morphological and environmental factors influence intercepts and slopes of the thermal reaction norms (thermal plasticity = slopes of thermal reaction norms) in *C. splendens* males, we constructed a series of linear mixed-effect models using all of the data from 2013 to 2016, while controlling for the random effect of year (Table S1). These models incorporated two environmental factors (ambient temperature and solar radiation) and two morphological traits (body mass and wing color patch darkness). The best performing model was a model that included that included ambient temperature, solar radiation, and body mass (Table S1; Model 4). A model that also included wing patch darkness was the second best performing model in terms of AIC (Table 1; Model 6), with a small but positive effect of wing patch darkness on thermal reaction norm slope (Table 1).

Sexual selection on the slopes of thermal reaction norms

We modelled the effect of the morphological traits and thermal variables (slope) on sexual selection in *C. spendens* and *C. virgo* males (Tables 2-3). Data from 2013 – 2016 was used, while we controlled for the random effect of year in a mixed effect model. For *C. splendens* males, a model that included a linear effect of wing length and wing patch darkness was the best performing model in terms of AIC (Table 2; Model 5). Model 6, was the second best performing model and included a non-significant overall negative effect of the slope of the thermal reaction norm (thermal plasticity) on male mating success. Thermal plasticity had an overall negative effect on mating success even after including body mass and forewing length in the analysis (Table 2; Model 7). In contrast, in *C. virgo*, we found evidence for sexual selection against thermal plasticity across all years, (Table 3).

Table 2. Thermal plasticity and morphology on mating success in *C. splendens*.

Models of how variation in morphology and thermal physiology influence mating success in *C. splendens* males. All models use data from all years (2013-2016; N = 2133), while controlling for year as a random factor. Across years, thermal plasticity was found to have an overall non-significant negative effect on mating success, even after controlling for the effect of body size (mass and wing length) (model 6). The best performing model (model 5), included only wing length and wing patch darkness, indicating that black color might primarily function as a sexual signal rather than in primarily serving in thermoregulation. Model effects that are significant or nearly significant (P < 0.10) models are indicated in **bold**.

Model	Term	Estimate	SE	P-value	AIC	Rank
1	thermal plasticity	-0.100	0.0685	0.1429	10974.46	8
2	wing length	0.136	0.0342	< 0.0001	10962.27	5
3	patch darkness.	0.321	0.0682	< 0.0001	10954.44	4
4	mass	0.2242	0.0683	0.0011	10965.88	6
5	wing length	0.1327	0.0341	0.0001	10946.24	1
5	patch darkness	0.3163	0.0680	< 0.0001		
6	wing length	0.1306	0.0341	0.0001	10949.82	2
6	patch darkness	0.3182	0.0679	< 0.0001		
6	thermal plasticity	-0.0952	0.0680	0.1615		
7	wing length	0.1166	0.0348	0.0008	10951.49	3
7	patch darkness	0.3042	0.0683	< 0.0001		
7	thermal plasticity	-0.0856	0.0681	0.2092		
7	mass	0.1366	0.0700	0.0511		
8	wing length	-0.1337	0.8265	0.8715	10969.85	7
8	patch darkness	0.3254	0.0900	0.0003		
8	thermal plasticity	-0.1455	0.0869	0.0943		
8	mass	0.1982	0.0763	0.0095		
8	wing length ²	0.0082	0.0281	0.7697		
8	patch darkness ²	0.0473	0.0701	0.4997		
8	thermal plasticity ²	0.0455	0.0435	0.2952		
8	mass ²	-0.1150	0.0495	0.0202		

Table 3. Thermal plasticity and morphology influence on mating success in *C. virgo*.

Modelling morphology and physiology influencing mating success in *C. virgo* males. All models use data from all years (2013–2016; N = 427), while controlling for the random effect of year. Thermal plasticity was found to have an overall negative effect on mating success, even after controlling for the effect of body size (mass and wing length)(model 5). The best performing model (model 1), included only thermal plasticity, which might indicate that heating rate is a composite trait that reflects variation in several underlying traits such as body size, mass, or physiology. Significant effects are indicated in **bold**.

Model	Term	Estimate	SE	P-value	AIC	Rank
1	thermal plasticity	-0.4811	0.1789	0.0074	2333.606	1
2	wing length	0.1261	0.0844	0.1360	2340.065	5
3	mass	-0.2824	0.1799	0.1173	2338.320	4
4	wing length	0.1222	0.0838	0.1456	2336.603	3
4	thermal plasticity	-0.4766	0.1787	0.0080		
5	wing length	0.1764	0.0874	0.0442	2335.774	2
5	thermal plasticity	-0.4708	0.1780	0.0085		
5	mass	-0.3894	0.1864	0.0373		
6	wing length	-0.3938	0.8293	0.6351	2348.984	6
6	thermal plasticity	-0.5183	0.2144	0.0160		
6	mass	-0.4209	0.1897	0.0271		
6	wing length ²	0.0187	0.0271	0.4901		
6	thermal plasticity ²	0.0579	0.1541	0.7074		
6	mass ²	-0.1721	0.2233	0.4412		

Survival selection on thermal performance

We also modelled the effect of various morphological traits and thermal performance on survival in *C. spendens* and *C. virgo* males (Table 3.; Table S3). Data from 2013 – 2016 was used, while incorporating year as a random factor in a mixed-model framework. For *C. splendens* males, the model that included a linear effect of patch density (darkness) was the best performing model in terms of AIC (Table 3 ; Model 3). Model 1, with only a linear effect thermal performance also performed well in terms of AIC (ranked 2nd). Thermal performance in this model had a non-significant negative effect on survival in *C. splendens* males. Thermal performance also continued to have an overall negative effect on survival even after controlling for mass and wing length (Table 3; Model 7). *C. virgo* was also modeled in a similar way and is presented (Table S2).

Selection on thermal performance compared to monthly temperature variation

We plotted the relationship between linear selection on survival and monthly mean temperature at our field site (Fig. 3). In *C. splendens* and *C. virgo* males, we see that there is increasing sexual selection on thermal plasticity at extreme temperatures (at around 22°C and 28°C) (Fig. 3A, C), and no to negative selection on thermal plasticity at medium monthly mean temperatures (23°C – 26°C). We also show survival selection on thermal plasticity in *C. virgo* and *C. splendens* males (Fig. 3 B, D). Modelling the sexual selection data using mixed effect models (controlling for the random effect of year) rather than splitting the data into months, revealed two nearly significant positive interaction terms between thermal plasticity and ambient temperature (Table 4; term: thermal plasticity x temp.²) in both *C. splendens* and *C. virgo* males. This significant interaction indicates that there is a underlying positive curvature to the fitness surface that is temperature dependent. There was no such significant effects of thermal plasticity in terms of survival selection (Table 4).

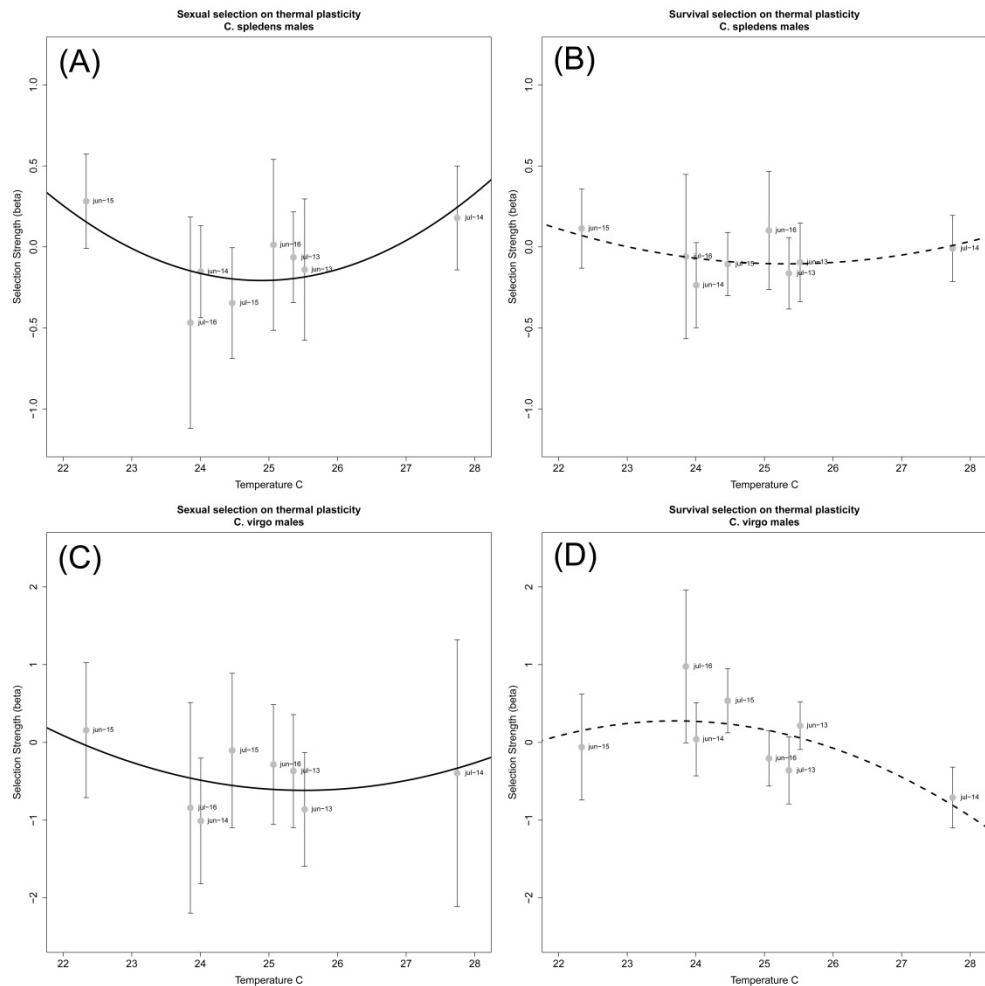


Figure3. Effect of ambient temperature on natural and sexual selection.

The relationship between linear selection on male mating success and male longevity against with monthly mean temperatures at our field site. In *C. splendens* and *C. virgo* males, the strength of sexual selection on thermal plasticity was higher at extreme temperatures (at around 22°C and 28°C) (A and C), and weak negative selection on thermal plasticity at intermediate temperatures (23°C – 26°C). The pattern for survival selection in both *C. virgo* and *C. splendens* males was less clear and not significant (B and D). See Table 4 for statistics.

Table 4. Effect of ambient temperature on natural and sexual selection.

Modeling the interactive effect between ambient temperature and sexual and natural selection on thermal plasticity in *C. splendens* and *C. virgo* males (CSM and CVM, respectively). We modeled the effect of ambient temperature, thermal plasticity and curvature of the plasticity fitness surface in relation to ambient temperature. All terms with $P < 0.1$ are shown in bold. The relationship between temperature and survival selection in *C. splendens* and *C. virgo* males is less clear, compared to the effect on sexual selection. Significant or near-significant ($P < 0.10$) terms are shown in **bold**.

Sp.Sex	Selection	Term	Estimate.	Std. Error.	P	Fig.
CSM	Sexual	thermal plasticity.	-0.2531	0.1032	0.0143	3A
		temperature	0.2714	0.0689	0.0001	
		thermal plasticity ²	0.0133	0.0225	0.5530	
		temperature ²	0.0461	0.0579	0.4264	
		thermal plasticity x temperature	-0.0664	0.0622	0.2854	
		thermal plasticity x temperature. ²	0.0941	0.0498	0.0589	
CSM	Survival	thermal plasticity	-0.0448	0.0711	0.5289	3B
		temperature	-0.1226	0.0475	0.0099	
		thermal plasticity ²	-0.0069	0.0155	0.6553	
		temperature ²	0.0177	0.0399	0.6570	
		thermal plasticity x temperature	-0.0238	0.0428	0.5786	
		thermal plasticity x temperature ²	0.0015	0.0343	0.9662	
CVM	Sexual	thermal plasticity	-0.9033	0.2955	0.0024	3C
		temperature	0.6804	0.1773	0.0001	
		thermal plasticity ²	0.0707	0.0803	0.3787	
		temperature ²	0.0219	0.1543	0.8873	
		thermal plasticity x temperature	-0.3934	0.1543	0.0367	
		thermal plasticity x temperature ²	0.2993	0.1543	0.0529	
CVM	Survival	thermal plasticity	0.1411	0.1404	0.3155	3D
		temperature	-0.0592	0.0843	0.4829	
		thermal plasticity ²	0.0400	0.0381	0.2954	
		temperature ²	-0.0443	0.0733	0.5460	
		thermal plasticity x temperature	0.0447	0.0892	0.6167	
		thermal plasticity x temperature ²	-0.1056	0.0892	0.1502	

Discussion

Ectotherms like damselflies and many other insects are highly dependent on ambient temperature to maintain beneficial internal body temperatures for flight and function (Angilletta 2009; Heinrich 2013). Here, we quantified natural and sexual selection on thermal plasticity in two sympatric and closely related species of damselflies.

We find that *C. virgo* and *C. splendens* individuals have an optimal mating rate between 20°C and 25°C (Table 1; Fig. 2; A,B). Although plots of mating rate of the two species appear to be different (Fig. 2; A,B), we found no statistically significant difference in how these two species responded to ambient temperature in terms of mating rate (Table 1: i.e. no significant Temperature² x Species-term). Survivorship in *C. splendens* males peaked at temperatures between 20 – 25°C, whereas higher or lower temperatures reduced male lifespan in the field (Fig. 1C). Similarly, *C. virgo* males had reduced lifespan at high and low temperatures, with a near-significant tendency for decreasing survival rate with increasing temperature (Table 1: Temperature² x Species: $P = 0.087$).

We quantified how selection operated on the thermal reaction norms in these two phenotypically similar species (Tables 2-3). We found that, across all the years of this study, there was some evidence of sexual selection against thermal plasticity in both *C. splendens* and *C. virgo* (Tables 2-3). Thus, males of both species which had shallower thermal reaction norms had the highest mating success, suggesting sexual selection for thermal canalization. In other words, sexual selection against thermal plasticity. In contrast, there was no evidence that natural selection through adult longevity favored either increased or decreased thermal plasticity (Tables S2-S3). These results might suggest that mating success, and potentially reproductive traits in general, are more sensitive to thermally stressful environments, which can lead to selection for thermal canalization. These results might also suggest that high thermal sensitivity and thermal plasticity could be a maladaptive form of plasticity that is currently opposed by sexual selection for thermal canalization.

Some evolutionary biologists have argued that phenotypic plasticity is mostly adaptive, and plasticity is often invoked as a creative factor in evolution that could enhance long-term adaptation, diversification and speciation (West-Eberhard 2003; Pigliucci et al. 2006; Pfennig et al. 2010; Moczek et al. 2011). These optimistic views about the adaptive nature of phenotypic plasticity do, however, largely rely on verbal arguments rather than on formal models. Rigorous theoretical and empirical studies demonstrating the adaptive nature of phenotypic plasticity in natural populations are far outnumbered by verbal arguments in many reviews claiming that plasticity is adaptive. More formal mathematical models of the evolution of phenotypic plasticity emphasize the need to directly quantify

selection on reaction norms themselves and relate such selection to patterns of environmental variation (Lande 2009; Chevin et al. 2010; Chevin and Lande 2011).

Selection on plastic traits, such as thermal reaction norms, can be formalized as selection on function-valued traits (Kingsolver et al. 2001; Stinchcombe et al. 2012). One could even argue that it is only when the slopes of a thermal reaction norms are subject to stabilizing selection around the population mean that the current level of plasticity should be viewed as adaptive. The fact that we did not find any evidence for stabilizing selection on reaction norm slope for either species in this study (Tables 2-3), suggests that the observed level of plasticity is not adaptive. Instead of stabilizing selection we found evidence for directional sexual selection *against* thermal plasticity and for environmental canalization (Tables 2-3). Our results should caution against assuming that just because traits exhibit phenotypic plasticity, that such plasticity must be necessarily adaptive. Instead, selection can operate to reduce plasticity and favor increased canalization. Other examples of maladaptive plasticity have recently been found or suggested for guppies (Ghalambor et al. 2015) and soapberry bugs (Cenzer 2017).

We hypothesized that selection on thermal plasticity would increase if individuals found themselves in environments which were on average higher or lower than their optimal range. In both *C. splendens* and *C. virgo* males, we found that, selection for increased thermal plasticity occurs when temperature conditions are less than around 23°C and greater than 26°C (Fig. 3A-C; Table 4). This implies that individuals that are better able to heat up in cold environments and cool down in warm environments will have higher mating success.

In environments that are already close to optimal body temperatures, there should be weak or negative selection on thermal plasticity, since individuals will already be at optimal body temperatures. If an individual is already at an optimal body temperature, any sudden change in ambient temperature should be resisted. However, if an individual is at a non-optimal temperature, that individual should quickly adapt to a sudden change in ambient temperature because that change will likely be in the direction towards an optimal temperature. We found less strong patterns when looking at survival selection on thermal plasticity, which implies that thermoregulatory ability might not be as important for survival (Fig 3; B, D and Table 4). From a life history perspective, thermal performance might be less important in survival selection. Those individuals at risk of dying from overheating, will likely seek out cooler micro-environments, and damselflies can typically survive cold nights of less than 10°C in Sweden.

Whether selection in extreme environments will lead to a change in mean thermal plasticity depends on the quantitative genetic architecture of thermal plasticity and correlated traits. Recent experimental work shows that there might be some limited amount of genetic variation in heating rate but not in thermal performance curves for mating rate in *Drosophila melanogaster* (Waller et al.

2017). Genetic covariance between thermal plasticity and other phenotypic traits might also affect expected evolutionary responses due to selection (Morrissey 2014). For example, it is known that body size in damselflies is under positive selection in the adult stage (Svensson and Waller 2017), and larger damselflies will also heat up slower (Table S1). The net effect of such selection for larger body size might restrict the evolution of thermal plasticity in damselflies because large body size will also increase thermal inertia (Table S1).

Conclusion

Thermal plasticity is an important trait in male damselflies for mating success in extreme temperature conditions. We found that, selection for increased thermal plasticity occurs when temperature conditions are on the extremes of this range (less than around 23°C and greater than 26°C). These results suggest that in environments that are already close to optimal body temperatures, there might even be selection against thermal plasticity, since individuals will already be at optimal body temperatures most of the time. Our results might also have some implications for the evolution of thermal traits in general and in less variable thermal environments, such as in the tropics. In thermally stable environments, selection might not remove shallow or steep thermal reaction norms, allowing the build-up of “cryptic” genetic variation that will only be exposed to selection following invasion in more thermally variable environments. Once organisms have invaded more variable environments, selection can then either favour reduced thermal plasticity and increased thermal inertia (as in this study), or increased thermal plasticity in other situations (see Lande 2009; 2014; Chevin and Lande 2010 for in-depth discussions).

References

- Angilletta, M.J. (2009). *Thermal adaptation: a theoretical and empirical synthesis*. Oxford University Press.
- Ashander, J., Chevin, L.M. and M.L. Baskett (2016). Predicting evolutionary rescue via evolving plasticity in stochastic environments. *Proc. R. Soc. B*. 283: e20161690.
- Blouin-Demers, G. and P. Nadeau. (2005). The cost–benefit model of thermoregulation does not predict lizard thermoregulatory behavior. *Ecology*. 86: 560-566.
- Buckley, L.B., Ehrenberger, J.C. and M.J. Angilletta. (2015). Thermoregulatory behaviour limits local adaptation of thermal niches and confers sensitivity to climate change. *Functional Ecology*. 29:1038-1047.
- Cenzer, M. L. (2017). Maladaptive Plasticity masks the effects of natural selection in the red-Shouldered soapberry bug. *American Naturalist*. 190:521-533.
- Charmantier, A., McCleery, R.H., Cole, L.R., Perrins, C., Kruuk, L.E. and B.C. Sheldon. (2008). Adaptive phenotypic plasticity in response to climate change in a wild bird population. *Science*. 320: 800-803.
- Chevin, L.M. and A.A. Hoffmann. (2017). Evolution of phenotypic plasticity in extreme environments. *Phil. Trans. R. Soc. B*. 372: e20160138.
- Chevin, L.M. and R. Lande. (2011). Adaptation to marginal habitats by evolution of increased phenotypic plasticity. *Journal of evolutionary biology*. 24: 1462-1476.
- Chevin, L.M. and R. Lande. (2011). Adaptation to marginal habitats by evolution of increased phenotypic plasticity. *Journal of evolutionary biology*. 24: 1462-1476.
- Chevin, L.M., Gallet, R., Gomulkiewicz, R., Holt, R.D. and S. Fellous. (2013). Phenotypic plasticity in evolutionary rescue experiments. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*. 368: e20120089.
- Chevin, L.M., Lande, R. and G. M. Mace. (2010). Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory. *PLoS biology*. 8:e1000357.
- Clark, P.U., Dyke, A.S., Shakun, J.D., Carlson, A.E., Clark, J., Wohlfarth, B., Mitrovica, J.X., Hostetler, S.W. and A.M. McCabe. (2009). The last glacial maximum. *Science*. 325: 710-714.
- Córdoba-Aguilar, A. (2002). Wing pigmentation in territorial male damselflies, *Calopteryx haemorrhoidalis*: a possible relation to sexual selection. *Animal Behaviour*. 63: 759-766.
- Córdoba-Aguilar, A., Salamanca-Ocaña, J.C. and M. Lopezaraiza. (2003). Female reproductive decisions and parasite burden in a calopterygid damselfly (Insecta: Odonata). *Animal Behaviour*. 66: 81-87.

- Deutsch, C.A., Tewksbury, J.J., Huey, R.B., Sheldon, K.S., Ghalambor, C.K., Haak, D.C. and P.R. Martin. (2008). Impacts of climate warming on terrestrial ectotherms across latitude. *Proceedings of the National Academy of Sciences*. 105: 6668-6672.
- Gavrilets, S. and S.M. Scheiner. (1993). The genetics of phenotypic plasticity. V. Evolution of reaction norm shape. *Journal of evolutionary biology*. 6:31-48.
- Ghalambor, C.K., Hoke, K.L., Ruell, E.W., Fischer, E.K., Reznick, D.N. and K. A. Hughes. (2015). Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature. *Nature*. 525: 372-375.
- Heinrich, B. (2013). *The hot-blooded insects: strategies and mechanisms of thermoregulation*. Springer Science & Business Media.
- Hoffmann, A.A. and C.M. Sgrò. (2011). Climate change and evolutionary adaptation. *Nature*. 470:479-485.
- Hoffmann, A.A., Hallas, R.J., Dean, J.A. and M. Schiffer. (2003). Low potential for climatic stress adaptation in a rainforest *Drosophila* species. *Science*. 301:100-102.
- Huey, R.B. and D. Berrigan. (2001). Temperature, demography, and ectotherm fitness. *American Naturalist*. 158: 204-210.
- Huey, R.B. and M. Slatkin. (1976). Cost and benefits of lizard thermoregulation. *The Quarterly Review of Biology*. 51: 363-384.
- Huey, R.B., Deutsch, C.A., Tewksbury, J.J., Vitt, L.J., Hertz, P.E., Pérez, H.J.Á. and T. Garland. (2009). Why tropical forest lizards are vulnerable to climate warming. *Proceedings of the Royal Society of London B: Biological Sciences*. 276: 1939-1948.
- Huey, R.B., Hertz, P.E. and , B. Sinervo. (2003). Behavioral drive versus behavioral inertia in evolution: a null model approach. *American Naturalist*. 161:357-366.
- Kearney, M., Shine, R. and W.P. Porter. (2009). The potential for behavioral thermoregulation to buffer “cold-blooded” animals against climate warming. *Proceedings of the National Academy of Sciences*. 106: 3835-3840.
- Kellermann, V., van Heerwaarden, B., Sgrò, C.M. and A.A. Hoffmann. (2009). Fundamental evolutionary limits in ecological traits drive *Drosophila* species distributions. *Science*. 325: 1244-1246.
- Kingsolver, J.G. and R.B. Huey. (2008). Size, temperature, and fitness: three rules. *Evolutionary Ecology Research*. 10: 251-268.
- Kingsolver, J.G., Gomulkiewicz, R. and P. A. Carter. (2001). Variation, selection and evolution of function-valued traits. *Microevolution Rate, Pattern, Process*. 8: 87-104.
- Lancaster, L.T., Dudaniec, R.Y., Hansson, B. and E. I. Svensson. (2015). Latitudinal shift in thermal niche breadth results from thermal release during a climate-mediated range expansion. *Journal of Biogeography*. 42:1953-1963.
- Lande, R. (2009). Adaptation to an extraordinary environment by evolution of phenotypic plasticity and genetic assimilation. *Journal of evolutionary biology*. 22:1435-1446.
- Lande, R. (2014). Evolution of phenotypic plasticity and environmental tolerance of a labile quantitative character in a fluctuating environment. *Journal of evolutionary biology*. 27:866-875.

- Liefting, M., Hoffmann, A.A. and J. Ellers. (2009). Plasticity versus environmental canalization: population differences in thermal responses along a latitudinal gradient in *Drosophila serrata*. *Evolution*. 63:1954-1963.
- Lynch, M. and B. Walsh. (1998). *Genetics and analysis of quantitative traits* (Vol. 1). Sunderland, MA: Sinauer.
- Marden, J., Kramer, M. and J. Frisch. (1996). Age-related variation in body temperature, thermoregulation and activity in a thermally polymorphic dragonfly. *Journal of experimental Biology*. 199: 529-535.
- Marden, J.H. (2000). Variability in the size, composition, and function of insect flight muscles. *Annual Review of Physiology*. 62: 157-178.
- Marden, J.H. and J.K. Waage. (1990). Escalated damselfly territorial contests are energetic wars of attrition. *Animal Behaviour*. 39: 954-959.
- Martin, T.L. and R.B. Huey. (2008). Why suboptimal is optimal: Jensen's inequality and ectotherm thermal preferences. *American Naturalist*. 171:102-118.
- May, M.L. (1978). Thermal adaptations of dragonflies. *Odonatologica*. 7: 27-47.
- May, M.L. (1979). Insect thermoregulation. *Annual review of entomology*. 24:313-349.
- May, M.L. (1991). Thermal adaptations of dragonflies, revisited. *Advances in Odonatology*. 5:71-88.
- Misof, B., Anderson, C.L. and H. Hadrys. (2000). A phylogeny of the damselfly genus *Calopteryx* (Odonata) using mitochondrial 16S rDNA markers. *Molecular Phylogenetics and Evolution*. 15: 5-14.
- Mitchell, K.A., Sgrò, C.M. and A.A. Hoffmann. (2011). Phenotypic plasticity in upper thermal limits is weakly related to *Drosophila* species distributions. *Functional Ecology*. 25: 661-670.
- Moczek, A.P., Sultan, S., Foster, S., Ledón-Rettig, C., Dworkin, I., Nijhout, H.F., Abouheif, E. and D. W. Pfennig. (2011). The role of developmental plasticity in evolutionary innovation. *Proceedings of the Royal Society of London B: Biological Sciences*. 278:2 705-2713.
- Morrissey, M.B. (2014) Selection and evolution of causally covarying traits. *Evolution*. 68.6:1748-1761.
- Norling U. (1984). Life history patterns in the northern expansion of dragonflies. *Advances in Odonatology* 2:127-156
- Pfennig, D. W., Wund, M. A., Snell-Rood, E. C., Cruickshank, T., Schlichting, C.D. and A. P. Moczek. (2010). Phenotypic plasticity's impacts on diversification and speciation. *Trends in ecology & evolution*. 25: 459-467.
- Pigliucci, M., Murren, C.J. and C. D. Schlichting. (2006). Phenotypic plasticity and evolution by genetic assimilation. *Journal of Experimental Biology*. 209: 2362-2367.
- Plaistow, S., and M.T. Siva-Jothy. (1996). Energetic constraints and male mate-securing tactics in the damselfly *Calopteryx splendens xanthostoma* (Charpentier). *Proceedings of the Royal Society of London B: Biological Sciences*. 263: 1233-1239.
- Polcyn, D. M. (1994). Thermoregulation during summer activity in Mojave Desert dragonflies (Odonata: Anisoptera). *Functional Ecology*. 1:441-449.

- Porter, K. (1982). Basking behaviour in larvae of the butterfly *Euphydryas aurinia*. *Oikos*. 38.3:308-312.
- Rantala, M.J., Koskimäki, J., Taskinen, J., Tynkkynen, K. and J. Suhonen. (2000). Immunocompetence, developmental stability and wingspot size in the damselfly *Calopteryx splendens* L. *Proceedings of the Royal Society of London B: Biological Sciences*. 267: 2453-2457.
- Sanborn, A. (2004). Thermoregulation in Insects. *Encyclopedia of Entomology*. Springer Netherlands.
- Sinervo, B., Mendez-De-La-Cruz, F., Miles, D.B., Heulin, B., Bastiaans, E., Villagrán-Santa Cruz, M., Lara-Resendiz, R., Martínez-Méndez, N., Calderón-Espinosa, M.L., Meza-Lázaro, R.N. and H. Gadsden. (2010). Erosion of lizard diversity by climate change and altered thermal niches. *Science*. 328: 894-899.
- Siva-Jothy, M.T. (2000). A mechanistic link between parasite resistance and expression of a sexually selected trait in a damselfly. *Proceedings of the Royal Society of London B: Biological Sciences*. 267: 2523-2527.
- Somero, G. N. (2010). The physiology of climate change: how potentials for acclimatization and genetic adaptation will determine ‘winners’ and ‘losers’. *Journal of Experimental Biology*. 213: 912-920.
- Stevenson, R.D., and R.K. Josephson. (1990). Effects of operating frequency and temperature on mechanical power output from moth flight muscle. *Journal of Experimental Biology*. 149:61-78.
- Stinchcombe, J.R., Kirkpatrick, M. and Function-valued Traits Working Group. (2012). Genetics and evolution of function-valued traits: understanding environmentally responsive phenotypes. *Trends in Ecology & Evolution*. 27: 637-647.
- Sunday, J.M., Amanda E.B., and K.D. Nicholas (2011). Global analysis of thermal tolerance and latitude in ectotherms. *Proceedings of the Royal Society of London B: Biological Sciences*. 278: 1823-1830.
- Svensson, E.I., and J.T. Waller. (2013). Ecology and sexual selection: evolution of wing pigmentation in calopterygid damselflies in relation to latitude, sexual dimorphism, and speciation. *American Naturalist* 182:174-195.
- Svensson, E.I., F., Eróukhmanoff, and M. Friberg. (2006). Effects of natural and sexual selection on adaptive population divergence and premating isolation in a damselfly. *Evolution*. 60: 1242-1253.
- Svensson, E.I., Kristoffersen, L., Oskarsson, K. and S. Bensch. (2004). Molecular population divergence and sexual selection on morphology in the banded demoiselle (*Calopteryx splendens*). *Heredity*. 93: 423-433.
- Trotta, V., Calboli, F.C., Ziosi, M., Guerra, D., Pezzoli, M.C., David, J.R. and S. Cavicchi. (2006). Thermal plasticity in *Drosophila melanogaster*: a comparison of geographic populations. *BMC Evolutionary Biology*. 6:67-86.
- True, J.R. (2003). Insect melanism: the molecules matter. *Trends in ecology & evolution*. 18: 640-647.
- Trullas, S.C., J.H., van Wyk, and J.R. Spotila. (2007). Thermal melanism in ectotherms. *Journal of Thermal Biology*. 32:235-245.

- Vedder, O., Bouwhuis, S. and B.C. Sheldon. (2013). Quantitative assessment of the importance of phenotypic plasticity in adaptation to climate change in wild bird populations. *PLoS Biology*. 11:e1001605.
- Waller, J.T., Kell, A., Ballesta, M., Giraud, A., Abbott, J. and E.I. Svensson. (2017). Limited genetic variation for male mating success reveals low evolutionary potential for thermal plasticity in *Drosophila melanogaster*. *bioRxiv*. e166801.
- West-Eberhard, M. J. (2003). *Developmental plasticity and evolution*. Oxford University Press.
- Wiens, J.J. (2016). Climate-related local extinctions are already widespread among plant and animal species. *PLoS biology*. 14:e2001104.

Supplementary Material

Table S1. Factors influencing heating rate in *C. splendens* males.

Modelling the factors affecting thermal plasticity (slope of thermal reaction norms) in *C. splendens* males. We used mixed models with data from all years (2013-2016; N = 2133), while controlling for the random effect of year. We tested models with two environmental factors (solar radiation and ambient temperature) and body mass and wing color patch darkness (patch density). All variables have been standardized to mean zero and unit variance. The best performing model (model 4) included a positive effect of solar radiation and ambient temperature and a negative effect of body mass. Wing patch darkness also had a positive effect on heating rate when included in the model (model 6).

Model	Term	Est.	SE	P-value	AIC	Rank
1	Temperature	0.0948	0.0064	< 0.0001	873.317	3
2	Solar radiation	0.0382	0.0067	< 0.0001	1049.35	5
3	Mass	-0.0094	0.0067	0.1616	1079.99	6
4	Temperature	0.0929	0.0064	< 0.0001	865.015	1
4	Solar radiation	0.0302	0.0064	< 0.0001		
4	Mass	-0.0205	0.0064	0.0014		
5	Temperature	0.0913	0.0064	< 0.0001	874.004	4
5	Solar radiation	0.0286	0.0064	< 0.0001		
5	Patch darkness	0.0070	0.0064	0.2725		
6	Temperature	0.0926	0.0064	< 0.0001	873.045	2
6	Solar radiation	0.0314	0.0064	< 0.0001		
6	Mass	-0.0217	0.0064	0.0008		
6	Patch darkness	0.0096	0.0064	0.1363		

Table S2. Thermal plasticity and survival in *C. splendens*.

Mixed models showing how morphology and thermal physiology influence survival in *C. splendens* males. All models use data from all years (2013-2016; N = 2133), while incorporating year as a random factor. Patch density had a large positive effect on survival in males (model 3), implying that it might be an honest indicator of condition. Thermal plasticity had an overall negative effect on survival (model 1), even after controlling for the effect of body size (wing length and mass) (models 6, 7, 8).

Model	Term	Est.	SE	P-value	AIC	Rank
1	thermal plasticity	-0.0634	0.0469	0.1769	9361.736	2
2	wing length	0.0232	0.0235	0.3243	9363.969	4
3	patch darkness	0.0832	0.0469	0.0762	9360.415	1
4	mass	-0.0148	0.0469	0.7527	9363.460	3
5	wing length	0.0223	0.0469	0.3421	9367.174	5
5	patch darkness	0.0823	0.0469	0.0796		
6	wing length	0.0210	0.0241	0.3728	9371.644	6
6	patch darkness	0.0835	0.0469	0.0751		
6	thermal plasticity	-0.0632	0.0469	0.1786		
7	wing length	0.0251	0.0241	0.2977	9377.193	7
7	patch darkness	0.0876	0.0472	0.0635		
7	thermal plasticity	-0.0660	0.0471	0.1613		
7	mass	-0.0396	0.0483	0.4124		
8	wing length	0.0912	0.5716	0.8732	9403.519	8
8	patch darkness	0.0827	0.0622	0.1839		
8	thermal plasticity	-0.0487	0.0601	0.4175		
8	mass	-0.0646	0.0528	0.2213		
8	wing length ²	-0.0021	0.0194	0.9122		
8	patch darkness ²	-0.0146	0.0485	0.7626		
8	therm. plasticity ²	-0.0127	0.0301	0.6738		
8	mass ²	0.0451	0.0342	0.1873		

Table S3. Mixed models of how morphology and thermal physiology influence survival (longevity) in *C. virgo* males.

Mixed models of how morphology and thermal physiology influence survival (longevity) in *C. virgo* males. All models use data from all years (2013-2016; N = 427), while incorporating year as a random factor. Body mass had a large positive effect on survival in males (model 3). Thermal plasticity had an overall positive effect on survival (model 1), even after controlling for the effect of body size (forewing length and body mass) (models 4, 5, 6).

Model	Term	Estimate	SE	P-value	AIC	Rank
1	thermal plasticity	0.0881	0.0832	0.2906	1683.223	2
2	wing length	0.0300	0.0391	0.4438	1685.268	3
3	mass	0.1307	0.0831	0.1165	1681.875	1
4	wing length	0.0307	0.0391	0.4328	1689.254	4
4	thermal plasticity	0.0892	0.0833	0.2847		
5	wing length	0.0138	0.0409	0.7357	1692.364	5
5	thermal plasticity	0.0874	0.0832	0.2941		
5	mass	0.1211	0.0871	0.1651		
6	wing length	-0.0533	0.3870	0.8904	1708.690	6
6	thermal plasticity	0.0029	0.1000	0.9768		
6	Mass	0.1120	0.0885	0.2067		
6	wing length ²	0.0022	0.0127	0.8614		
6	thermal plasticity ²	0.1067	0.0719	0.1387		
6	mass ²	-0.0661	0.1042	0.5263		

List of Papers

- I. Waller, J., & Svensson, E.I. (2016). The measurement of selection when detection is imperfect: How good are naïve methods?. *Methods in Ecology and Evolution*, 7(5), 538-548.
- II. Waller, J., Willink B., Tschol M., & Svensson E.I. (2018). The Odonate Phenotypic Database: a new open data resource for comparative studies of an old insect order. www.odonatephenotypicdatabase.org. (Manuscript).
- III. Waller, J., & Svensson, E.I. (2017). Body size evolution in an old insect order: No evidence for Cope's Rule in spite of fitness benefits of large size. *Evolution*.
- IV. Waller, J. (2018). Is the blunderbuss a misleading visual metaphor for stasis and punctuated evolution? (Manuscript).
- V. Waller, J., & Svensson, E.I. (2017). Biotic and abiotic drivers of global body size distributions and macroevolutionary diversification in an old insect order. (Manuscript).
- VI. Waller, J., Kell, A., Ballesta, M., Giraud, A., Abbott, J., & Svensson, E. (2017). Limited genetic variation for male mating success reveals low evolutionary potential for thermal plasticity in *Drosophila melanogaster*. *bioRxiv*, 166801.
- VII. Waller, J. & Svensson, E.I. (2018). Sexual and natural selection on thermal plasticity and thermal canalization in insects at high latitudes. (Manuscript).

